



Preface

Acknowledgements

Welcome to East, a software platform for the statistical design, simulation and monitoring of clinical trials.

The current release of East (version 6.4) was developed by a team comprising (in alphabetical order): Gordhan Bagri, Dhaval Bapat, Priyanka Bhosle, Jim Bolognese, Sudipta Basu, Jaydeep Bhattacharyya, Swechhya Bista, Apurva Bodas, Pushkar Borkar, V. P. Chandran, Soorma Das, Pratiksha Deoghare, Aniruddha Deshmukh, Namita Deshmukh, Yogesh Dhanwate, Suraj Ghadge, Pranab Ghosh, Karen Han, Aarati Hasabnis, Pravin Holkar, Munshi Imran Hossain, Abhijit Jadhav, Yogesh Jadhav, Prachi Jagtap, Paridhi Jain, Yannis Jemiai, Ashwini Joshi, Nilesh Kakade, Janhavi Kale, Aditya Kamble, Anthiyur Kannappan, Parikshit Katikar, Uday Khadilkar, Kapildev Koli, Yogita Kotkar, Hrishikesh Kulkarni, Mandar Kulkarni, Mangesh Kulkarni, Shailesh Kulthe, Charles Liu, Lingyun Liu, Shashank Maratkar, Cyrus Mehta, Pradoshkumar Mohanta, Manashree More, Tejal Motkar, Ankur Mukherjee, Nabeela Muzammil, Neelam Nakadi, Vijay Nerkar, Sandhya Paranjape, Gaurangi Patil, Vidyadhar Phadke, Anup Pillai, Shital Pokharkar, Vidyagouri Prayag, Achala Sabane, Sharad Sapre, Rohan Sathe, Pralay Senchaudhuri, Rhiannon Sheaparé, Pradnya Shinde, Priyadarshan Shinde, Sumit Singh, Sheetal Solanki, Chitra Tirodkar, Janhavi Vaidya, Shruti Verma, Pantelis Vlachos, Suchita Wageshwari, Kiran Wadje, Ritika Yadav.

Others contributors to this release include Asmita Ghatnekar, Sam Hsiao, Brent Rine, Ajay Sathe, Chinny Swamy, Nitin Patel, Yogesh Gajjar, Shilpa Desai.

Other contributors who worked on previous releases of East: Gayatri Bartake, Ujwala Bamishte, Apurva Bhingare, Bristi Bose, Chandrashekhar Budhwant, Krisnaiah Byagari, Vibhavari Deo, Rupali Desai, Namrata Deshpande, Yogesh Deshpande, Monika Ghatage, Ketan Godse, Vishal Gujar, Shashikiran Halvagal, Niranjan Kshirsagar, Kaushal Kulkarni, Nilesh Lanke, Manisha Lohokare, Jaydip Mukhopadhyay, Abdulla Mulla, Seema Nair, Atul Paranjape, Rashmi Pardeshi, Sanket Patekar, Nabarun Saha, Makarand Salvi, Abhijit Shelar, Amrut Vaze, Suryakant Walunj, Sanhita Yeolekar.

We thank all our beta testers for their input and obvious enthusiasm for the East software. They are acknowledged by name in Appendix Z.

We owe a debt of gratitude to Marvin Zelen and to Swami Sarvagatananda, special

people whose wisdom, encouragement and generosity have inspired Cytel for over two decades.

Finally, we dedicate this software package to our families and to the memory of our dearly departed Stephen Lagakos and Aneesh Patel.

Our Philosophy

We would like to share with you what drives and inspires us during the research and development stages of the East software.

Empower, do not Frustrate

We believe in making simple, easy-to-use software that empowers people.

We believe that statisticians have a strategic role to play within their organization and that by using professionally developed trial design software they will utilize their time better than if they write their own computer programs in SAS or R to create and explore complex trial designs. With the help of such software they can rapidly generate many alternative design options that accurately address the questions at hand and the goals of the project team, freeing time for strategic discussions about the choice of endpoints, population, and treatment regimens.

We believe that software should not frustrate the user's attempt to answer a question. The user experience ought to engage the statistician and inspire exploration, innovation, and the quest for the best design. To that end, we believe in the following set of principles:

- **Fewer, but Important and Useful Features** It is better to implement fewer, but important and useful features, in an elegant and simple-to-use manner, than to provide a host of options that confuse more than they clarify. As Steve Jobs put it: 'Innovation is not about saying "Yes" to everything. It's about saying "No" to all but the most crucial features.'
- **Just because we Can, doesn't mean we Should** Just because we can provide functionality in the software, doesn't mean we should.
- **Simplify, Simplify, Simplify** Find and offer simple solutions - even for the most complex trial design problems.
- **Don't Hurry, but continually Improve** Release new solutions when they are ready to use and continually improve the commercial releases with new features, bug fixes, and better documentation.
- **Provide the best Documentation and Support** Our manuals are written like textbooks, to educate, clarify, and elevate the statistical knowledge of the user.

Preface

Our support is provided by highly competent statisticians and software engineers, focusing on resolving the customer's issue, and being mindful of the speed and quality requirements. We believe that delivering *delightful customer support is essential to our company's lifeblood*.

Finally, we listen to our customers constantly and proactively through countless informal and formal interactions, software trainings, and user group meetings. This allows us to follow all the principles laid out above in the most effective manner.

Assess

It is essential to be able to assess the benefits and flaws of various design options and to work one's way through a sensitivity analysis to evaluate the robustness of design choices. East can very flexibly generate multiple fixed sample size, group sequential, and other adaptive designs at a click of a button. The wealth of design data generated in this manner requires new tools to preview, sort, and filter through in order to make informed decisions.

Share

Devising the most innovative and clever designs is of no use if the statistician is unable to communicate in a clear and convincing manner what the advantages and characteristics of the design are for the clinical trial at hand. We believe statistical design software tools should also be communication tools to share the merits of various trial design options with the project team and encourage dialog in the process.

The many graphs, tables, simulation output, and other flexible reporting capabilities of East have been carefully thought out to provide clear and concise communication of trial design options in real time with the project team.

Trust

East has been fully validated and intensely tested. In addition, the East software package has been in use and relied upon for almost 20 years. East has helped design and support countless actual studies at all the major pharmaceutical and biotech companies, academic research centers, and government institutions.

We use and rely on our software every day in our consulting activities to collaborate with our customers, helping them optimize and defend their clinical trial designs. This also helps us quickly identify things that are frustrating or unclear, and improve them fast - for our own sake and that of our customers.

What's New in East 6.4

Version 6.4 of East introduces some important new features:

1. **Multi-arm multi-stage designs** East now offers the ability to design multi-arm multi-stage studies with options for early stopping, dose selection, and sample size re-estimation. The group sequential procedures (Gao et al., 2014) have been implemented for normal endpoint whereas the p-value combination approaches (Posch et al. 2005) have been implemented for both normal and binomial endpoints. See Chapters 17, 18 and 29 for more details.
2. **Multiple endpoints designs for binomial endpoints** Gatekeeping procedures to control family-wise type-1 error when testing multiple families of binomially distributed endpoints are now available in East for fixed sample (1-look) designs. East will also use the intersection-union test when testing a single family of endpoints. See Chapter 16 and 28 for more details.
3. **Multi-arm designs for survival endpoints** Designs for pairwise comparisons of treatment arms to control have been added for survival endpoints. See Chapter 51 for more details.
4. **Enrollment and event prediction** East now includes options to predict enrollment and events based on accumulating blinded data and summary statistics. Prediction based on unblinded data was already implemented in the previous version so the current version provides both options - Unblinded as well as Blinded. See Chapter 68 for more details.
5. **Dual agent dose-escalation designs** This version of East adds methods to the Escalate module for dual-agent dose-escalation designs, including the Bayesian logistic regression model (BLRM; Neuenschwander et al., 2014), and the Product of Independent beta Probabilities dose Escalation (PIPE; Mander et al., 2015). Numerous feature enhancements have also been made to the existing single-agent dose escalation designs. See Chapter 32 for more details.
6. **Bayesian probability of success (assurance) and predictive power for survival designs**
East 6.4 will now calculate assurance (O'Hagan et al., 2005), or Bayesian probability of success, and predictive power for survival endpoints. See Chapter 48 for more details.
7. **Interim monitoring using Muller and Schafer method** East6.4 will now provide the capability of monitoring clinical trials using the adaptive approach. It can be done using the Muller and Schafer method. Currently, this feature is

Preface

available for Survival Endpoint tests only. See Chapter 56 for more details.

8. **General usability enhancements** Numerous enhancements have been made to the software to improve the user experience and workflow.

What's New in East 6.3

Version 6.3 of East introduces some important new features:

1. **Updates to Promising Zone designs: Ratio of Proportions designs; Müller and Schäfer type-1 error control method; Estimation**
East 6.3 introduces Promising Zone designs for the ratio of proportions. East 6.3 also implements the method of Müller and Schäfer (2001) to control type-1 error for adaptive unblinded sample size re-estimation designs. This is available for simulation and interim monitoring. Also estimation using Repeated Confidence Intervals (RCI) and Backward Image Confidence Intervals (BWCI) (Gao, Liu & Mehta, 2013) are available in Müller and Schäfer simulations. See Chapter 52 for more details.
2. **Multiple endpoint designs**
Parallel gatekeeping procedures to control family-wise type-1 error when testing multiple families of normally distributed endpoints are now available in East for fixed sample (1-look) designs. East will also use the intersection-union test when testing a single family of endpoints. See Chapter 16 for more details.
3. **Exact designs for binomial endpoints**
East now includes the ability to use the exact distribution when computing power and samples size for binomial endpoints. This applies for all binomial tests in the case of fixed designs. In addition, group sequential exact designs are available for the single proportion case, and the Simon's two-stage optimal and minimax designs (Simon, 1989) have been implemented that allow for early futility stopping while optimizing the expected sample size and the maximum sample size, respectively. See Chapter 33 for more details.
4. **Dose escalation designs**
East 6.3 now includes a module for the design, simulation, and monitoring of modern dose-escalation clinical trials. Model-based dose-escalation methods in this module include the Continual Reassessment Method (mCRM; Goodman et al., 1995), the Bayesian logistic regression model (BLRM; Neuenschwander et al., 2008), and the modified Toxicity Probability Interval (mTPI; Ji et al., 2010). See Chapter 32 for more details.
5. **Predictive interval plots, conditional simulations, , and enrolment/events**

prediction

East 6.3 now includes a module that offers the ability to simulate and forecast the future course of the trial based on current data. This includes conditional simulations to assess expected treatment effects and associated repeated confidence intervals at future looks (also called Predicted Interval Plots or PIP; Li et al. 2009), as well as the probability of finishing with a successful trial (conditional power). You can also plan and simulate clinical trials with greater precision using different accrual patterns and response information for different regions/sites. East allows you to make probabilistic statements about accruals, events, and study duration using Bayesian models and accumulating data. See Chapters 65, 66 and 67 for more details.

6. Sample size and information calculators

Sample size and information calculators have been added back into East to allow easy calculation of the two quantities. See Chapter 59 for more details.

7. Exporting/Importing between East and East Procs

East 6.3 designs can now be exported to work with the newly released East Procs. The output from East Procs can be imported back into East 6.3 for use in the East Interim Monitoring dashboard and to conduct conditional inference and simulations. See Chapters 69 for more details.

8. Changes to East input

Many changes have been implemented in East to enhance the user experience in providing input for their designs. These changes include the ability to specify multiple values of input parameters for survival designs (most notably the Hazard Ratio), the ability to directly convert many fixed sample designs into group sequential designs with the use of the Sample Size based design option, and the ability to convert an ANOVA design into a Multiple Comparison to Control design.

9. Changes to East output

Display of East output has been changed in many ways, including color coding of input and output, ability to collapse and expand individual tables, greater decimal display control, and more exporting options for results (e.g. ability to export graphs directly into Microsoft Power Point).

What's New in East 6.2

Version 6.2 of East introduces some important new features:

1. Promising Zone Designs using CHW and CDL type-1 error control methods

Preface

East 6.2 introduces Promising Zone Designs from East 5.4 for differences of means, proportions, and the log-rank test. The methods of Cui, Hung, and Wang (1999) and Chen, DeMets, and Lan (2003) are implemented for adaptive unblinded sample size re-estimation designs and available for simulation and interim monitoring.

2. **Multiple endpoint designs** Serial gatekeeping procedures to control family-wise type-1 error when testing multiple families of normally-distributed endpoints are now available in East for fixed sample (1-look) designs.
3. **Power and sample size calculations for count data** East now offers power analysis and sample size calculations for count data in fixed sample (1-look) designs. Specifically, East provides design capabilities for:
 - (a) Test of a single Poisson rate
 - (b) Test for a ratio of Poisson rates
 - (c) Test for a ratio of Negative Binomial rates
4. **Precision-based sample size calculations** Sample size calculations are now available based on specification of a confidence interval for most tests provided in East.

What's New in East 6.1

Version 6.1 of East introduces some important new features:

1. **Bayesian probability of success (assurance) and predictive power**

For one-sample and two-sample continuous and binomial endpoints, East 6.1 will now compute Assurance (O'Hagan et al., 2005) or Bayesian probability of success, a Bayesian version of power, which integrates power over a prior distribution of the treatment effect, giving an unconditional probability that the trial will yield a significant result. When monitoring such a design using the Interim Monitoring dashboard, East 6.1 will also compute Bayesian predictive power using the pre-specified prior distribution on the treatment effect. This computation will be displayed in addition to the fiducial version of predictive power, which uses the estimated treatment effect and standard error to define a Gaussian prior distribution.
2. **Stratification in simulation of survival endpoints**

When simulating a trial design with a time-to-event endpoint, East 6.1 accommodates data generation in a stratified manner, accounting for up to 3 stratification variables and up to 25 individual strata. The fraction of subject data generated in each stratum, and the survival response generation mechanism for each stratum, can be flexibly adjusted. In addition, stratified versions of the logrank statistic and other test statistics available for analysis of the simulated data are provided.

3. **Integration of R code into simulations**

East 6.1 simulations now include the option to use custom R code to define specific elements of the simulation runs. R code can be used to modify the way the subjects are accrued, how they are randomized, how their response data are generated, and how the test statistic is computed.

4. **Reading East 5.4 workbooks**

East 5.4 workbooks can be read into East 6.1 after conversion using the utility provided in the program menu. Go to the start menu and select:

Programs > East Architect > File Conversion > East5 to East6

5. **Floating point display of sample size**

East 6.1 now has a setting to choose whether to round sample sizes (at interim and final looks) up to the nearest integer, or whether to display them as a floating point number, as in East 5. (See

6. **Enhancement to the Events vs. Time plot**

This useful graphic for survival designs has been updated to allow the user to edit study parameters and create a new plot directly from a previous one, providing the benefit of quickly assessing the overall impact of input values on a design prior to simulation. (See

7. **Interim monitoring (IM) dashboard**

The capability to save snapshots of the interim monitoring (IM) dashboard is now supported in East 6.1. At each interim look of a trial, updated information can be saved and previous looks can be easily revisited. Alternatively, prior to employing actual data this functionality could be used to compare multiple possible scenarios, providing the user a sense of how a future trial could unfold.

8. **Enhancement to the Logrank test**

For trials with survival endpoints, East 6.1 allows the user to simultaneously create multiple designs by specifying a range of values for key parameters in the Logrank test. (See Subsection

9. **Enhancement to binomial designs**

For studies with discrete outcomes, East 6.1 allows the user to simultaneously create multiple designs by specifying a range of values for key parameters.

What's New in East 6.0 on the Architect Platform

East Architect is version 6.0 of the East package and builds upon earlier versions of the

software. The transition of East to the next generation platform that is Architect has abandoned all prior dependencies of Microsoft Excel. As a result the user interface is very different leading to a new user experience and workflow. Although you might find that there is a learning curve to getting comfortable with the software, we trust that you will find that the new platform provides for a superior user experience and improved workflow.

Preface

The Architect platform also adds data management and analysis capabilities similar to those found in Cytel Studio, StatXact, and LogXact, as well as a powerful reporting tool we call Canvas, which provides flexible and customizable reports based on design and simulation information.

Version 6.0 of East introduces some important new features in addition to the new platform environment. Here is a selection:

1. **New designs** A large number of fixed sample designs have been added for various endpoints and trial types. These were present in the SiZ software and have now been fully integrated into East.
2. **Multi-arm designs** Designs for pairwise comparisons of treatment arms to control have been added for differences of means and differences of proportions. These designs are mostly simulation-based and provide operating characteristics for fixed sample studies using multiplicity adjusting procedures such as Dunnett's, Bonferroni, Sidak, Hochberg, Fallback, and others.
3. **Creation of multiple designs or simulations at once:** East Architect provides the ability to create multiple designs or to run multiple simulation scenarios at once, by specifying lists or sequences of values for specific parameters rather than single scalars. This capability allows the user to explore a greater space of possibilities or to easily perform sensitivity analysis. Accompanying tools to preview, sort, and filter are provided to easily parse the large output generated by East.
4. **Response lag, accrual, and dropouts for continuous and discrete endpoints:** Designs created for continuous and discrete endpoints now have the option for the user to specify a response lag (between randomization and observation of the endpoint), as well as an accrual rate and dropout rate for the study population. As a result, some terminology has been introduced to distinguish between the number of subjects who need to be enrolled in the study (Sample Size) and the number of subjects whose endpoint must be observed in order to properly power the study (Completers).
5. **Flexibility in setting up boundaries** Both the efficacy and futility rules of a design need not be present at each and every look anymore. The user can specify whether a look includes either the efficacy stopping rule or the futility rule or both. Therefore, a design can be set up where at the first look only futility stopping is possible, whereas at later looks both efficacy and futility or maybe only efficacy stopping is allowed. In addition, the futility rule can now be specified on two new scales, which are the standardized treatment scale and the conditional power scale.
6. **Predictive power** Predictive power is now provided as an alternative to conditional power in the interim monitoring sheet of the software. Further

details about how this is implemented can be found in the appendix C.

7. **Comparing designs** One can compare multiple designs either graphically or in tabular format simply by selecting them and choosing a plot or table output button.
8. **Improvements in algorithms** Many improvements have been made to the way computations are performed, both to improve accuracy and speed, but also to provide more intuitive results. For example, older versions of East used an approximation to conditional power based on ignoring all future looks but the final one. This approximation has been dropped in favor of computing the exact value of conditional power. Many other changes have been made that might result in different values being computed and displayed in East Architect as compared to earlier versions of the software. For greater details about the changes made, please refer to the "Read Me" notes that accompany the software release.

What's New in East 5

After East 5 (version 5.0) was released, a few upgrades have been issued. The details are:

1. In the current release of version 5.4, the module EastSurvAdapt has been added.
2. In the previous version 5.3, the module EastAdapt was substantially revised.
3. In the earlier version 5.2, the module EastExact was released.
4. In the still earlier version 5.1, several improvements were introduced in EastSurv module.

The details of these modules can be found in the respective chapters of the user manual.

East 5 upgraded the East system in several important ways in direct response to customer feedback. Six important extensions had been developed in East 5:

1. **Designs using t-tests:**

In previous versions of East, the single look design was treated as a special case of a group sequential design. Thus the same large sample theory was used to power and size these traditional types of designs. Recognizing this solution not to be entirely satisfactory for small sample trials, in East 5, we have implemented single-look t-test designs for continuous data. (Sections 8.1.4, 8.2.4, 9.1.3, and 11.1.3)

2. **New boundaries:**

East 5 provides two new procedures for specifying group sequential boundaries.

- Generalized Haybittle-Peto boundaries allow the user to specify unequal p-values at each interim look for a group sequential plan. East will

Preface

- recalculate the final p-value in order to preserve the type I error. (Section 38.1)
 - The cells for entering the cumulative alpha values of an interpolated spending function can be automatically populated with the cumulative alpha values of any of the published spending functions available to East, and subsequently edited to suit user requirements. For example, a 4-look Lan and DeMets O'Brien-Fleming spending function can be modified so that the critical value at the first look is less conservative than usual. (Section 38.3.1)
3. **Interim monitoring and simulation for single-look designs:**
Interim monitoring and simulation sheets have been provided for all single look designs in East 5.
 4. **Improvement to Charts:**
Many improvements to existing charts in East have been implemented in this version.
 - Scaling in the Duration vs. Accrual chart has been corrected to provide a better tool for the user.
 - The use of semi-log scaling has enabled us to represent many charts on the natural scale of the treatment effect. This concerns mostly any ratio and odds ratio metrics such as the relative risk, the hazard ratio, and the odds ratio. Boundaries on the relative risk scale for example are now available in East 5.
 - Boundaries can also be visualized on the score scale.
 - Charts can be summarized in tabular form. Option is given to the user to generate tables of power vs. sample size, power vs. treatment effect, events vs. time, and so on. These tables can easily be copied and pasted into external applications like Microsoft Word and Excel (Section 4.5)
 5. **Improved usability:**
Much attention in East 5 was spent to improve the user's experience within the environment.
 - A graph sheet allows the user to compare up to 16 charts side by side. Charts for any number of plans within a workbook can be exported to the graph sheet. (Section 5.3)
 - The scratch sheet is a full-fledged Microsoft Excel sheet that can be brought up within the East application . (Section 4.4)
 - The split view option enables the user to see two sheets of the same workbook simultaneously. This can be useful if one window pane contains a scratch sheet where side calculations may be done based on numbers in

the other window pane. Another use can be to have two or plans to show up on one pane and their graphsheets containing boundaries or other charts to show up on another pane for easy comparison. (Section 4.8)

- Messages in the help menu, pop-up help, and context sensitive help have been revised and rendered more informative to the user.
- The default appearance of charts can be specified by the user through the preferences settings menu item. (Section 4.7)

6. Installation validation:

East 5 includes an installation validation procedure that will easily check that the software has been properly installed on the user's system. (Section 2.3)

Finally, there has been an important reorganization of the East manual, which now comprises seven volumes organized as follows: (1) The East System (2) Continuous Endpoints (3) Binomial and Categorical Endpoints (4) Time-to-Event Endpoints (5) Adaptive Designs (6) Special Topics (7) Appendices. Page numbers are continuous through volumes 1-7. Each volume contains a full table of contents and index to the whole manual set.

Preface to East 4

East 4 was a very large undertaking involving over 20 developers, documenters, testers and helpers over a two-year period. Our goal was to produce one single powerful design and monitoring tool with a simple, intuitive, point and click, menu driven user interface, that could cover the full range of designs commonly encountered in a clinical trial setting, for either fixed sample or group sequential designs. The resulting product, East 4, extends the East system for flexible design and interim monitoring in four major ways as listed below.

1. Broad Coverage:

Previous versions of East dealt primarily with the design of two-arm group sequential trials to detect a difference of means for normal and binomial endpoints and a hazard ratio for survival endpoints. East 4 extends these capabilities to other settings.

- Easily design and monitor up to 34 different clinical trial settings including one-, two- and K-sample tests; linear, logistic and Cox regression; longitudinal designs; non-inferiority and bioequivalence designs; cross-over and matched-pair designs; nonparametric tests for continuous and ordered categorical outcomes.
- Comparisons between treatment and control groups can be in terms of differences, ratios or odds ratios.

Preface

- Non-inferiority trials can be designed to achieve the desired power at superiority alternatives
2. **New Stopping Boundaries and Confidence Intervals:**
 - Non-binding futility boundaries. Previously futility boundaries could not be overruled without inflating the type-1 error. New non-binding futility boundaries preserve power and type-1 error and yet can be overruled if desired.
 - Asymmetric two-sided efficacy boundaries. You can allocate the type-1 error asymmetrically between the upper and lower stopping boundaries, and can spend it at different rates with different error spending functions. This will provide added flexibility for aggressive early stopping if the treatment is harmful and conservative early stopping if the treatment is beneficial.
 - Futility boundaries can be represented in terms of conditional power. This brings greater objectivity to conditional power criteria for early stopping.
 - Two sided repeated confidence intervals are now available for one-sided tests with efficacy and futility boundaries. Previously only one-sided confidence bounds were available.
 - Interactive repeated confidence intervals are provided at the design stage to aid in sample size determination and selection of stopping boundaries.
 3. **New Analytical and Simulation Tools for Survival Studies:**

EastSurv is an optional new module, fully integrated into the East system, that extends East's design capabilities to survival studies with non-uniform accrual, piecewise exponential distributions, drop outs, and fixed length of follow-up for each subject. Designs can be simulated under general settings including non-proportional hazard alternatives.
 4. **Design and Simulation of Adaptive Trials:**

EastAdapt is an optional new module, fully integrated into the East system, that permits data-dependent changes to sample size, spending functions, number and spacing of interim looks, study objectives, and endpoints using a variety of published flexible approaches.

In addition to these substantial statistical capabilities, East 4 has added numerous improvements to the user interface including clearer labeling of tables and graphs, context sensitive help, charts of power versus sample size and power versus number of events, convenient tools for calculating the test statistics to be entered into the interim monitoring worksheet for binomial endpoints, and the ability to type arithmetic expressions into dialog boxes and into design, interim monitoring and simulation worksheets.

Preface to East 3

East 3 is a major upgrade of the East-2000 software package for design and interim monitoring of group sequential clinical trials. It has evolved over a three-year period with regular input from our East-2000 customers. The main improvements that East 3 offers relative to East-2000 are greater flexibility in study design, better tracking of interim results, and more powerful simulation capabilities. Many of our East-2000 customers expressed the desire to create group sequential designs that are ultra-conservative in terms of stopping early for efficacy, but which can be quickly terminated for futility. The extremely wide selection of spending functions and stopping boundaries in East 3, combined with its interactive Excel-based spreadsheet user interface for comparing multiple designs quickly and effortlessly, make such designs possible. The interim monitoring module of East 3 has been completely revised, with a “dashboard” user interface that can track the test statistic, error spent, conditional power, post-hoc power and repeated confidence intervals on a single worksheet, over successive interim monitoring time points, for superior trial management and decision making by a data monitoring committee. Finally, we have enhanced the simulation capabilities of East 3 so that it is now possible to evaluate the operating characteristics not only of traditional group sequential designs, but also of adaptive designs that permit mid-course alterations in the sample size based on interim estimates of variance or treatment effect. A list of the substantial new features in East 3 relative to East-2000 is given below. (The items on this list beginning with ‘(*)’ are optional extras.)

New Design Features

1. Design of non-inferiority trials.
2. Design of trials with unequally spaced looks.
3. Use of Lan and DeMets (1983) error spending functions to derive stopping boundaries.
4. (*) Flexible stopping boundaries derived from the gamma spending function family (Hwang, Shih and DeCani, 1990) and the rho spending function family (Kim and DeMets, 1987).
5. Haybittle-Peto stopping boundaries (Haybittle, 1971).
6. (*) Boundaries derived from user-specified spending functions with interpolation.
7. Boundaries for early stopping for futility only.
8. Graphical and numerical representation of stopping boundaries on other scales besides the standard normal scale; e.g., boundaries expressed on the p-value scale, effect size scale, and conditional power scale.
9. Computing power for a fixed sample size.
10. Chart displaying the number of events as a function of time (for survival studies).

Preface

New Interim Monitoring Features

1. Detailed worksheet for keeping track of interim monitoring data and providing input to the data monitoring committee.
2. Simultaneous view of up to four thumbnail charts on the interim monitoring worksheet. Currently one may select any four charts from, the stopping boundary chart, the error spending chart, the conditional power chart, the post-hoc power chart, and the repeated confidence intervals chart. You can also expand each thumbnail into a full-sized chart by a mouse click.
3. Computation of repeated confidence interval (Jennison and Turnbull, 2000) at each interim look.

New Simulation Features

1. (*) Simulation of actual data generated from the underlying normal or binomial model instead of simulating the large sample distribution of the test statistic.
2. (*) Simulation on either the maximum sample size scale, or the maximum information scale.
3. (*) Simulation of the adaptive design due to Cui, Hung and Wang (1999).

New User Interface Features

1. Full integration into the Microsoft Excel spreadsheet for easy generation and display of multiple designs, interim monitoring or simulation worksheets, and production of reports.
2. Save design details and interim monitoring results in Excel worksheets for easy electronic transmission to regulatory reviewers or to end-users.
3. Create custom calculators in Excel and save them with the East study workbook.

Preface to East-2000

For completeness we repeat below the preface that we wrote for the East-2000 software when it was released in April, 2000.

Background to the East-2000 Development The precursor to East-2000 was East-DOS an MS-DOS program with design and interim monitoring capabilities for normal, binomial and survival end points. When East-DOS was released in 1991 its user interface and statistical features were adequate to the needs of its customer base. MS-DOS was still the industry standard operating system for desktop computers. Group sequential designs were not as popular then as they are now. The role of data and safety monitoring boards (DSMB's) in interim monitoring was just beginning to emerge. FDA and industry guidelines on the conduct of group sequential studies were in the early draft stage. Today the situation is very different. Since the publication of

the ICH-E9 guidance on clinical trials by the FDA and regulatory bodies in Europe and Japan, industry sponsors of phase-III clinical trials are more favorably inclined to the group sequential approach. For long-term mortality studies especially, interim monitoring by an independent DSMB is almost mandatory. As the popularity of group sequential studies has increased so has the demand for good software to design and monitor such studies. For several years now we have been flooded with requests from our old East-DOS customers to move away from the obsolete MS-DOS platform to Microsoft Windows and to expand the statistical capabilities of the software. We have responded by developing East-2000, a completely re-designed Windows package with unparalleled design, simulation and interim monitoring capabilities.

What's New in East-2000 The East-2000 software adds considerable functionality to its MS-DOS predecessor through a superior user interface and through the addition of new statistical methods.

New User Interface East-2000 is developed on the Microsoft Windows platform. It supports a highly interactive user interface with ready access to stopping boundary charts, error spending function charts, power charts and the ability to present the results as reports in Microsoft Office.

1. *Interactivity* Designing a group sequential study is much more complex than designing a fixed sample study. The patient resources needed in a group sequential setting depend not only on the desired power and significance level, but also on how you will monitor the data.

How many interim looks are you planning to take? What stopping boundary will you use at each interim look? Does the stopping boundary conform to how you'd like to spend the type-1 error at each look? Do you intend to stop early only for benefit, only for futility, or for both futility and benefit? In a survival study, how long are you prepared to follow the patients?

These design and monitoring decisions have profound implications for the maximum sample size you must commit up-front to the study, the expected sample size under the null and alternative hypotheses, and the penalty you will have to pay in terms of the nominal p-value needed for declaring significance at the final look. To take full advantage of the group sequential methodology and consider the implications of potential decisions you must have highly interactive software available, both at the study design stage and at the interim monitoring stage. East-2000 is expressly developed with this interactivity in mind. Its intuitive form-fill-in graphical user interface can be an invaluable tool for visualizing how these design and monitoring decisions will affect the operating characteristics of the study.

Preface

2. *Charts* By clicking the appropriate icon on the East toolbar you can view stopping boundary charts, study duration charts, error spending function charts, conditional and post-hoc power charts, and exit probability tables. The ease with which these charts can be turned on and off ensures that they will be well utilized both at the design and interim monitoring phases of the study.
3. *Reports* All worksheets, tables and charts produced by East-2000 can be copied and pasted into Microsoft Word, Excel and PowerPoint pages thus facilitating the creation of annotated reports describing the study design and interim monitoring schedule.

New Statistical Methods East-2000 has greatly expanded the design and interim monitoring capabilities previously available in East-DOS. In addition East-2000 provides a simulation module for investigating how the power of a sequential design is affected by different assumptions about the magnitude of the treatment difference. Some highlights from these new capabilities are listed below.

1. *Design* Whereas East-DOS only provided design capabilities for normal, binomial and survival end points East-2000 makes it possible to design more general studies as well. This is achieved through the use of an inflation factor. The inflation factor determines the amount by which the sample size of a fixed sample study should be inflated so as to preserve its type-1 error in the presence of repeated hypothesis tests. It is thus possible to use any external software package to determine the fixed sample size of the study, input this fixed sample size into the design module of East-2000 and have the sample size inflated appropriately. These general capabilities are discussed in Chapter 8.
2. *Interim Monitoring* A major new feature in the interim monitoring module of East-2000 is the computation of adjusted p-values, confidence intervals and unbiased parameter estimates at the end of the sequential study. Another important feature is the ability to monitor the study on the Fisher information scale and thereby perform sample-size re-estimation if initial assumptions about the data generating process were incorrect. Chapter 9 provides an example of sample-size re-estimation for a binomial study in which the initial estimate of the response rate of the control drug was incorrect.
3. *Simulation* East-2000 can simulate an on-going clinical trial and keep track of the frequency with which a stopping boundary is crossed at each interim monitoring time-point. These simulations can be performed under the null hypothesis, the alternative hypothesis or any intermediate hypothesis thus permitting us to evaluate how the various early stopping probabilities are affected by miss-specifications in the magnitude of the treatment effect.

Continuous Development of East East-2000 will undergo continuous development with major new releases expected on an annual basis and smaller improvements regularly posted on the Cytel web site. We will augment the software and implement new techniques based on the recommendations of the East Advisory Committee, and as the demand for them is expressed by our customers. The following items are already on the list:

- Easy links to fixed-sample design packages so as to extend the general methods in Chapter 8;
- Analytical and simulation tools to convert Fisher information into sample size and thereby facilitate the information based design and interim monitoring methods of Chapter 9, especially for sample-size re-estimation.

We will build a forum for discussing East related issues on the Cytel web site, www.cytel.com. Interesting case studies, frequently asked questions, product news and other related matters will be posted regularly on this site.

Roster of East Consultants Cytel offers consulting services to customers requiring assistance with study design, interim monitoring or representation on independent data and safety monitoring boards. Call us at 617-661-2011, or email sales@cytel.com, for further information on this service.

Table of Contents

| | |
|---|-----------|
| Preface | ii |
| 1 The East System | 1 |
| 1 Introduction to Volume 1 | 2 |
| 2 Installing East 6 | 3 |
| 3 Getting Started | 7 |
| 4 Data Editor | 55 |
| 2 Continuous Endpoints | 71 |
| 5 Introduction to Volume 2 | 73 |
| 6 Tutorial: Normal Endpoint | 79 |
| 7 Normal Superiority One-Sample | 91 |
| 8 Normal Noninferiority Paired-Sample | 113 |
| 9 Normal Equivalence Paired-Sample | 128 |
| 10 Normal Superiority Two-Sample | 141 |
| 11 Nonparametric Superiority Two Sample | 179 |
| 12 Normal Non-inferiority Two-Sample | 185 |
| 13 Normal Equivalence Two-Sample | 211 |

Table of Contents

| | | |
|----|---|-----|
| 14 | Normal: Many Means | 232 |
| 15 | Multiple Comparison Procedures for Continuous Data | 240 |
| 16 | Multiple Endpoints-Gatekeeping Procedures | 265 |
| 17 | Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs | 285 |
| 18 | Two-Stage Multi-arm Designs using p-value combination | 309 |
| 19 | Normal Superiority Regression | 332 |
| 3 | Binomial and Categorical Endpoints | 342 |
| 20 | Introduction to Volume 3 | 344 |
| 21 | Tutorial: Binomial Endpoint | 350 |
| 22 | Binomial Superiority One-Sample | 363 |
| 23 | Binomial Superiority Two-Sample | 394 |
| 24 | Binomial Non-Inferiority Two-Sample | 474 |
| 25 | Binomial Equivalence Two-Sample | 535 |
| 26 | Binomial Superiority n-Sample | 549 |
| 27 | Multiple Comparison Procedures for Discrete Data | 577 |
| 28 | Multiple Endpoints-Gatekeeping Procedures for Discrete Data | 601 |
| 29 | Two-Stage Multi-arm Designs using p-value combination | 621 |
| 30 | Binomial Superiority Regression | 644 |
| 31 | Agreement | 649 |
| 32 | Dose Escalation | 658 |

| | | |
|-----------|---|------------|
| 4 | Exact Binomial Designs | 708 |
| 33 | Introduction to Volume 8 | 709 |
| 34 | Binomial Superiority One-Sample – Exact | 714 |
| 35 | Binomial Superiority Two-Sample – Exact | 736 |
| 36 | Binomial Non-Inferiority Two-Sample – Exact | 751 |
| 37 | Binomial Equivalence Two-Sample – Exact | 767 |
| 38 | Binomial Simon’s Two-Stage Design | 774 |
| | | |
| 5 | Poisson and Negative Binomial Endpoints | 784 |
| 39 | Introduction to Volume 4 | 785 |
| 40 | Count Data One-Sample | 790 |
| 41 | Count Data Two-Samples | 799 |
| | | |
| 6 | Time to Event Endpoints | 819 |
| 42 | Introduction to Volume 6 | 820 |
| 43 | Tutorial: Survival Endpoint | 826 |
| 44 | Superiority Trials with Variable Follow-Up | 865 |
| 45 | Superiority Trials with Fixed Follow-Up | 908 |
| 46 | Non-Inferiority Trials Given Accrual Duration and Accrual Rates | 934 |
| 47 | Non-Inferiority Trials with Fixed Follow-Up | 950 |
| 48 | Superiority Trials Given Accrual Duration and Study Duration | 966 |
| 49 | Non Inferiority Trials Given Accrual Duration and Study Duration | 984 |

Table of Contents

| | | |
|----|--|------|
| 50 | A Note on Specifying Dropout parameters in Survival Studies | 994 |
| 51 | Multiple Comparison Procedures for Survival Data | 999 |
| 7 | Adaptive Designs | 1019 |
| 52 | Introduction To Adaptive Features | 1020 |
| 53 | The Motivation for Adaptive Sample Size Changes | 1027 |
| 54 | The Cui, Hung and Wang Method | 1055 |
| 55 | The Chen, DeMets and Lan Method | 1160 |
| 56 | Muller and Schafer Method | 1221 |
| 57 | Conditional Power for Decision Making | 1350 |
| 8 | Special Topics | 1387 |
| 58 | Introduction to Volume 8 | 1388 |
| 59 | Design and Monitoring of Maximum Information Studies | 1393 |
| 60 | Design and Interim Monitoring with General Endpoints | 1423 |
| 61 | Early Stopping for Futility | 1434 |
| 62 | Flexible Stopping Boundaries in East | 1460 |
| 63 | Confidence Interval Based Design | 1493 |
| 64 | Simulation in East | 1552 |
| 65 | Predictive Interval Plots | 1575 |
| 66 | Enrollment/Events Prediction - At Design Stage (By Simulation) | 1609 |
| 67 | Conditional Simulation | 1658 |

| | |
|---|-------------|
| 68 Enrollment/Events Prediction - Analysis | 1675 |
| 69 Interfacing with East PROCs | 1787 |
| 9 Analysis | 1795 |
| 70 Introduction to Volume 9 | 1798 |
| 71 Tutorial: Analysis | 1806 |
| 72 Analysis-Descriptive Statistics | 1827 |
| 73 Analysis-Analytics | 1837 |
| 74 Analysis-Plots | 1854 |
| 75 Analysis-Normal Superiority One-Sample | 1890 |
| 76 Analysis-Normal Noninferiority Paired-Sample | 1901 |
| 77 Analysis-Normal Equivalence Paired-Sample | 1907 |
| 78 Analysis-Normal Superiority Two-Sample | 1913 |
| 79 Analysis-Normal Noninferiority Two-Sample | 1926 |
| 80 Analysis-Normal Equivalence Two-Sample | 1941 |
| 81 Analysis-Nonparametric Two-Sample | 1956 |
| 82 Analysis-ANOVA | 1976 |
| 83 Analysis-Regression Procedures | 1987 |
| 84 Analysis-Multiple Comparison Procedures for Continuous Data | 2024 |
| 85 Analysis-Multiple Endpoints for Continuous Data | 2055 |
| 86 Analysis-Binomial Superiority One-Sample | 2060 |

Table of Contents

| | | |
|-----------|---|-------------|
| 87 | Analysis-Binomial Superiority Two-Sample | 2069 |
| 88 | Analysis-Binomial Noninferiority Two-Sample | 2088 |
| 89 | Analysis-Binomial Equivalence Two-Samples | 2106 |
| 90 | Analysis-Discrete: Many Proportions | 2111 |
| 91 | Analysis-Binary Regression Analysis | 2131 |
| 92 | Analysis- Multiple Comparison Procedures for Binary Data | 2180 |
| 93 | Analysis-Comparison of Multiple Comparison Procedures for Continuous Data- Analysis | 2207 |
| 94 | Analysis-Multiple Endpoints for Binary Data | 2211 |
| 95 | Analysis-Agreement | 2216 |
| 96 | Analysis-Survival Data | 2219 |
| 97 | Analysis-Multiple Comparison Procedures for Survival Data | 2240 |
| 10 | Appendices | 2267 |
| A | Introduction to Volume 10 | 2269 |
| B | Group Sequential Design in East 6 | 2271 |
| C | Interim Monitoring in East 6 | 2313 |
| D | Computing the Expected Number of Events | 2334 |
| E | Generating Survival Simulations in EastSurv | 2345 |
| F | Spending Functions Derived from Power Boundaries | 2347 |
| G | The Recursive Integration Algorithm | 2352 |
| H | Theory - Multiple Comparison Procedures | 2353 |

| | | |
|---|--|------|
| I | Theory - Multiple Endpoint Procedures | 2368 |
| J | Theory-Multi-arm Multi-stage Group Sequential Design | 2374 |
| K | Theory - MultiArm Two Stage Designs Combining p-values | 2394 |
| L | Technical Details - Predicted Interval Plots | 2404 |
| M | Enrollment/Events Prediction - Theory | 2409 |
| N | Dose Escalation - Theory | 2412 |
| O | R Functions | 2427 |
| P | East 5.x to East 6.4 Import Utility | 2478 |
| Q | Technical Reference and Formulas: Single Look Designs | 2484 |
| R | Technical Reference and Formulas: Analysis | 2542 |
| S | Theory - Design - Binomial One-Sample Exact Test | 2605 |
| T | Theory - Design - Binomial Paired-Sample Exact Test | 2611 |
| U | Theory - Design - Simon's Two-Stage Design | 2614 |
| V | Theory-Design - Binomial Two-Sample Exact Tests | 2617 |
| W | Classification Table | 2638 |
| X | Glossary | 2639 |
| Y | On validating the East Software | 2657 |
| Z | List of East Beta Testers | 2686 |
| | References | 2695 |
| | Index | 2719 |

Volume 1 *The East System*

- 1 Introduction to Volume 1* [2](#)
- 2 Installing East 6* [3](#)
- 3 Getting Started* [7](#)
- 4 Data Editor* [55](#)

1 Introduction to Volume 1

This volume contains chapters which introduce you to East software system.

Chapter 2 explains the hardware and operating system requirements and the installation procedures. It also explains the installation validation procedure.

Chapter 3 is a tutorial for introducing you to East software quickly. You will learn the basic steps involved in getting in and out of the software, selecting various test options under any of the endpoints, designing a study, creating and comparing multiple designs, simulating and monitoring a study, invoking the graphics, saving your work in files, retrieving previously saved studies, obtaining on-line help and printing reports. It basically describes the menu structure and the menus available in East software, which is a menu driven system. Almost all features are accessed by making selections from the menus.

Chapter 4 discusses the Data Editor menu of East 6 which allows you to create and manipulate the contents of your Case Data and Crossover Data. This menu is in use while working with the **Analysis** menu as well as with some other features like PIP or Conditional Simulations.

These features are illustrated with the help of a simple worked example of a binary endpoint trial.

2 *Installing East 6*

2.1 System Requirements to run East 6

The minimum hardware/operating system/software requirements for East 6 (standalone version of the software or the East client in case of concurrent version) are listed below:

- In case of Standalone version and East clients in case of concurrent version, the following operating systems are supported:
 - Windows 7 (32-bit / 64 bit)
 - Windows 8 (32-bit / 64 bit)
 - Windows 8.1 (32-bit / 64-bit)
 - Windows 10 (32-bit / 64-bit)
 - All of above for computers with English, European and Japanese versions of Windows.
- In case of concurrent user version, the following server operating systems are supported:
 - Windows 7 (32-bit / 64 bit)
 - Windows 8 (32-bit / 64 bit)
 - Windows 8.1 (32-bit / 64-bit)
 - Windows 10 (32-bit / 64-bit)
 - All of above for computers with English, European and Japanese versions of Windows
 - Windows Server 2008 (32-bit / 64-bit)
 - Windows Server 2012
 - Citrix
 - * XenApp 6.0 on Windows 2008
 - * XenApp 6.5 on Windows 2008
 - * XenApp 7.6 on Windows 2008
 - * XenApp 7.6 on Windows 2012
- Further, East has the following hardware/software requirements:
 - CPU -1 GHz or faster x86 (32 bit) or x64 (64 bit) processor
 - Memory - Minimum 1 GB of RAM
 - Hard Drive - Minimum 5 GB of free hard disk space
 - Display - 1024 x 768 or higher resolution

2 Installing East 6

- Microsoft .Net Framework 4.0 Full (this will be installed as a part of prerequisites if your computer does not have it)
- Microsoft Visual C++ 2010 SP1 (this will be installed as a part of prerequisites if your computer does not have it) Installer 4.5
- Internet Explorer 9.0 or above
- A stable internet connection is required during installation so that prerequisites like the
- East is compatible and supported with R versions between 2.9.0 to 3.2.3. East may or may not work well with later versions of R. If R is not installed, the ability to include custom R functions to modify specific simulation steps will not be available. The R integration feature is an Add-on to East and is required only to integrate custom R functions with East. But note that this feature doesn't affect any of the core functionalities of East.

2.2 Other Requirements

Users with Windows 7 or above: East uses the font Verdana. Generally Verdana is a part of the default fonts installed by Windows. However, sometimes this font may not be available on some computers, especially if a language other than English has been selected. In such cases, the default fonts need to be restored. To restore fonts, go to Control Panel → Fonts → Font settings. Click the button “Restore default font settings”. This will restore all default fonts including Verdana. Note that this must be done before the first use of East.

Users with Windows 8.1 On some computers with Windows 8.1, problems may be observed while uninstalling East, especially if the user has upgraded from the previous version using a patch. This is because of a security update KB2962872 (MS14-037) released by Microsoft for Internet Explorer versions 6, 7, 8, 9, 10 and 11. Microsoft has fixed this issue and released another security update [KB2976627 \(MS14-051\)](#) for Internet Explorer which replaces the old problematic update. So it is recommended that users who are affected by this issue install security update [KB2976627 \(MS14-051\)](#) on their computers.

2.3 Installation

IMPORTANT: Please follow the steps below if you are installing a standalone/single user version of East. If you are installing a concurrent version, please refer to the document “Cytel License Manager Setup.pdf” for detailed installation instructions.

1. **Uninstalling Previous Versions** If any version (including a beta or demo) of East 6 is currently installed on your PC, please uninstall it completely or else the

installation of the current version will not proceed correctly. To uninstall the earlier version of East 6, go to the **Start** Menu and select:

All Programs → **Cytel Architect** → **East 6.x** → **Uninstall**
Or

All Programs → **East Architect** → **Uninstall East Architect**

depending upon the version installed on your computer.

2. **Installing Current Version** You will need to be an administrator of your computer in order to perform the following steps. If you do not have administrator privileges on your computer, please contact your system administrator / IT.

In order to install East, please follow these steps:

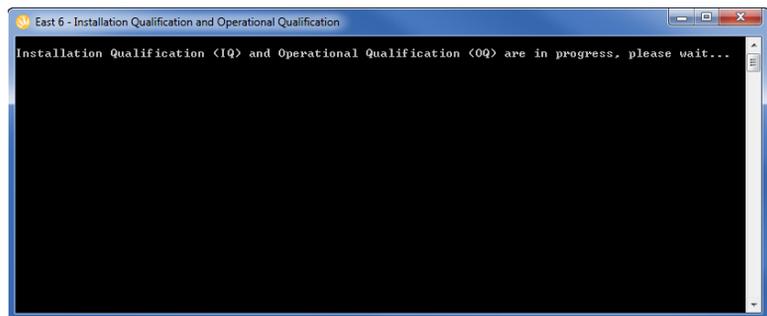
- (a) If you received an email containing a link for downloading the setup, please follow the link and download the setup. This will be a zipped folder. Unzip this folder completely.
- (b) In the setup folder, locate the program **setup.exe** and double-click on it. Follow the instructions on the subsequent windows.

2.4 Installation Qualification and Operational Qualification

To perform the installation and operational qualification of East 6, go to the **Start** Menu and select

All Programs → **Cytel Architect** → **East 6.4** → **Installation Qualification (IQ)**.

You will be presented with the following dialog box.



It will take a few minutes to complete. At the end of the process, the status of the installation qualification will appear. Press Enter (or any other key) to open the

2 Installing East 6

validation log. Similarly, one can run the Operational Qualification (OQ). If the validation is successful, the log file will contain a detailed list of all files installed by East on your computer and other details related to IQ and OQ.

If the validation fails, the validation log file will contain detailed error messages. Please contact your system administrator with the log file.

- **IQ (Installation Qualification) script:** This script verifies whether the software is completely and correctly installed on the system or not. It does this by checking whether all the software components, XML and DLL files are in place.
- **OQ (Operational Qualification) script:** This script runs some representative test cases covering all the major modules/features of East and compares the runtime results to the benchmarks (benchmarks are validated results stored internally in the OQ program). It ensures the quality and consistency of the results in the new version.
- **Manual Examples:** In addition to IQ/OQ, if more testing is to be done, refer to the user manual and reproduce the results for some representative examples/modules. The flow of examples is easy to follow. Some examples in the manual require additional files (datasets) which are available to you in the Samples folder.
- **Validation Chapter:** There is a chapter in this manual dedicated to describe how every feature was validated within Cytel. Refer to the appendix chapter [Y](#) on "Validating East Software". This covers validation strategies for all the features available in East 6.

3

Getting Started

East has evolved over the past several years with MS Excel[®] as the user interface. The East on MS Excel[®] did not integrate directly with any other **Cytel** products. Under the **Architect** platform, East is expected to coexist and integrate seamlessly with other **Cytel** products such as SiZ, and Compass. Architect is a common platform designed to support various Cytel products. It provides a user-friendly, Windows-standard graphical environment, consisting of tabs, icons, and dialog boxes, with which you can design, simulate and analyze. Throughout the user manual, this product is referred to as East 6.

One major advantage of East 6 is the facility for creating multiple designs. This is achieved by giving multiple inputs of the parameters as either comma separated, or in a range such as **(a:b:c)** with **a** as the initial value, **b** as the last value and **c** as the step size. If you give multiple values for more than one parameter, East creates all possible combinations of the input parameters. This is an immense advancement over earlier versions of East, where you had to create one design at a time. Furthermore, one could not compare different types of designs (e.g., superiority vs. noninferiority designs). Similarly, graphical comparison of designs with different numbers of looks was difficult with earlier versions of East. All such comparisons are readily available in East 6.

Another new feature is the option to add assumptions for accruals and dropouts at the design stage. Previously, this was available only for survival endpoint trials, but has been extended to continuous and discrete endpoints in East 6. Information about accrual rates, response lag, and dropouts can be given whether designing or simulating a trial. This makes more realistic, end-to-end design and simulation of a trial possible. Section 3.6 discusses all the above features under the **Design** menu with the help of a case study, CAPTURE.

Simulations help to develop better insight into the operating characteristic of a design. In East 6, the simulation module has now been enhanced to allow fixed or random allocation to treatment and control, and different sample sizes. Such options were not possible with earlier versions of East. Section 3.7 briefly describes the **Simulations** in East 6.

Section 3.8 discusses capability to flexibly monitoring a group sequential trial using the **Interim Monitoring** feature of East 6.

We have also provided powerful data editors to create, view, and modify data. A wide variety of statistical tests are now a part of East 6, which enables you to conduct

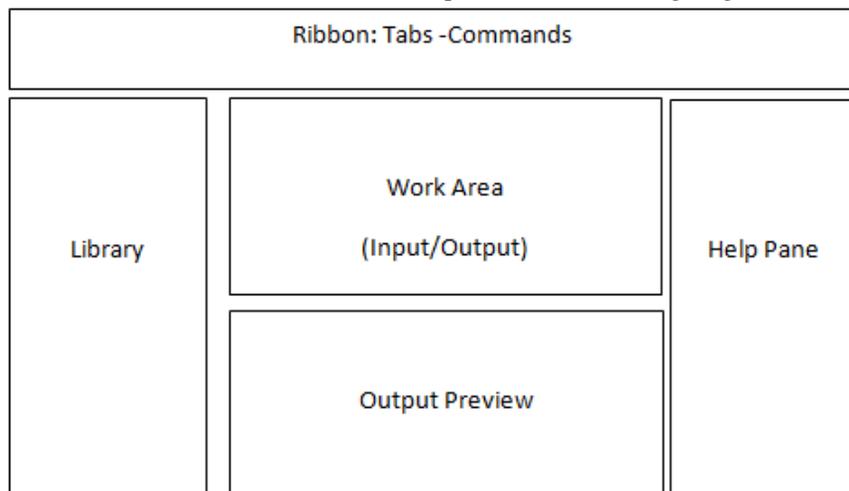
3 Getting Started

statistical analysis of interim data for continuous, discrete and time to event endpoints. Sections 3.4 and 3.5 briefly describes the **Data Editor** and **Analysis** menus in East 6. The purpose of this chapter is to familiarize you with the East 6 user interface.

3.1 Workflow in East

In this section, the architecture of East 6 is explained. The logical workflow in which the different parts of the user interface co-ordinate with each other is discussed.

The basic structure of the user interface is depicted in the following diagram.



Besides the top **Ribbon**, there are four main windows in East 6 namely, (starting from left), the **Library** pane, the **Input / Output** window, the **Output Preview** window and the **Help** pane. Note that both, the **Library** and the **Help Pane** can be auto-hidden temporarily or throughout the session, allowing the other windows to occupy larger area on the screen for display.

Initially, **Library** shows only the **Root** node. As you work with East, several nodes corresponding to designs, simulation scenarios, data sets and related analyses can be managed using this panel. Various nodes for outputs and plots are created in the **Library**, facilitating work on multiple scenarios at a time. The width of the **Library** window can be adjusted for better readability.

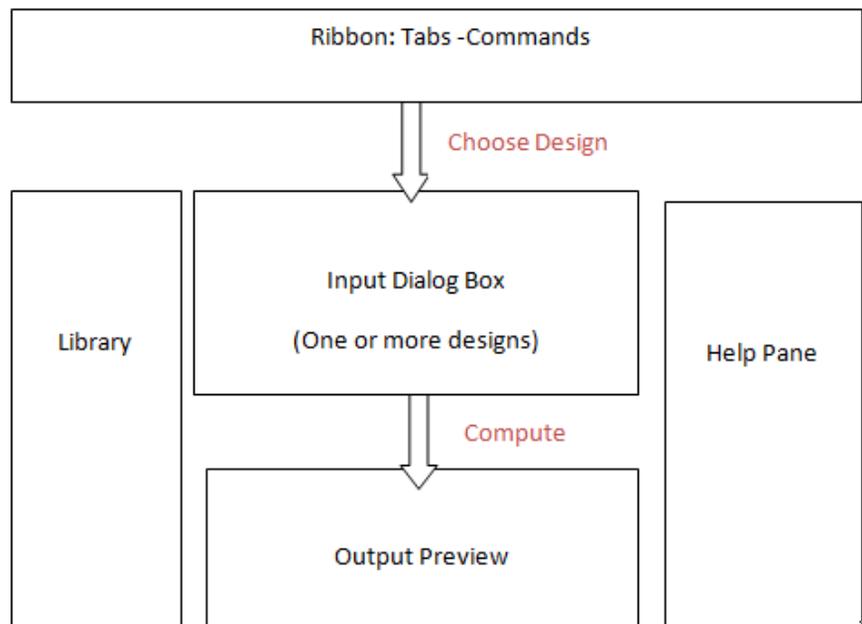
The central part of the user interface, the **Input / Output** window, is the main work area where you can-

- Enter input parameters for design computation create and compare multiple designs, view plots
- Simulate a design under different scenarios
- Perform interim analysis on a group sequential design look by look and view the results, receive decisions such as stopping or continuing during the execution of a trial
- Open a data on which you want to perform analysis, enter new data, view outputs, prepare a report etc.

This is the area where the user interacts with the product most frequently.

The **Output Preview** window compiles several outputs together in a grid like structure where each row is either a design or simulation run. This area is in use only when working with Design or Simulations.

When the **Compute** or **Simulate** button is clicked, all requested design or simulation results are computed and are listed row wise in the **Output Preview** window:



By clicking different rows of interest while simultaneously holding the **Ctrl** key, either a single or multiple designs can be displayed in the **Output Summary** in vertical

3 Getting Started

manner or side-by-side comparison can be done.

The screenshot shows two windows from a software application. The top window is titled 'Output Summary' and displays a detailed comparison of three designs: Des1, Des2, and Des3. The bottom window is titled 'Output Preview' and shows a table summarizing the parameters for each design.

| | | | | Output Summary | | |
|---|--|--|--|------------------|------------------|------------------|
| | | | | Des1 | Des2 | Des3 |
| | | | | MN-25-D1 | MN-25-D1 | MN-25-D1 |
| Test Parameters | | | | | | |
| Design Type | | | | Superiority | Superiority | Superiority |
| No. of Looks | | | | 1 | 3 | 3 |
| Test Type | | | | 1-Sided | 1-Sided | 1-Sided |
| Specified α | | | | 0.025 | 0.025 | 0.025 |
| Power | | | | 0.9 | 0.9 | 0.9 |
| Model Parameters | | | | | | |
| Allocation Ratio (nt/nc) | | | | 1 | 1 | 1 |
| Input Method | | | | Individual Means | Individual Means | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | | | | 0.3 | 0.3 | 0.3 |
| Mean Control (μ_c) | | | | 0 | 0 | 0 |
| Mean Treatment (μ_t) | | | | 0.3 | 0.3 | 0.3 |
| Std. Deviation (σ) | | | | 1 | 1 | 1 |
| Test Statistic | | | | Z | Z | Z |
| Boundary Parameters | | | | | | |
| Spacing of Looks | | | | | Equal | Equal |
| Efficacy Boundary | | | | | LD (OF) | LD (OF) |
| Accrual & Dropout Parameters | | | | | | |
| Accrual Rate | | | | | | 8 |
| Response Lag | | | | | | 2 |
| Probability of Dropout | | | | | | 0.1 |
| Sample Size | | | | | | |
| Maximum | | | | 467 | 473 | 526 |
| Expected Under H0 | | | | | 472.032 | 525.024 |
| Expected Under H1 | | | | | 379.185 | 431.008 |

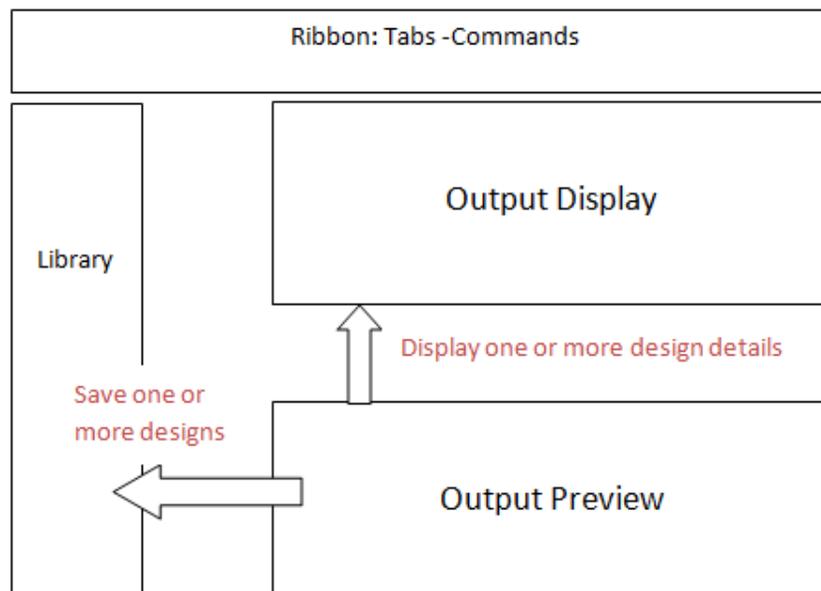
| | | | | | | | | | | | | | | Output Preview | | | | |
|------|-------------|--------------|-----------|--------------------|-------|-------|-------------|----|------------------------|----------|----------|------------------|---------|-----------------------|----------|----------------|------------------|-------------------|
| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | nc | Prop. Treatment (Alt.) | δ | Variance | Input Method | μ_c | Mean Treatment (Alt.) | σ | Test Statistic | Spacing of Looks | Efficacy Boundary |
| Des1 | Superiority | 1 | 1-Sided | 0.025 | 0.9 | 1 | 467 | | 0.3 | | | Individual Means | 0 | 0.3 | 1 | Z | Equal | LD (OF) |
| Des2 | Superiority | 3 | 1-Sided | 0.025 | 0.9 | 1 | 473 | | 0.3 | | | Individual Means | 0 | 0.3 | 1 | Z | Equal | LD (OF) |
| Des3 | Superiority | 3 | 1-Sided | 0.025 | 0.9 | 1 | 526 | | 0.3 | | | Individual Means | 0 | 0.3 | 1 | Z | Equal | LD (OF) |

Note that the active window and the **Output Preview** can be minimized, maximized, or resized. If you want to focus on the **Output Summary**, click the  icon in the top-right corner of the main window. The Output will be maximized as shown below:

The screenshot shows the 'Output Summary' window maximized, displaying the parameters for 'Des2' in detail.

| | | Des2 |
|---|--|------------------|
| Mnemonic | | MN-25-D1 |
| Test Parameters | | |
| Design Type | | Superiority |
| No. of Looks | | 3 |
| Test Type | | 2-Sided |
| Specified α | | 0.05 |
| Power | | 0.9 |
| Model Parameters | | |
| Input Method | | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | | 0.3 |
| Mean Control (μ_c) | | 0 |
| Mean Treatment (μ_t) | | 0.3 |
| Std. Deviation (σ) | | 1 |
| Test Statistic | | Z |
| Allocation Ratio (nt/nc) | | 1 |
| Boundary Parameters | | |
| Efficacy Boundary | | LD (OF) |
| Spacing of Looks | | Equal |
| Sample Size | | |
| Maximum | | 473 |
| Expected Under H0 | | 471.064 |
| Expected Under H1 | | 379.185 |

Any of the designs/simulations in the **Output Preview** window can be saved in the **Library**, as depicted in the following workflow diagram.



Double click any of these nodes and the detailed output of the design will be displayed. This will include all relevant input and output information. Right clicking any design node in the **Library** will allow you to perform various operations on the design such as interim monitoring and simulation.

The **Help** pane displays the context sensitive help for the control currently under the focus. This help is available for all the controls in the **Input / Output** window. This pane also displays the design specific help which discusses the purpose of the selected test, the published literature referred while developing it and the chapter/section numbers of this user manual to quickly look-up for more details. This pane can be hidden or locked by clicking the pin in its corner.

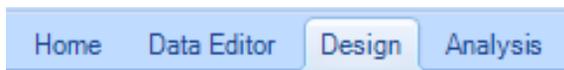
All the windows and features mentioned above are described in detail with the help of an illustration in the subsequent sections of this chapter.

3.2 A Quick Overview of User Interface

Almost all the functionalities of East 6 are invoked by selecting appropriate menu items and icons from the **Ribbon**. The interface consists of four windows as described

3 Getting Started

in the previous section and four major menu items. These menu items are:



- **Home.** This menu contains typical file-related Windows sub-menus. The **Help** sub-menu provides access to this manual.
- **Data Editor.** This menu will be available once a data set is open, providing several sub-menus used to create, manage and transform data.
- **Design.** This menu provides a sub-menu for each of the study designs which can be created using East 6. The study designs are grouped according to nature of the response. The tasks like Simulations and Interim Monitoring are available for almost all the study designs under this menu.
- **Analysis.** This menu provides a sub-menu for each of the analysis procedure that can be carried out in East 6. The tests are grouped according to the nature of the response. There are also options for basic statistics and plots.

3.3 Home Menu

3.3.1 File

3.3.2 Importing workbooks from East5.4

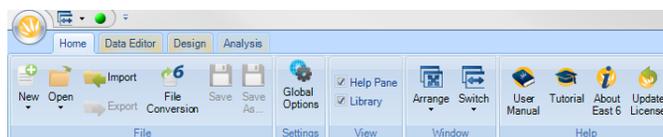
3.3.3 Settings

3.3.4 View

3.3.5 Window

3.3.6 Help

The **Home** menu contains icons that are logically grouped under *File*, *Settings*, *View*, *Window* and *Help*. These icons can be used for specific tasks.



3.3.1 File



New

Click this icon to create new case data or crossover data. A new workbook or log can also be created.



Open

Click this icon to open a saved data set, workbook, or log file.



Import

Click this icon to import external files created by other programs.



Export

Click this icon to export files in various formats.



Save

Click this icon to save the current files or workbooks.



Save As...

Click this icon to save a file or workbook with different name.

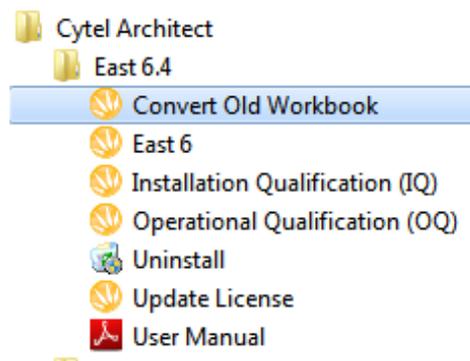
3.3.2 Importing workbooks from East5.4

East allows the conversion of workbooks previously created in East 5.4 (and above) to be imported into East 6 for further development. In order to open a workbook with the .es5 extension given by previous versions of East, it must first be converted to a file with the .cywx extension that will be recognized by East 6. This is easily accomplished

through the **Covert Old Workbook** utility. Click the  icon under **Home** menu to see the location of this utility.

From the **Start** Menu and select:

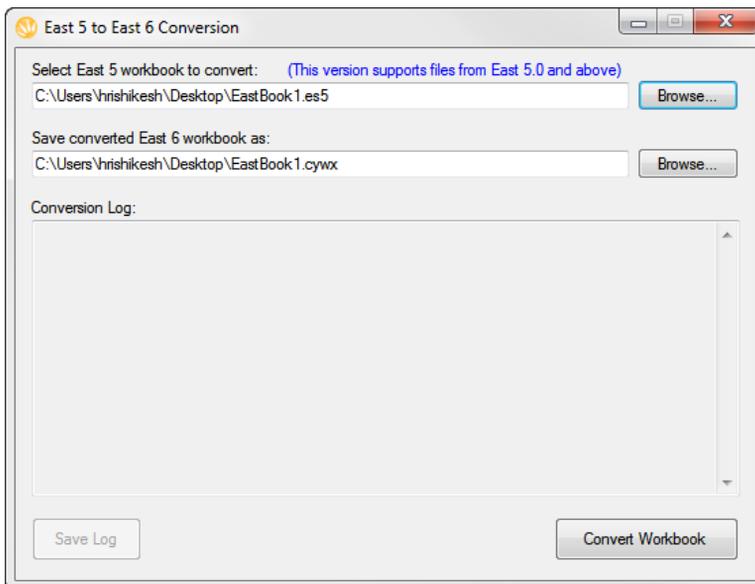
All Programs → **Cytel Architect** → **East 6.x** → **Convert Old Workbook**



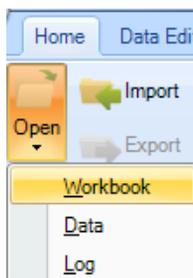
We can see the following window which accepts East5.4 workbook as input and outputs a workbook of East6. Click the **Browse** buttons to choose the East 5.4 file to

3 Getting Started

be converted and the file to be saved with **.cywx** extension of East 6 version.



To start the conversion, click **Convert Workbook**. Once complete, the file can be opened as a workbook in East 6 as shown below:

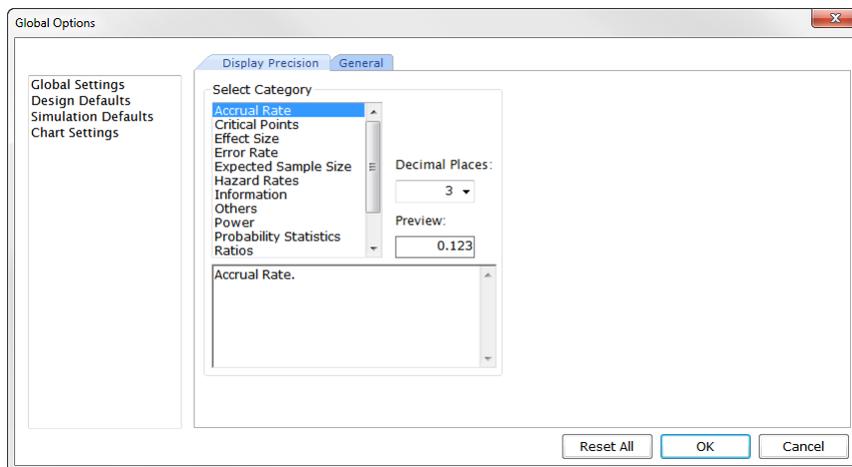


In order to convert files from East 5.3 or older versions, open the file in East 5.4, save it with a new name say with a suffix *.East5.4* and then convert this 5.4 file to 6.x as explained above. To get East 5.4 or any help regarding file conversion, contact Cytel at support@cytel.com.

3.3.3 Settings



Click the **Global Options** icon in the **Home** menu to adjust default values in East 6.



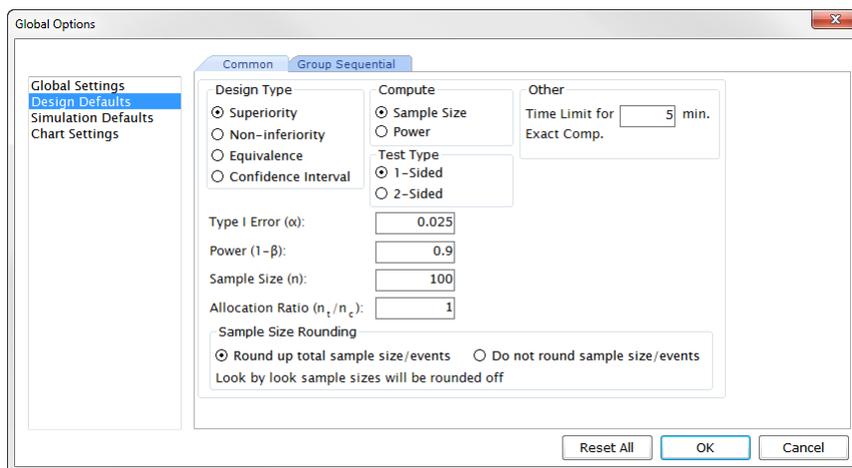
The options provided in the **Display Precision** tab are used to set the decimal places of numerical quantities. The settings indicated here will be applicable to all tests in East 6 under the **Design** and **Analysis** menus.

All these numerical quantities are grouped in different categories depending upon their usage. For example, all the average and expected sample sizes computed at simulation or design stage are grouped together under the category "Expected Sample Size". So to view any of these quantities with greater or lesser precision, select the corresponding category and change the decimal places to any value between 0 to 9.

The **General** tab has the provision of adjusting the paths for storing workbooks, files, and temporary files. These paths will remain throughout the current and future sessions even after East is closed. This is the place where we need to specify the installation directory of the R software in order to use the feature of R Integration in East6.

3 Getting Started

The **Design Defaults** is where the user can change the settings for trial design:

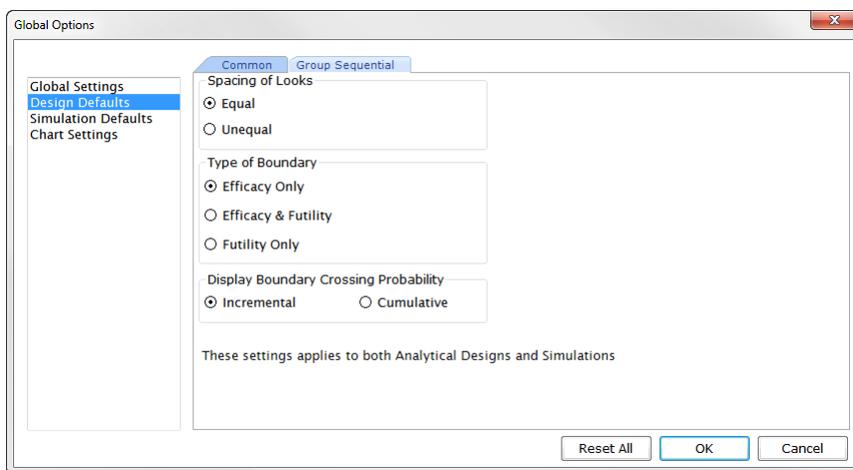


Under the **Common** tab, default values can be set for input design parameters.

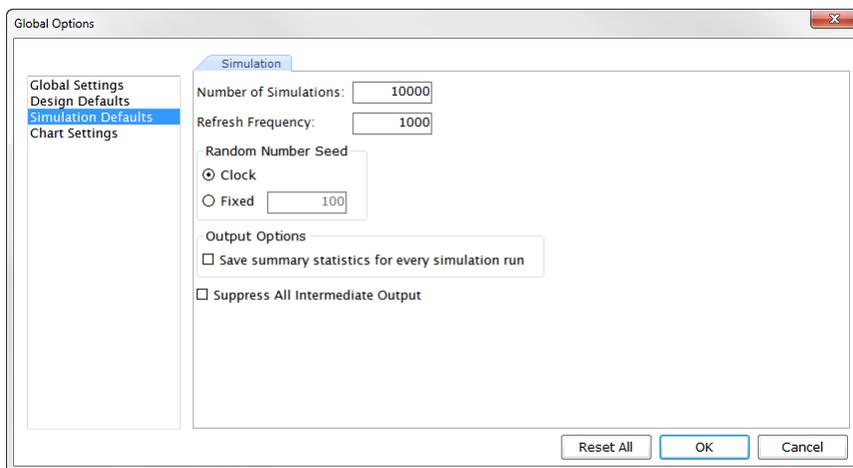
You can set up the default choices for the design type, computation type, test type and the default values for type-I error, power, sample size and allocation ratio. When a new design is invoked, the input window will show these default choices.

- **Time Limit for Exact Computation**
 This time limit is applicable only to exact designs and charts. Exact methods are computationally intensive and can easily consume several hours of computation time if the likely sample sizes are very large. You can set the maximum time available for any exact test in terms of minutes. If the time limit is reached, the test is terminated and no exact results are provided. Minimum and default value is 5 minutes.
- **Type I Error for MCP**
 If user has selected 2-sided test as default in global settings, then any MCP will use half of the alpha from settings as default since MCP is always a 1-sided test.
- **Sample Size Rounding**
 Notice that by default, East displays the integer sample size (events) by rounding up the actual number computed by the East algorithm. In this case, the look-by-look sample size is rounded off to the nearest integer. One can also see the original floating point sample size by selecting the option "Do not round sample size/events".

Under the **Group Sequential** tab, defaults are set for boundary information. When a new design is invoked, input fields will contain these specified defaults. We can also set the option to view the Boundary Crossing Probabilities in the detailed output. It can be either Incremental or Cumulative.



Simulation Defaults is where we can change the settings for simulation:



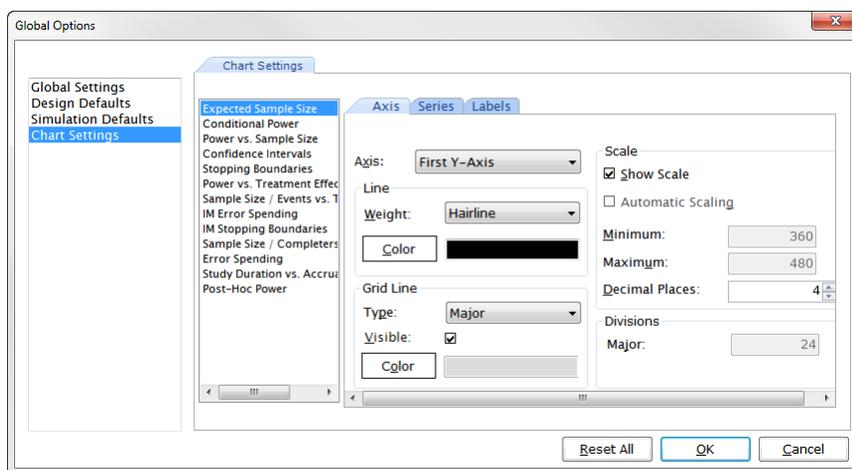
If the checkbox for "Save summary statistics for every simulation" is checked, then East simulations will by default save the per simulation summary data for all the

3 Getting Started

simulations in the form of a case data.

If the checkbox for "Suppress All Intermediate Output" is checked, the intermediate simulation output window will be always suppressed and you will be directed to the **Output Preview** area.

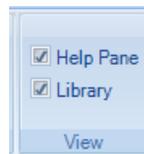
The **Chart Settings** allows defaults to be set for the following quantities on East6 charts:



We suggest that you do not alter the defaults until you are quite familiar with the software.

3.3.4 View

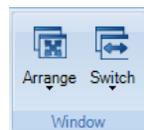
The **View** submenu consists of enabling or disabling the **Help** and **Library** panes by (un)checking the respective check boxes.



3.3.5 Window

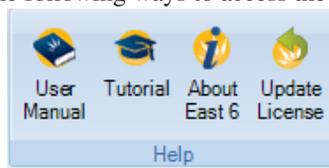
The **Window** submenu contains an **Arrange** and **Switch** option. This provides the

ability to view different standard arrangements of available windows (Design Input Output, Log, Details, charts and plots) and to switch the focus from one window to another.



3.3.6 Help

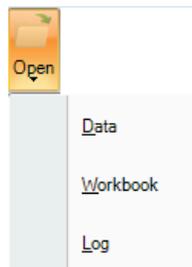
The **Help** group provides the following ways to access the East6 documentation:



- **User Manual:** Invoke the current East 6 user manual.
- **Tutorial:** Invoke the available East 6 tutorials.
- **About East 6:** Displays the current version and license information for the installed software.
- **Update License:** Use this utility to update the license file which you will be receiving from Cytel.

3.4 Data Editor Menu

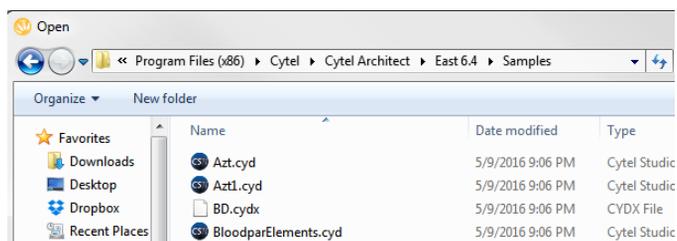
All submenus under the **Data Editor** menu are enabled once a new or existing data set is open. The **Open** command under the **Home** menu shows the list of items that can be opened:



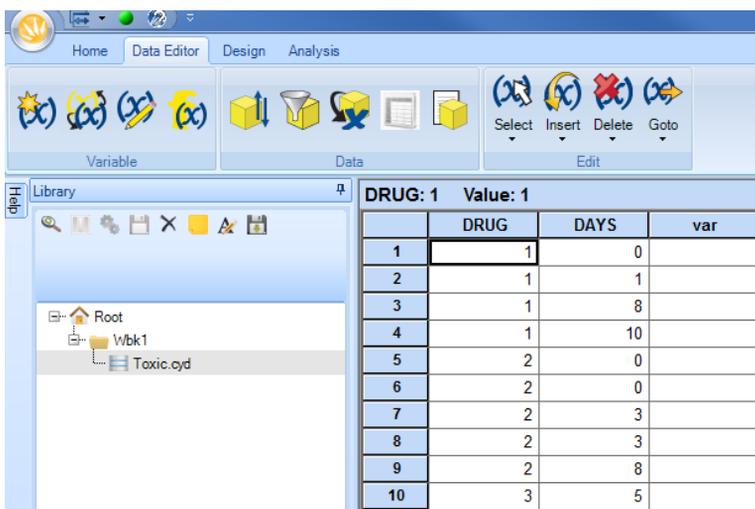
Suppose East 6 is installed in the directory **C:/Program Files (x86)/Cytel/Cytel**

3 Getting Started

Architect/East 6.4 on your machine. You can find sample datasets in the **Samples** under this directory.



Suppose, we open the file named **Toxic** from the **Samples** folder. The data is displayed in the main window under the **Data Editor** menu as shown:



Here the columns represent the variable and the rows are the different records. Placing the cursor on a cell containing data will enable all submenus under the **Data Editor** menu. The submenus are grouped into three sections, **Variable**, **Data** and **Edit**. Here we can modify and transform variables, perform operations on case data, and edit a case or variable in the data.

The icons in the **Variable** group are:



Creates a new variable at the current column position.

-  Renames the current variable.
-  Modifies the currently selected variable.
-  Transforms the currently selected variable.

Numerous algebraic, statistical functions are available which can be used to transform the variable. This feature can also be used to generate a data randomly from distributions such as Normal, Uniform, Chi-Square etc.

The following functions are available in the **Data** group:

-  Sorts case data in ascending or descending order.
-  Filter cases from the case data as per specified criteria.
-  Converts case data to crossover data.
-  Converts crossover data to case data.
-  Displays case data contents to the log window.

For the **Edit** group the following options are available:

-  Selects a case or variable.
-  Inserts a case or variable.
-  Deletes a case or variable.
-  Navigates to a specified case.

3.5 Analysis Menu

3.5.1 Basic Plots

3.5.2 Crossover Plots

The **Analysis** menu allows access to analytical tests which can be performed in East 6.



The tests available in the **Analysis** menus are grouped according to the nature of the response variable. Click an icon to select the test available in a drop down menu.

3 Getting Started

- **Basic Statistics** - This part contains tests to compute basic statistics and frequency distribution from a dataset.
- **Continuous** - This part groups analysis tests for continuous response.
- **Discrete** - This part groups all analysis tests for discrete response.
- **Events** - This group contains tests for time to event outcomes
- **Predict** - This group contains different procedures to predict the future course of the trial given the current subject level data or summary data. Refer to chapter 68 for more details.

3.5.1 Basic Plots



Bar and pie charts for categorical data.



Plots such as area, bubble, scatter plot and normality plots for continuous data.



Plots related to frequency distributions such as histogram, stem and leaf plots, cumulative plots.

3.5.2 Crossover Plots

This menu provides plots applicable to 2x2 crossover data.



Subject plots.



Summary plots.



Diagnostic plots.

All the tests under **Analysis** menu are discussed in detail under **Volume 8** of this manual.

3.6 Design Menu

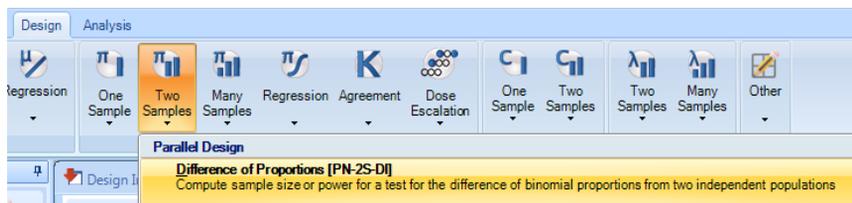
- 3.6.1 Design Input-Output Window
- 3.6.2 Creating Multiple Designs
- 3.6.3 Filter Designs
- 3.6.4 What is a Workbook?
- 3.6.5 Group Sequential Design for the CAPTURE Trial
- 3.6.6 Adding a Futility Boundary
- 3.6.7 Accrual Dropout Information
- 3.6.8 Output Details

This section discusses with the help of the CAPTURE trial the various East features mentioned so far in this chapter. This was a randomized clinical trial of placebo versus Abciximab for patients with refractory unstable angina. Results from this trial were presented at a workshop on clinical trial data monitoring committees *Randomised placebo-controlled trial of abciximab before and during coronary intervention in refractory unstable angina: the CAPTURE study, THE LANCET: Vol 349 - May 17, 1997.*

Let us design, simulate and monitor the CAPTURE trial using East6. The goal of this study is to test the null hypothesis, H_0 , that the Abciximab and placebo arms both have an event rate of 15%, versus the alternative hypothesis, H_1 , that Abciximab reduces the event rate by 5%, from 15% to 10%. It is desired to have a 2-Sided test with three looks at the data, a type-1 error, α as 0.05 and a power, $(1 - \beta)$ as 0.8.

We shall start with a fixed sample design and then extend it to group sequential design. In this process, we demonstrate the useful features of Architect one by one.

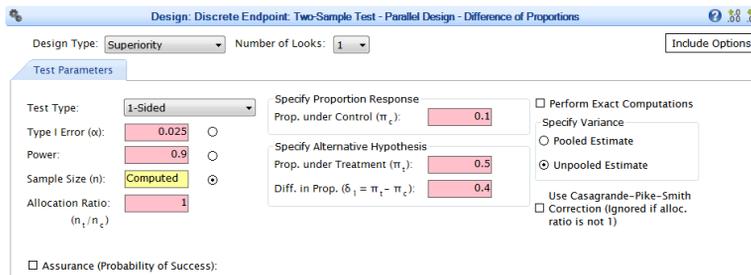
To begin, click **Design** menu, then **Two Samples** on the **Discrete** group, and then click **Difference of Proportions**.



Below the top ribbon, there are three windows: the **Input/Output**, the **Library**, and the **Help**. All these windows are explained in section 3.1 on Workflow of East. Both the **Library** and the **Help** can be hidden temporarily or throughout the session. The

3 Getting Started

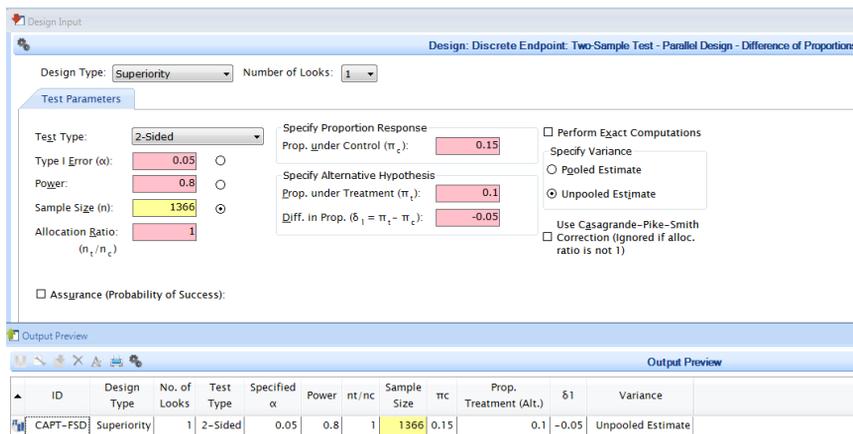
input window for **Difference of Proportions** test appears as shown below:



The design specific help can be accessed by clicking the  icon after invoking a design. This help is available for all the designs in East6.

3.6.1 Design Input-Output Window

This window is used to enter various design specific input parameters in the input fields and drop-down options available. Let us enter the following inputs for the CAPTURE Trial and create a fixed sample design. *Test Type* as 2-Sided, *Type I Error* as 0.05, *Power* as 0.8, π_c as 0.15 and π_t as 0.1. On clicking **Compute** button, a new row for this design gets added in the **Output Preview** window. Select this row and click the  icon. Rename this design as **CAPT-FSD** to indicate that it is a fixed sample design for the CAPTURE trial.



3.6.2 Creating Multiple Designs

Before finalizing on any particular study design, the statisticians might want to assess the operating characteristics of the trial under different conditions and over a range of parameter values. For example, when we are working on time-to-event trials, we want to see the effect of different values of hazard ratio on the overall power and duration of the study.

East makes it easy to rapidly generate and assess multiple options, to perform sensitivity analysis, and select the optimal plan. We can enter multiple values for one or more input parameters and East creates designs for all possible combinations. These designs can then be compared in a tabular as well as graphical manner.

Following are the three ways in which we can enter the multiple values:

- Comma-separated values: (0.8, 0.9, 0.95)
- Colon-separated range of values: (0.8 to 0.9 in steps of 0.05 can be entered as 0.8:0.9:0.05)
- Combined values: (0.7, 0.8, 0.85: 0.95: 0.01)

Multiple values can be entered only in the cells with **pink** background color.

Now suppose, we want to create designs for two values of **Type I Error**, three values of **Power** and four values of π_t : 0.1, 0.2 : 0.3 : 0.05. Without changing other parameters, let us enter these ranges for the three parameters as shown below:

On clicking **Compute** button, East will create $2 \times 3 \times 4 = 24$ designs for the CAPTURE Trial. To view all the designs in the **Output Preview** window, maximize it

3 Getting Started

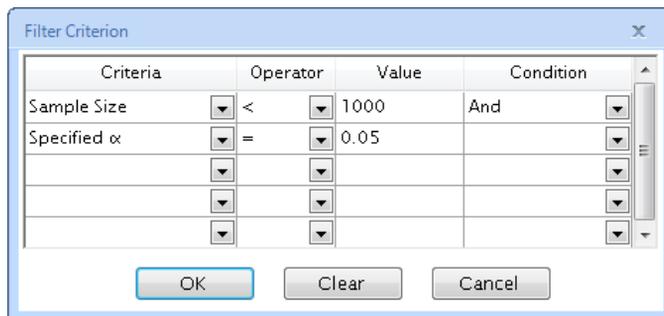
from the right-hand top.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | πc | Prop. Treatment (Alt.) | $\delta 1$ | Variance |
|-------|-------------|--------------|-----------|--------------------|-------|-------|-------------|---------|------------------------|------------|-------------------|
| Des1 | Superiority | 1 | 2-Sided | 0.025 | 0.8 | 1 | 1654 | 0.15 | 0.1 | -0.05 | Unpooled Estimate |
| Des2 | Superiority | 1 | 2-Sided | 0.025 | 0.8 | 1 | 2187 | 0.15 | 0.2 | 0.05 | Unpooled Estimate |
| Des3 | Superiority | 1 | 2-Sided | 0.025 | 0.8 | 1 | 599 | 0.15 | 0.25 | 0.1 | Unpooled Estimate |
| Des4 | Superiority | 1 | 2-Sided | 0.025 | 0.801 | 1 | 286 | 0.15 | 0.3 | 0.15 | Unpooled Estimate |
| Des5 | Superiority | 1 | 2-Sided | 0.05 | 0.8 | 1 | 1366 | 0.15 | 0.1 | -0.05 | Unpooled Estimate |
| Des6 | Superiority | 1 | 2-Sided | 0.05 | 0.8 | 1 | 1806 | 0.15 | 0.2 | 0.05 | Unpooled Estimate |
| Des7 | Superiority | 1 | 2-Sided | 0.05 | 0.8 | 1 | 495 | 0.15 | 0.25 | 0.1 | Unpooled Estimate |
| Des8 | Superiority | 1 | 2-Sided | 0.05 | 0.801 | 1 | 236 | 0.15 | 0.3 | 0.15 | Unpooled Estimate |
| Des9 | Superiority | 1 | 2-Sided | 0.025 | 0.85 | 1 | 1870 | 0.15 | 0.1 | -0.05 | Unpooled Estimate |
| Des10 | Superiority | 1 | 2-Sided | 0.025 | 0.85 | 1 | 2472 | 0.15 | 0.2 | 0.05 | Unpooled Estimate |
| Des11 | Superiority | 1 | 2-Sided | 0.025 | 0.85 | 1 | 677 | 0.15 | 0.25 | 0.1 | Unpooled Estimate |
| Des12 | Superiority | 1 | 2-Sided | 0.025 | 0.851 | 1 | 323 | 0.15 | 0.3 | 0.15 | Unpooled Estimate |
| Des13 | Superiority | 1 | 2-Sided | 0.05 | 0.85 | 1 | 1563 | 0.15 | 0.1 | -0.05 | Unpooled Estimate |
| Des14 | Superiority | 1 | 2-Sided | 0.05 | 0.85 | 1 | 2066 | 0.15 | 0.2 | 0.05 | Unpooled Estimate |
| Des15 | Superiority | 1 | 2-Sided | 0.05 | 0.85 | 1 | 566 | 0.15 | 0.25 | 0.1 | Unpooled Estimate |
| Des16 | Superiority | 1 | 2-Sided | 0.05 | 0.851 | 1 | 270 | 0.15 | 0.3 | 0.15 | Unpooled Estimate |
| Des17 | Superiority | 1 | 2-Sided | 0.025 | 0.9 | 1 | 2160 | 0.15 | 0.1 | -0.05 | Unpooled Estimate |
| Des18 | Superiority | 1 | 2-Sided | 0.025 | 0.9 | 1 | 2855 | 0.15 | 0.2 | 0.05 | Unpooled Estimate |
| Des19 | Superiority | 1 | 2-Sided | 0.025 | 0.9 | 1 | 782 | 0.15 | 0.25 | 0.1 | Unpooled Estimate |
| Des20 | Superiority | 1 | 2-Sided | 0.025 | 0.901 | 1 | 373 | 0.15 | 0.3 | 0.15 | Unpooled Estimate |
| Des21 | Superiority | 1 | 2-Sided | 0.05 | 0.9 | 1 | 1829 | 0.15 | 0.1 | -0.05 | Unpooled Estimate |
| Des22 | Superiority | 1 | 2-Sided | 0.05 | 0.9 | 1 | 2417 | 0.15 | 0.2 | 0.05 | Unpooled Estimate |
| Des23 | Superiority | 1 | 2-Sided | 0.05 | 0.9 | 1 | 662 | 0.15 | 0.25 | 0.1 | Unpooled Estimate |
| Des24 | Superiority | 1 | 2-Sided | 0.05 | 0.901 | 1 | 316 | 0.15 | 0.3 | 0.15 | Unpooled Estimate |

3.6.3 Filter Designs

Suppose we are interested in designs with some specific input/output values, we can set up a criterion by using **Filter** functionality by clicking the  icon available on the top right corner of **Output Preview** window. For example, we want to see designs with *Sample Size* less than 1000 and *Type I Error*

equal to 0.05.



The qualified designs appear in the **Output Preview** window as shown below:

| | ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | nc | Prop. Treatment (Alt.) | δ1 | Variance |
|--|-------|-------------|--------------|-----------|-------------|-------|-------|-------------|------|------------------------|------|-------------------|
| | Des7 | Superiority | 1 | 2-Sided | 0.05 | 0.8 | 1 | 495 | 0.15 | 0.25 | 0.1 | Unpooled Estimate |
| | Des8 | Superiority | 1 | 2-Sided | 0.05 | 0.801 | 1 | 236 | 0.15 | 0.3 | 0.15 | Unpooled Estimate |
| | Des15 | Superiority | 1 | 2-Sided | 0.05 | 0.85 | 1 | 566 | 0.15 | 0.25 | 0.1 | Unpooled Estimate |
| | Des16 | Superiority | 1 | 2-Sided | 0.05 | 0.851 | 1 | 270 | 0.15 | 0.3 | 0.15 | Unpooled Estimate |
| | Des23 | Superiority | 1 | 2-Sided | 0.05 | 0.9 | 1 | 662 | 0.15 | 0.25 | 0.1 | Unpooled Estimate |
| | Des24 | Superiority | 1 | 2-Sided | 0.05 | 0.901 | 1 | 316 | 0.15 | 0.3 | 0.15 | Unpooled Estimate |

The filter criteria can be edited or cleared by again clicking the Filter icon. On clearing the above criterion, all the 24 designs are displayed back. Before we proceed, let us first delete these recently created 24 designs, leaving behind **CAPT-FSD** and then minimize the **Output Preview** window from the right-hand top.

One or more rows in the can be deleted by selecting them and clicking the icon. Use the **Ctrl** key and mouse click to select specific rows. Use the **Shift** key and mouse click to select all the rows in the range. Use the combination **Ctrl + A** to select all the rows. The resulting **Output Preview** is shown below:

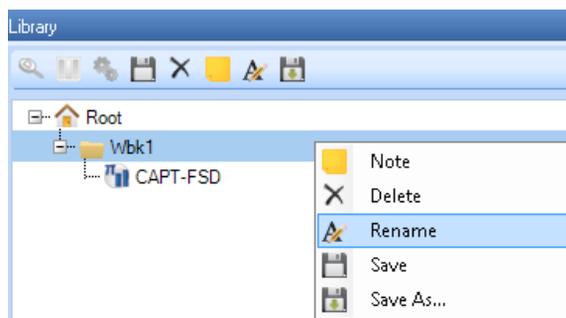
| | ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | nc | Prop. Treatment (Alt.) | δ1 | Variance |
|--|----------|-------------|--------------|-----------|-------------|-------|-------|-------------|------|------------------------|-------|-------------------|
| | CAPT-FSD | Superiority | 1 | 2-Sided | 0.05 | 0.8 | 1 | 1366 | 0.15 | 0.1 | -0.05 | Unpooled Estimate |

It is advisable to save this design or any work which you would like to refer in future in an East **Workbook**. The next subsection briefly discusses about use of workbooks.

3 Getting Started

3.6.4 What is a Workbook?

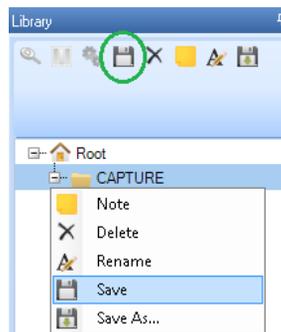
A *Workbook* is a storage construct managed by East for holding different types of generated outputs. The user designs a trial, simulates it, monitors it at several interim looks, conducts certain analyses, draws plots, etc. All of these outputs can be kept together in a workbook which can be saved and retrieved for further development when required. . Note that a single workbook can also contain outputs from more than one design. Select the design **CAPT-FSD** in the **Output Preview** window and click the  icon. When a design is saved to the library for the first time, East automatically creates a workbook named **Wbk1** which can be renamed by right-clicking the node.



Let us name it as **CAPTURE**. Now this is still a temporary storage which means if we exit out of East without saving it permanently, the workbook will not be available in future.

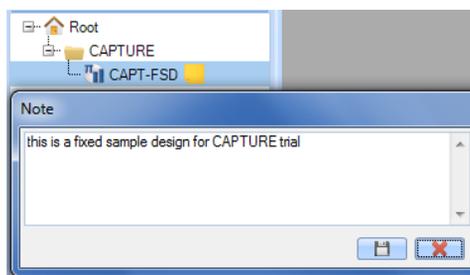
Note that Workbooks are not saved automatically on your computer; they are to be saved by either right-clicking the node in the Library and selecting Save or

clicking the  icon.



In addition, the user will be prompted to save contents of the **Library** while closing East 6.

Many a times, we wish to add some specific comments to a design or any other output window. These comments are useful for future references. One can do that by attaching a *Note* to any node by selecting it and clicking on the  icon. A small window will pop up where comments can be stored.



Once saved, a yellow icon against the design node will indicate the presence of a note. If you want to view or remove the note, right click the design node, select **Note**, and clear the contents.

The tabs available on the status bar at the bottom left of the screen can be used to navigate between the active windows of East.

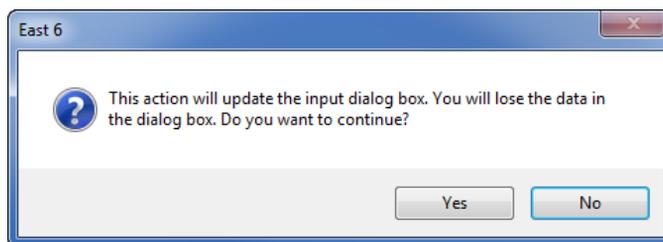
3 Getting Started



For example, if you wish to return to the design inputs, click the **Input** button which will take you the latest Input window you worked with. As we proceed further, more such tabs will appear enabling us to navigate from one screen of East to another.

3.6.5 Group Sequential Design for the CAPTURE Trial

Select the design **CAPT-FSD** and click the  icon in the **Library** to modify the design. On clicking this icon, following message will pop up. Click "Yes" to continue.



Let us extend this fixed sample design to a group sequential design by changing the *Number of Looks* from 1 to 3. It means that we are planning to take 2 interim looks and one final look at the data while monitoring the study.

Number of Looks:

An additional tab named **Boundary** is added which allows us to enter inputs related to

the boundary family, look spacing and error spending functions.

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.05

Futility

Boundary Family: None

Spacing of Looks: Equal Unequal

Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.333 | 0.000 | 3.710 | -3.710 |
| 2 | 0.667 | 0.012 | 2.511 | -2.511 |
| 3 | 1.000 | 0.050 | 1.993 | -1.993 |

Let the boundary family be **Spending Functions** and the alpha spending function, **Lan-DeMets** with the parameter **OF**. Click on **Compute** to create the three-look design and rename it as **CAPT-GSD**.

As you go on creating multiple designs in East, the output preview area can become too busy to manage. Thus, you can also select the designs you are interested in, save them in the workbook and then rename them appropriately. The **Output Preview** window now looks as shown below:

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | mc | Prop. Treatment (Alt.) | δ_1 | Variance | Spacing of Looks | Efficacy Boundary | Expected SS (H0) | Expected SS (H1) |
|----------|-------------|--------------|-----------|--------------------|-------|-------|-------------|------|------------------------|------------|-------------------|------------------|-------------------|------------------|------------------|
| CAPT-FSD | Superiority | 1 | 2-Sided | 0.05 | 0.8 | 1 | 1366 | 0.15 | 0.1 | -0.05 | Unpooled Estimate | | | | |
| CAPT-GSD | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | 1384 | 0.15 | 0.1 | -0.05 | Unpooled Estimate | Equal | LD (OF) | 1378.32 | 1182.678 |

Notice that **CAPT-GSD** requires 18 subjects more than **CAPT-FSD** to achieve 80% power. This view gives us the horizontal comparison of two designs. Save the design **CAPT-GSD** in the workbook.

One can also compare these designs in a vertical manner. Select the two designs by clicking on one of them, pressing **Ctrl** and then clicking on the other one. Next, click

3 Getting Started

the  icon.

| | CAPT-FSD | CAPT-GSD |
|--|-------------------|-------------------|
| Mnemonic | PN-2S-DI | PN-2S-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 3 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.8 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Proportion under Control (π_c) | 0.15 | 0.15 |
| Proportion under Treatment (π_t) | 0.1 | 0.1 |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.05 | -0.05 |
| Variance | Unpooled Estimate | Unpooled Estimate |
| Boundary Parameters | | |
| Spacing of Looks | | Equal |
| Efficacy Boundary | | LD (OF) |
| Sample Size | | |
| Maximum | 1366 | 1384 |
| Expected Under H0 | | 1378.32 |
| Expected Under H1 | | 1182.678 |

This is the **Output Summary** window of East which compares the two designs vertically. We can easily copy this display from East to MS Excel and modify/save it further in any other format. To do that, right click anywhere in the **Output Summary** window, select **Copy All** option and paste the copied data in an Excel workbook. The table gets pasted as two formatted columns.

Let us go back to the input window of **CAPT-GSD** (select the design and click the  icon) and activate the **Boundary** tab. By default, the boundary values in the table at the bottom of this tab are displayed on **Z Scale**. We can also view these boundaries on other scales such as: *Score Scale*, *δ Scale* and *p-value Scale*.

Spacing of Looks: Equal Unequal Efficacy Boundary: Z Scale  

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.333 | 0.000 | 3.710 | -3.710 |
| 2 | 0.667 | 0.012 | 2.511 | -2.511 |
| 3 | 1.000 | 0.050 | 1.993 | -1.993 |

Let us view the efficacy boundaries for **CAPT-GSD** on a p-value scale.

Spacing of Looks: Equal Unequal Efficacy Boundary: p-value Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|-------|
| | | | Upper | Lower |
| 1 | 0.333 | 0.000 | 0.000 | 0.000 |
| 2 | 0.667 | 0.012 | 0.006 | 0.006 |
| 3 | 1.000 | 0.050 | 0.023 | 0.023 |

The final p-value required to attain statistical significance at level 0.05 is 0.0463. This is sometimes regarded as the penalty for taking two interim looks at the data.

Also observe that, although the maximum sample size for this design is 1384, the expected sample size under alternative that $\delta = -0.05$ is much less, 1183. However, there is very little saving under the null hypothesis that $\delta = 0$. The sample size in this case is 1378.

Therefore, it might be beneficial to consider replacing the lower efficacy boundary by a futility boundary. Also, sometimes we might wish to stop a trial early because the effect size observed at an interim analysis is too small to warrant continuation. This can be achieved by using β -spending function and introducing a futility boundary at the design stage.

3.6.6 Adding a Futility Boundary

Select the design **CAPT-GSD** and click  icon to edit it. Change the *Test Type* from **2-Sided** to **1-Sided** and also the *Type I Error* from 0.05 to 0.025. Go to **Boundary** tab and add the futility boundaries by using γ (-2) spending function.

Test Parameters **Boundary**

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.025

Futility

Boundary Family: Spending Functions

Spending Function: Gamma Family

Parameter (γ): -2

Type II Error (β): 0.2

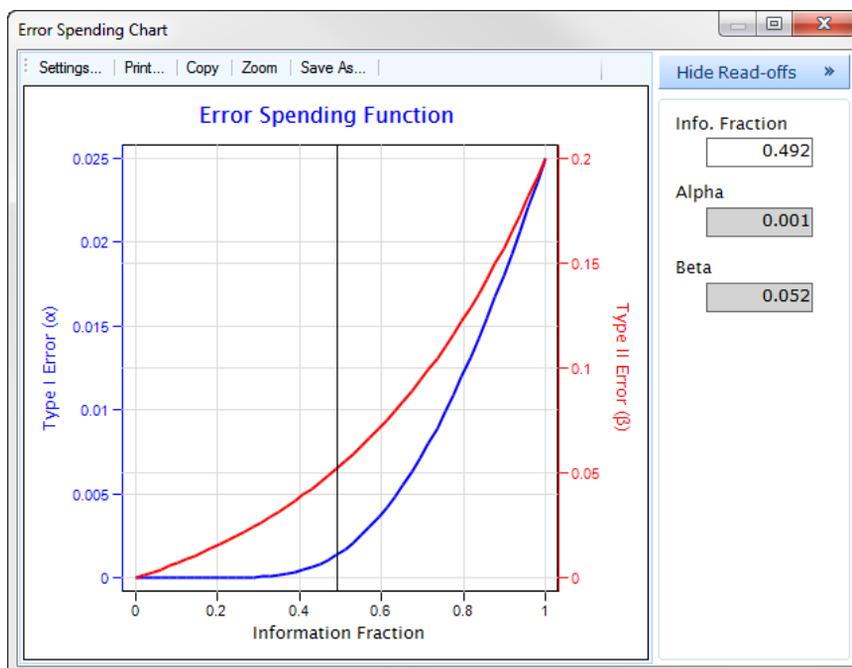
Non-Binding
 Binding

Spacing of Looks: Equal Unequal Boundary Scale: Z Scale

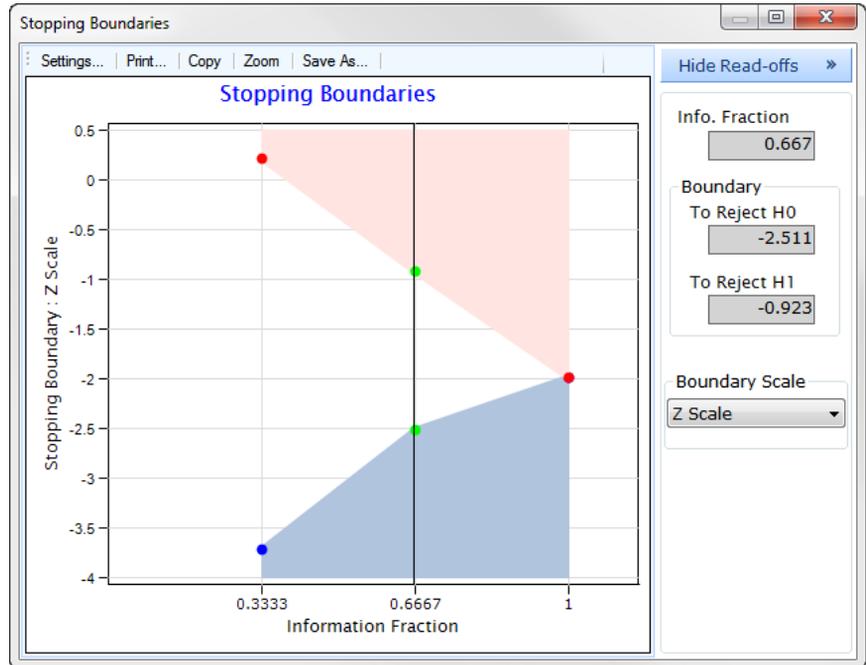
| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Cum. β Spent | Futility Boundary |
|--------|----------------|-------------------------------------|-------------------------------------|---------------------|-------------------|--------------------|-------------------|
| 1 | 0.333 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.000 | -3.710 | 0.030 | 0.216 |
| 2 | 0.667 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.006 | -2.511 | 0.087 | -0.923 |
| 3 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.025 | -1.993 | 0.200 | -1.993 |

3 Getting Started

Before we create this design, we can see the error spending chart and the boundaries chart for the CAPTURE trial with efficacy as well as futility boundaries. This gives us a way to explore different boundary families and error spending functions and deciding upon the desired combination before even creating a design. Click the  icon to view the Error Spending Chart.



Click the  icon to view the Boundaries Chart.

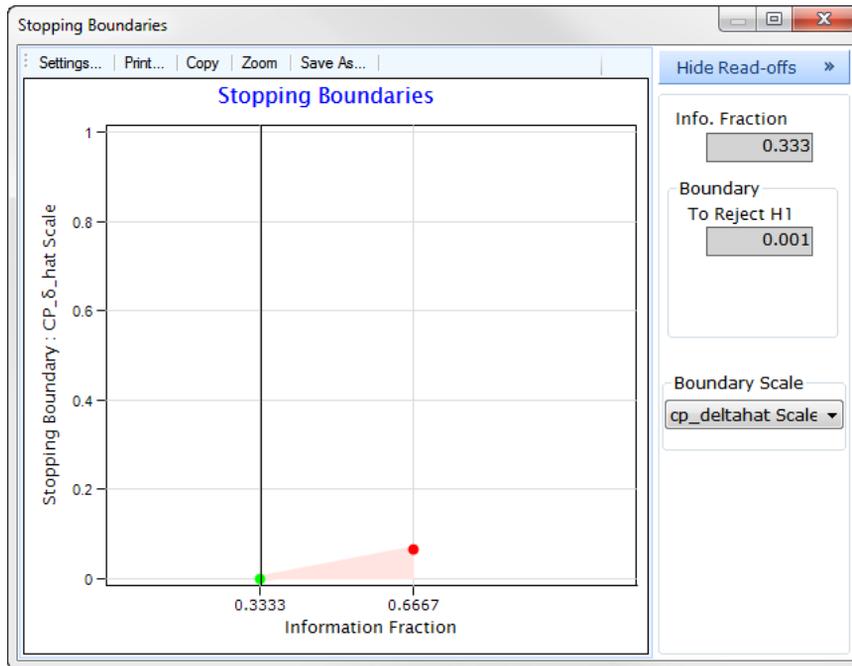


The shaded region in light pink corresponds to the critical region for futility and the one in light blue corresponds to the critical region for efficacy.

We can also view the boundaries on conditional power scale in presence of a futility boundary. Select the entry named *cp_deltahat Scale* from the dropdown **Boundary**

3 Getting Started

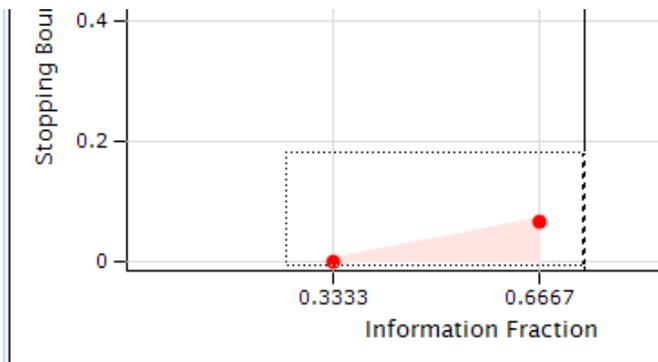
Scale. The chart is be updated and the boundaries are displayed on CP scale.



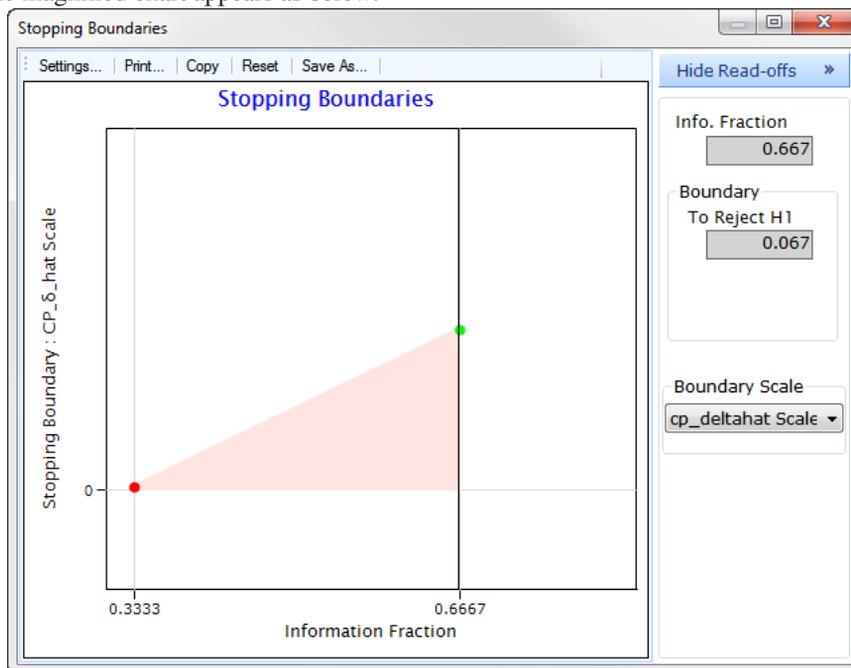
Zooming the Charts

To zoom into any area of the chart, click and drag the mouse over that area. After clicking Zoom button, click on the plot at the top left corner of the area you want to magnify, keep the mouse button pressed and drag the mouse over the desired area. This draws a rectangle around that area. Now leave the mouse button and East magnifies the

selected area. You can keep doing this to zoom in further.



The magnified chart appears as below:



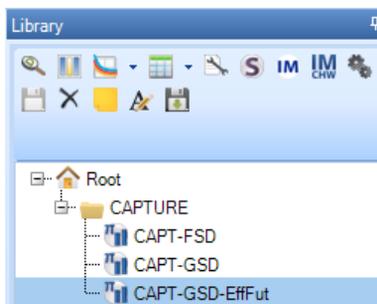
Note that after zooming, the Zoom button changes to Reset. When you click it, the plot

3 Getting Started

is reset back to the original shape.

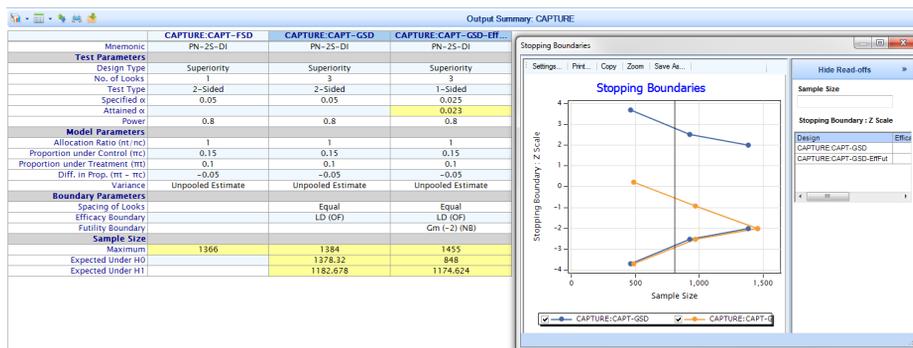
Let us compute the third design for the CAPTURE trial and rename it as **CAPT-GSD-EffFut**. Save it in the workbook.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | tTC | Prop. Treatment (Alt.) | δ | Variance | Spacing of Looks | Efficacy Boundary | Futility Boundary | Expected SS (H0) | Expected SS (H1) | Attained α |
|-----------------|-------------|--------------|-----------|--------------------|-------|-------|-------------|------|------------------------|----------|-------------------|------------------|-------------------|-------------------|------------------|------------------|-------------------|
| CAPT-FSD | Superiority | 1 | 2-Sided | 0.05 | 0.8 | 1 | 1366 | 0.15 | 0.1 | -0.05 | Unpooled Estimate | Equal | LD (OF) | | | | |
| CAPT-GSD | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | 1384 | 0.15 | 0.1 | -0.05 | Unpooled Estimate | Equal | LD (OF) | 1378.32 | 1182.678 | | |
| CAPT-GSD-EffFut | Superiority | 3 | 1-Sided | 0.025 | 0.8 | 1 | 1455 | 0.15 | 0.1 | -0.05 | Unpooled Estimate | Equal | LD (OF) | Gm (-2) (NB) | 848 | 1174.624 | 0.023 |



Click the icon to compare all the three designs side-by-side as explained above.

Along with the side-by-side comparison, let us compare the two group sequential designs graphically. Press **Ctrl** and click on **CAPT-FSD**. Notice that the remaining two designs are still highlighted which means they are selected and **CAPT-FSD** is unselected. Now click the icon and select **Stopping Boundaries** to view the graphical comparison of boundaries of the two designs.



As we can see, the design **CAPT-GSD** uses an upper efficacy boundary whereas **CAPT-GSD-EffFut** uses an upper futility boundary. We can turn ON and OFF the boundaries by checking the boxes available in the legends.

Before we proceed, let us save this third design in the workbook. We can also create several workbooks in the **Library** and then compare multiple designs across the workbooks. This is an advantage of working with workbooks in East6.

3.6.7 Accrual / Dropout option for Continuous and Discrete Endpoints

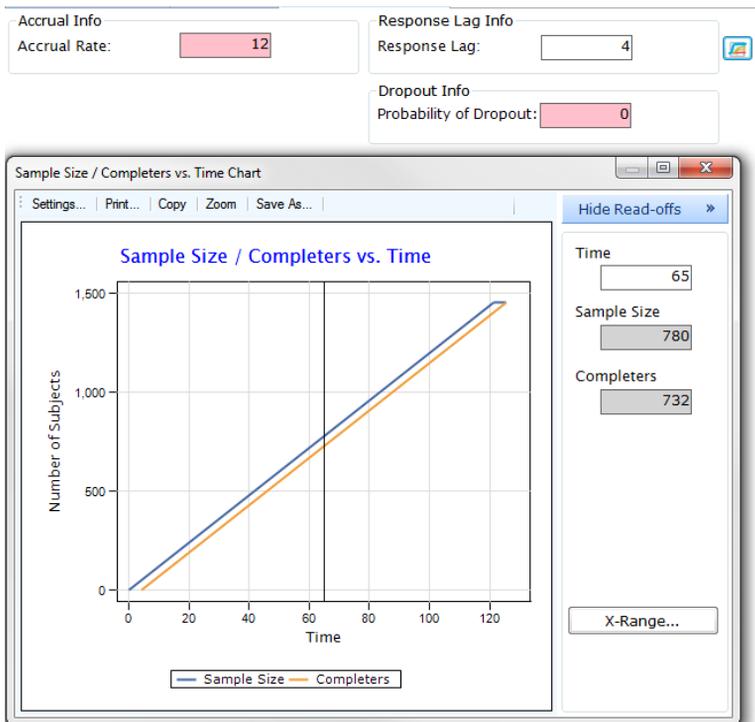
In the earlier versions of East, the option to incorporate the accrual and dropout information was available only for tests under time-to-event/survival endpoint. East 6 now provides this option for almost all the tests under Continuous and Discrete endpoints as well. Let us see the use it in CAPTURE trial. Select the design **CAPT-GSD-EffFut** from the **Library** and edit it to add the accrual-dropout information. From the **Design Parameters** tab, add the option **Accrual/Dropout Info** by clicking on **Include Options** button.

The screenshot shows the 'Accrual / Dropouts' configuration panel. It is divided into three sections: 'Accrual Info', 'Response Lag Info', and 'Dropout Info'.
- 'Accrual Info': 'Accrual Rate' is set to 8.
- 'Response Lag Info': 'Response Lag' is set to 0.
- 'Dropout Info': 'Probability of Dropout' is set to 0.

Let the accrual rate be 12 subjects/week. Suppose we expect the response to be observed after 4 weeks from the recruitment. Let us create a design by first assuming that there will not be any dropouts during the course of trial. We will then introduce some dropouts and compare the two designs. After entering the above inputs, click on

3 Getting Started

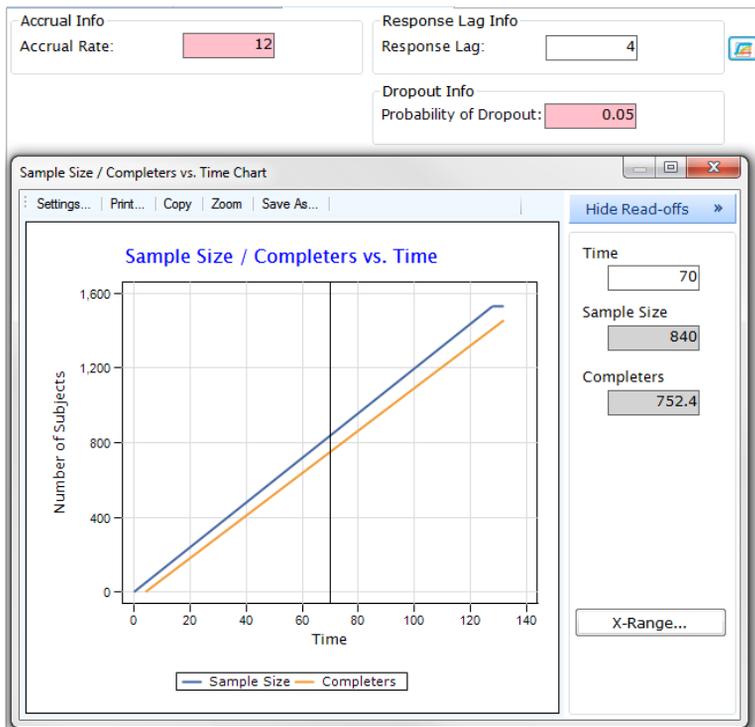
the  icon to see how the subjects will accrue and complete the study.



Close the chart, create the design by clicking the **Compute** button, save it in the workbook **CAPTURE** and rename it as **CAPT-GSD-NoDrp** to indicate that there are no dropouts in this design. Notice that in this design, the maximum sample size and maximum number of completers is same as there is no dropout.

Let us now introduce dropouts. Suppose there is a 5% chance of a subject dropping out

of the trial.

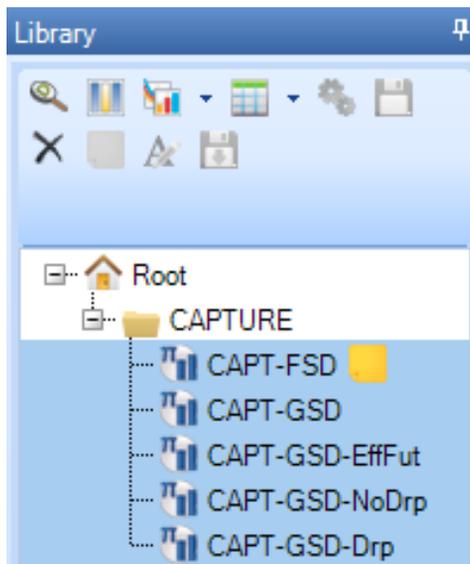


Notice that the two lines are not parallel anymore because of the presence of dropouts. Click **Compute** button to create this design. Save the design in the workbook **CAPTURE** and rename it as **CAPT-GSD-Drp**. Compare this design with **CAPT-GSD-NoDrp** by selecting the two designs and clicking on  icon

Notice the inflation in sample size for **CAPT-GSD-Drp**. This design will require additional 80 subjects to obtain data on 1455 subjects (1455 completers). Let us now compare all the five designs saved in the workbook. Select them all

3 Getting Started

together and click the  icon.



The resulting screen will look as shown below:

| | CAPTURE-CAPT-FSD | CAPTURE-CAPT-GSD | CAPTURE-CAPT-GSD-Eff... | CAPTURE-CAPT-GSD-No... | CAPTURE-CAPT-GSD-Drp |
|---|-------------------|-------------------|-------------------------|------------------------|----------------------|
| Mnemonic | PN-25-DI | PN-25-DI | PN-25-DI | PN-25-DI | PN-25-DI |
| Test Parameters | | | | | |
| Design Type | Superiority | Superiority | Superiority | Superiority | Superiority |
| No. of Looks | 1 | 3 | 3 | 3 | 3 |
| Test Type | 2-Sided | 2-Sided | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 | 0.025 | 0.025 | 0.025 |
| Attained α | | | 0.023 | 0.023 | 0.023 |
| Power | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 |
| Model Parameters | | | | | |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 | 1 | 1 |
| Proportion under Control (trc) | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 |
| Proportion under Treatment (trt) | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| Diff. in Prop. (trt - trc) | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 |
| Variance | Unpooled Estimate | Unpooled Estimate | Unpooled Estimate | Unpooled Estimate | Unpooled Estimate |
| Boundary Parameters | | | | | |
| Spacing of Looks | Equal | Equal | Equal | Equal | Equal |
| Efficacy Boundary | | LD (OF) | LD (OF) | LD (OF) | LD (OF) |
| Futility Boundary | | | Gm (-2) (NB) | Gm (-2) (NB) | Gm (-2) (NB) |
| Accrual & Dropout Parameters | | | | | |
| Accrual Rate | | | | 12 | 12 |
| Response Lag | | | | 4 | 4 |
| Probability of Dropout | | | | 0 | 0.05 |
| Sample Size | | | | | |
| Maximum | 1366 | 1384 | 1455 | 1455 | 1532 |
| Expected Under H0 | | 1378.32 | 848 | 888.177 | 933.474 |
| Expected Under H1 | | 1182.678 | 1174.624 | 1199.958 | 1262.455 |
| Completers | | | | 1455 | 1455 |
| Maximum | | | | 848 | 848 |
| Expected Under H0 | | | | 1174.624 | 1174.624 |
| Expected Under H1 | | | | | |
| Study Duration | | | | | |
| Maximum | | | | 125.25 | 131.667 |

We can see additional quantities in the design **CAPT-GSD-Drp**. These correspond to

the information on total number of completers and the study duration which are computed by taking into account the non-zero response lag and possibility of dropouts. Also notice the trend in **Maximum Sample Size** across all these designs. We can see that it increases as more constraints are added to the study. But if we see values of **Expected Sample Size** under null and alternative, there is a significant potential saving.

You can also save this output summary window comparing three designs in the library by clicking the  icon

3.6.8 Output Details

In the earlier part of this chapter, we have seen the design output at two different places: **Output Preview** (horizontal view) and **Output Summary** (vertical view). The final step in the East6 design workflow is to see the detailed output in the form of an **HTML** file.

Select the design **CAPT-GSD-Drp** from the **Library** and click the  icon. Alternatively, one can also double-click on any of the nodes in the **Library** to see the

3 Getting Started

details.

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Test Parameters | |
|-----------------------------------|-------------------|
| Design ID | CAPT-GSD-Drp |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Attained α | 0.023 |
| Power | 0.8 |
| Model Parameters | |
| Prop. under Control (π_c) | 0.15 |
| Prop. under Treatment (π_t) | 0.1 |
| $\delta = \pi_t - \pi_c$ | |
| Under H0 | 0 |
| Under H1 | -0.05 |
| Allocation Ratio (n_t/n_c) | 1 |
| Variance | Unpooled Estimate |
| Casagrande-Pike-Smith Correction | Not Applied |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Futility Boundary | Gm (-2) (NB) |
| Accrual / Dropouts Parameters | |
| Accrual Rate | 12 |
| Response Lag | 4 |
| Probability of Dropout | 0.05 |

Sample sizes and completers have been rounded.

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|-------------------------|-------------|---------------|-----------|
| Sample Size (n) | | | |
| Maximum | 766 | 766 | 1532 |
| Expected H1 | 631,253 | 631,202 | 1,262,455 |
| Expected H0 | 466,944 | 466,63 | 933,474 |
| Completers (s) | | | |
| Maximum | 728 | 727 | 1455 |
| Expected H1 | 587,573 | 587,051 | 1,174,624 |
| Expected H0 | 424,288 | 423,711 | 848 |
| Dropouts (d) | | | |
| Maximum | 38 | 39 | 77 |
| Expected H1 | 31,013 | 31,485 | 62,497 |
| Expected H0 | 22,567 | 22,73 | 45,297 |
| Maximum Information (i) | | | 3344,828 |

Accrual and Study Duration

| | Accrual Duration | Study Duration |
|-------------|------------------|----------------|
| Maximum | 127,667 | 131,667 |
| Expected H1 | 103,093 | 107,093 |
| Expected H0 | 74,441 | 78,441 |

Completers, Sample Size, Dropouts, Pipeline and Analysis Times: Look by Look

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Completers (s) | Dropouts (d) | Pipeline (p.s-d) | Analysis Time | Boundary Crossing Probability (Incremental) | | | |
|--------|--------------------------|-----------------|----------------|--------------|------------------|---------------|---|----------|----------|----------|
| | | | | | | | Under H0 | | Under H1 | |
| | | | | | | | Efficacy | Futility | Efficacy | Futility |
| 1 | 0.333 | 559 | 485 | 26 | 48 | 46,583 | 1.035E-4 | 0.414 | 0.021 | 0.03 |
| 2 | 0.667 | 1070 | 970 | 52 | 48 | 89,167 | 0.006 | 0.417 | 0.42 | 0.058 |
| 3 | 1 | 1532 | 1455 | 77 | 0 | 131,667 | 0.017 | 0.146 | 0.36 | 0.113 |

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Cumulative α Spent | Cumulative β Spent | Boundaries | |
|--------|--------------------------|-----------------|---------------------------|--------------------------|------------|------------|
| | | | | | Efficacy Z | Futility Z |
| 1 | 0.333 | 559 | 1.035E-4 | 0.03 | -3.71 | 0.216 |
| 2 | 0.667 | 1070 | 0.006 | 0.087 | -2.511 | -0.923 |
| 3 | 1 | 1532 | 0.026 | 0.2 | -1.993 | -1.993 |

The output details are broadly divided into two panels. The left panel consists of all the input parameters and the right panel consists of all the design output quantities in the tabular format. These tables will be explained in detail in subsequent chapters of this manual.

Click the Save icon to save all the work done so far. This is the end of introduction to the **Design Menu**. The next section discusses another very useful feature called **Simulations**.

3.7 Simulations in East6

A simulation is a very useful way to perform sensitivity analysis of the design assumptions. For instance - What happens to the power of the study when the δ value is not the same as specified at the design stage?

We will now simulate design **CAPT-GSD-Drp**. Select this design from the library and click the **S** icon. Alternatively, you can right-click this design in the **Library**, and select **Simulate**.

The default view of input window for simulations is as shown below:

The screenshot shows the 'Response Generation' tab of the simulation input window. It includes the following fields and a table:

- Trial Type: Superiority
- Test Type: 1-Sided
- Sample Size (n): 1532
- Specify Variance:
 - Pooled Estimate
 - Unpooled Estimate

| Look # | Info. Fraction | Cum. α Spent | Efficacy Z | Futility Z |
|--------|----------------|---------------------|------------|------------|
| 1 | 0.333 | 0.000 | -3.710 | 0.216 |
| 2 | 0.667 | 0.006 | -2.511 | -0.923 |
| 3 | 1.000 | 0.025 | -1.993 | -1.993 |

Notice that the value of δ on the **Response Generation** tab is -0.05. This corresponds to the difference in proportions under the alternative hypothesis. You may either keep this default value for the simulation or change it if you wish to simulate the study with a different value of δ . Let us run some simulations by changing the value of δ . We will run simulations over a range of values for π_t , say, 0.1, 0.125 and 0.14. Enter the values as shown below:

The 'Specify Proportion' section shows the following values:

- Prop. under Control (π_c): 0.15
- Prop. under Treatment (π_t): .1, .125, 0.14

Before running simulations, let us have a quick look at the **Simulation Control** tab where we can change the number of simulations, save the simulation data in East format or in a csv format and some more useful things. You can manipulate the simulations with the following actions:

- Enter the number of simulations you wish to run in the "Number of Simulations" field. The default is 10000 simulations.
- Increase/ Decrease the "Refresh Frequency" field to speed up or slow down the simulations. The default is to refresh the screen after every 1000 simulations.
- Set the Random Number Seed to **Clock** or **Fixed**. The default is **Clock**.
- Select the checkbox of "Suppress All Intermediate Output" to suppress the intermediate output.

3 Getting Started

To see the intermediate results after a specific number of simulations, select the checkbox of "Pause after Refresh" and enter the refresh frequency accordingly. The checkbox of "Stop At End" is selected by default to display the summary results at the end of all the simulations a corresponding item gets created in the **Output Preview** window. One can uncheck this box and save the simulation node directly in the **Output Preview** window.

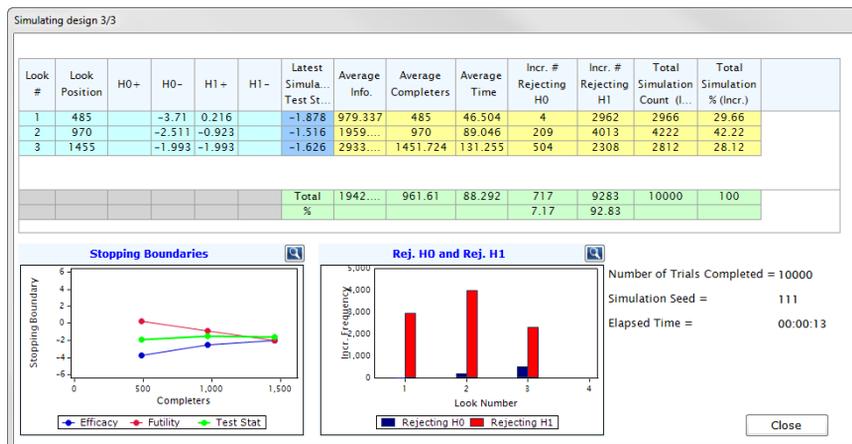
One can also save the summary statistics for each simulation run and the subject level simulated data in the form of a Case Data or a Comma Separated File. Select the checkboxes accordingly and provide the file names and paths while using the CSV option. If you are saving the data as Case Data, the corresponding data file will be associated with the simulation node. It can be accessed by saving the simulation node from **Output Preview** to the workbook in **Library**.

For now, let us keep the **Simulation Control** tab as shown below:

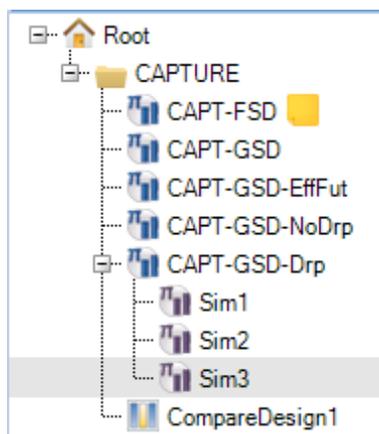
The screenshot shows the 'Simulation Controls' tab of a software interface. It contains several input fields and checkboxes. On the left side, there are fields for 'Number of Simulations' (set to 10000) and 'Refresh Frequency' (set to 1000). Below these is a 'Random Number Seed' section with radio buttons for 'Clock' and 'Fixed' (selected), with a text input field containing '111'. Further down are checkboxes for 'Suppress All Intermediate Output', 'Pause after Refresh', and 'Stop at End' (which is checked). On the right side, there is an 'Output Options' section with a dropdown menu for 'Output Type' set to 'Case Data'. Below this are two checkboxes: 'Save summary statistics for every simulation run' (unchecked) and 'Save subject level data for' (unchecked) with a text input field containing '1'. A note below states 'Note: Max. 100,000 records will be saved.'

Click the **Simulate** button on right hand side bottom to run the simulations. Three scenarios corresponding to three values of π_t are simulated one after the other and in the end, the following output window appears. This is the **Simulation Intermediate Output window** which shows the results from **last simulated scenario**. The two plots

on this window are useful to see how the study performed over 10000 simulations.



Click the **Close** button on this intermediate window which takes us to the **Output Preview** window. Save these three simulation rows in the workbook **CAPTURE**. Since we simulated the design **CAPT-GSD-Drp**, the three simulation nodes get saved as child nodes of this design. This is the hierarchy which is followed throughout the East6 software.



A full picture of the CAPTURE trial design with accrual/dropout information and its simulations can be viewed easily. Select the three simulation nodes and the parent

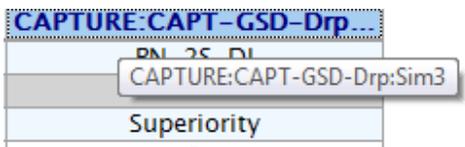
3 Getting Started

design node in the **Library** and click the  icon.

| | CAPTURE:CAPT-GSD-Drp | CAPTURE:CAPT-GSD-Drp... | CAPTURE:CAPT-GSD-Drp... | CAPTURE:CAPT-GSD-Drp... |
|---|----------------------|-------------------------|-------------------------|-------------------------|
| Mnemonic | PN-2S-DI | PN-2S-DI | PN-2S-DI | PN-2S-DI |
| Test Parameters | | | | |
| Design Type | Superiority | Superiority | Superiority | Superiority |
| No. of Looks | 3 | 3 | 3 | 3 |
| Test Type | 1-Sided | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | | | |
| Attained α | 0.023 | | | |
| Power | 0.8 | 0.805 | 0.258 | 0.072 |
| Model Parameters | | | | |
| Allocation Ratio (nt / nc) | 1 | | | |
| Proportion under Control (π_c) | 0.15 | | | |
| Proportion under Treatment (π_t) | 0.1 | | | |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.05 | | | |
| Variance | Unpooled Estimate | | | |
| Proportion under Control Data (π_c) | | 0.15 | 0.15 | 0.15 |
| Proportion under Treatment Data (π_t) | | 0.1 | 0.125 | 0.14 |
| Boundary Parameters | | | | |
| Spacing of Looks | Equal | User Specified | User Specified | User Specified |
| Efficacy Boundary | LD (OF) | User Specified | User Specified | User Specified |
| Futility Boundary | Gm (-2) (NB) | User Specified | User Specified | User Specified |
| Accrual & Dropout Parameters | | | | |
| Accrual Rate | 12 | 12 | 12 | 12 |
| Response Lag | 4 | 4 | 4 | 4 |
| Probability of Dropout | 0.05 | 0.05 | 0.05 | 0.05 |
| No. of Accrual Periods | | 1 | 1 | 1 |
| Sample Size | | | | |
| Maximum | 1532 | 1532 | 1532 | 1532 |
| Expected Under H0 | 933,474 | | | |
| Expected Under H1 | 1262,455 | | | |
| Completers | | | | |
| Maximum | 1455 | 1455 | 1455 | 1455 |
| Expected Under H0 | 848 | | | |
| Expected Under H1 | 1174,624 | | | |
| Study Duration | | | | |
| Maximum | 131.667 | | | |
| Simulation Results (Overall) | | | | |
| Average Study Duration | | 106.835 | 102.925 | 88.292 |
| Average Sample Size | | 1262.016 | 1214.796 | 1047.899 |
| Average Completers | | 1173.081 | 1128.516 | 961.61 |

Note the drop in simulated power as the difference between the two arms decreased. This is because, the sample size of 1532 was insufficient to detect the δ value -0.025 and -0.01. It shows the effect of mis-specifying the alternative hypothesis. It did achieve the power of 80% for the first case with δ equal to -0.05 which was actually the δ at the design stage. This is called simulating the design under *Alternative*. We can also simulate a design under *Null* by entering π_t equal to 0.15, same as π_c and verify that the type I error is preserved.

The column width in the comparison mode is fixed and the heading appears in the format *workbook_name.design_name.Sim*. If this string is longer than the fixed width then you may not be able to see the complete heading. In that case, you can hover the mouse on cell of column heading to see the complete heading.



Thus, simulations in East6 are one of the very powerful tools which help us to verify the operating characteristics of the design.

The next section introduces us to another key feature of East6 - **Interim Monitoring**. Let us see monitor the CAPTURE Trial using this feature.

3.8 Interim Monitoring

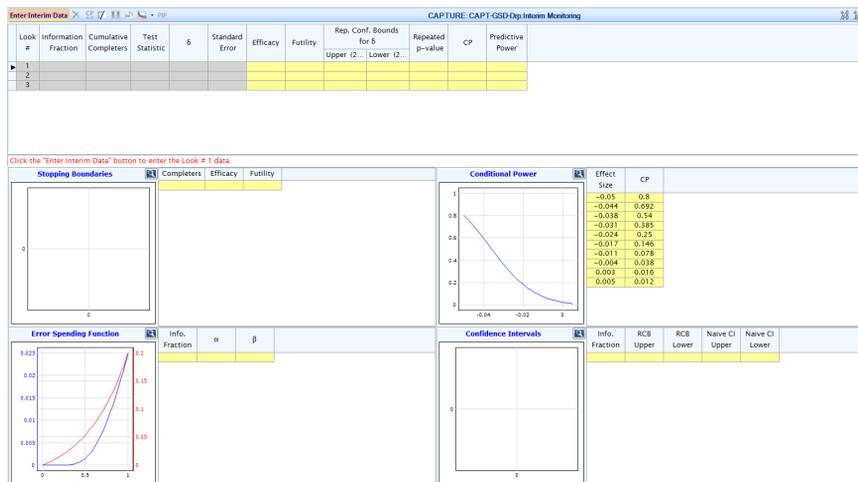
Interim monitoring is critical for the management of group sequential trials, and there are many reasons why flexibility in both design and monitoring is necessary. Administrative schedules may call for the recalculation of statistical information and unplanned analyses at arbitrary time points, while the need for simultaneously preserving both the type-1 error and power of the study must be maintained. East provides the capability to flexibly monitor a group sequential trial using the **Interim Monitoring**.

The **IM dashboard** provides a coherent visual display of many output values based on interim information. In addition to important statistical information, included are tables and graphs for stopping boundaries, conditional power, error spending and confidence intervals for each interim look. All of this information is useful in tracking the progress of a trial for decision making purposes, as well as allowing for improvements to a study design adaptively.

Consider the monitoring of **CAPT-GSD-Drp** of the CAPTURE trial. Select this design from the **Library** and click the  icon. The adaptive version of IM dashboard can be invoked by clicking the  icon. But for this example, we will use regular IM dashboard. A node named Interim Monitoring gets associated with the design in the **Library** and a

3 Getting Started

blank IM dashboard is opened up as shown below:



Suppose we have to take the first look at the data based on 485 completers.

The interim data on these subjects is to be entered in **Test Statistic Calculator** which can be opened by clicking **Enter Interim Data** button. Open this calculator and click OK with default parameters.

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Read from Analysis Node

Select Workbook:

Select Analysis Node:

Input for Binomial end point

Completers and Responses

| | Control | Treatment |
|------------------------|---------|-----------|
| Cumulative Completers: | 242 | 243 |
| Cumulative Response: | 36 | 24 |

Completers (Overall): 485

Input for Binomial end point

Estimate of δ : -0.05

$\delta = (\pi_t - \pi_c)$

Standard Error of Estimate of δ : 0.03

Output

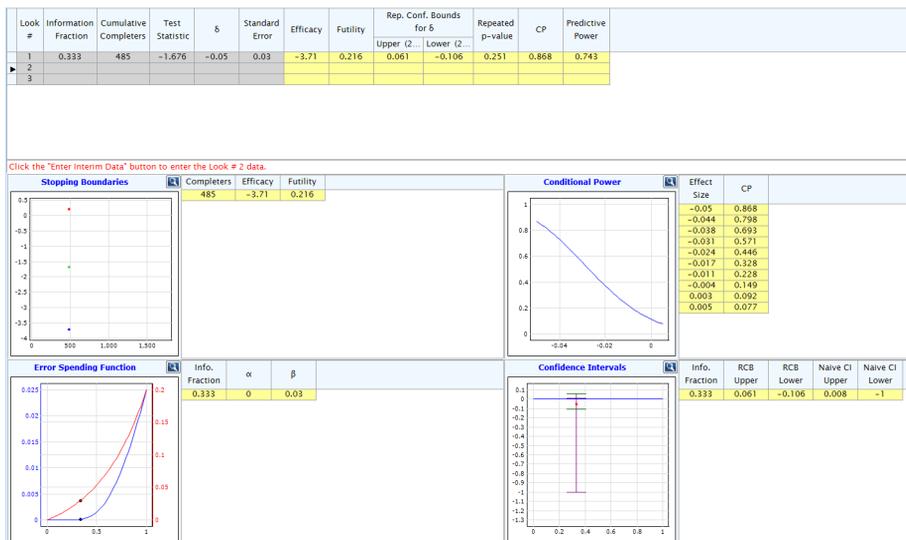
Test Statistic: -1.676

Recalc OK Cancel

If we have run any analysis procedure on the interim data then the test statistic calculator can read in the information from Analysis node. Select the appropriate workbook and the node and hit Recalc to see the interim inputs. Alternatively, for binomial endpoint trials, we can enter the interim data in terms of the number of responses on each arm and East computes the difference in proportion and its standard error. Alternatively, we can directly enter the and its standard error which can be the output of some external computation. The inputs on the test statistic calculator depend upon the type of trial you are monitoring.

3 Getting Started

The resulting screen is as shown below:



The output quantities for the first look are computed in that row and all the four charts are updated based on the look1 data.

There some more advanced features like Conditional Power calculator, Predicted Intervals Plot, Conditional Simulations available from the IM dashboard. These are explained in later sections of this manual.

Let us take the second look at 970 subjects. Open the test statistic calculator and leaving all other parameters default, change the number of responses on Treatment arm

to 30. Click the OK button. The screen will look as shown below:

| Look # | Information Fraction | Cumulative Completers | Test Statistic | δ | Standard Error | Efficacy | Futility | Rep. Conf. Bounds for δ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|-----------------------|----------------|----------|----------------|----------|----------|--------------------------------|--------------|------------------|-------|------------------|
| | | | | | | | | Upper (2...) | Lower (2...) | | | |
| 1 | 0.333 | 485 | -1.676 | -0.05 | 0.03 | -3.71 | 0.216 | 0.061 | -0.106 | 0.251 | 0.868 | 0.743 |
| 2 | 0.667 | 970 | -4.441 | -0.087 | 0.02 | -2.511 | -0.923 | -0.038 | -0.119 | 0 | 1 | 1 |

Boundary Crossed

Since the value of Test Statistic is \leq the critical point for efficacy, H_0 is rejected.

Although boundary has been crossed, East gives you choice either to stop the study or to continue entering further looks. Please make your decision.

stop the study and bar further looks input

allow the study to continue

Click the "Edit Interim Data" button to edit

Stopping Boundaries

| Final Inference | |
|--|--------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0 |
| Adj. Pt. Est. for δ | -0.085 |
| Adj. 95% CI for δ | |
| Upper Confidence Bound | -0.043 |
| Lower Confidence Bound | -0.124 |
| Post-Hoc Power | |

East tells us that the null hypothesis is rejected at second look, provides an option to stop the trial and conclude efficacy of the drug over the control arm. It computes the final inference in the end. At this stage, it also provides another option to continue entering data for future looks. But the final inference is computed only once.

In the last part of this chapter, we shall see how to capture a snapshot of any ongoing interim monitoring of a trial.

The **IM dashboard** can also be used as a tool at design time, where we can construct and analyze multiple possible trial scenarios before actual data is collected. The feature to save a snapshot of information at interim looks can be employed to allow a user the benefit of quickly comparing multiple scenarios under a variety of assumptions. This option increases the flexibility of both, design and interim monitoring process. At each interim look, a snapshot of the updated information in the dashboard can be saved for the current design in the workbook.

Click the  icon located at the top of **IM Dashboard** window to save the current contents of the dashboard:

A new snapshot node is added under the **Interim Monitoring** node in the library. The **Interim Monitoring** window is the input window which can't be printed whereas it

3 Getting Started

snapshot is in the HTML format which can be printed and shared.

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

Snapshot of Look 2 of Interim Monitoring:

| Look # | Information Fraction | Cumulative Completers | Test Statistic | δ | Standard Error | Efficacy | Futility | Rep. Conf. Bounds for δ | | Repeated p-value | INLP | CP | Predictive Power |
|--------|----------------------|-----------------------|----------------|----------|----------------|----------|----------|--------------------------------|--------------|------------------|------|-------|------------------|
| | | | | | | | | Upper (20%) | Lower (2.5%) | | | | |
| 1 | 0.333 | 485 | -1.676 | -0.05 | 0.03 | -3.71 | 0.216 | 0.061 | -0.106 | 0.251 | 1397 | 0.968 | 0.743 |
| 2 | 0.667 | 970 | -4.441 | -0.087 | 0.02 | -2.511 | -0.923 | -0.038 | -0.119 | 1.799E-4 | NA | NA | NA |

Final Inference

| | |
|----------------------------|----------|
| Final Outputs at Look # | 2 |
| Adj. p-Value | 1.067E-4 |
| Adj. Pt. Est. for δ | -0.085 |
| Adj. 95% CI for δ | |
| Upper Confidence Bound | -0.043 |
| Lower Confidence Bound | -0.124 |
| Post-Hoc Power | NA |

To illustrate the benefit of the snapshot feature, it is often the case that actual trial data is not available at design time. Determining a reasonable estimate of nuisance parameters, such as the variance, rather than making strong assumptions of its certainty may be desired. The ability to quickly compare potential results under a variety of different estimates of the variance by easily looking at multiple interim snapshots of a study can be a powerful tool.

Other examples could include sample size re-estimation where initial design assumptions may be incorrect or using hypothetical interim data to compare relevant treatment differences.

With this, we come to an end of the chapter on getting started with East6. The subsequent chapters in this manual discuss in detail with the help of case studies all the features available in the software. The theory part of all the design and analysis procedures is explained in Appendix A of this manual.

4 Data Editor

Data Editor allows you to manipulate the contents of your data. **East** caters to Case Data and Crossover Data. Depending on the type of data, a corresponding set of menu items becomes available in the **Data Editor** menu.

4.1 Case Data

4.1.1 Data Editor

Capabilities for Case Data

4.1.2 Creating Variables

4.1.3 Variable Type Setting

4.1.4 Editing Data

4.1.5 Filter Cases

The **Data editor** window for case data is a spreadsheet-like facility for creating or editing case data files. A case data file is organized as a sequence of **records** called **cases** one below the other. Each record is subdivided into a fixed number of **fields**, called **variables**. The name assigned to that field is referred to as the **variable name**. Each such name identifies a specific variable across all the cases. Each cell holds a value of a variable for a case. The top line of the Data editor holds the **variable names**. Case data is the most common format to enter and store data. If you plan to share data with any other package you need to use case data editor.

4.1.1 Data Editor Capabilities for Case Data

The Data Editor is used to create a new Case Data file or to edit one that was previously saved. You can:

- Create new variables
- Change names and attributes of existing variables
- Alter the column width
- Alter the row height
- Type in new case data records
- Edit existing case data records
- Insert new variables into the data set
- Remove variables from the data set
- Select or reject subsets of the data
- Transform variables
- List data in the log window
- Calculate summary measures from the variables

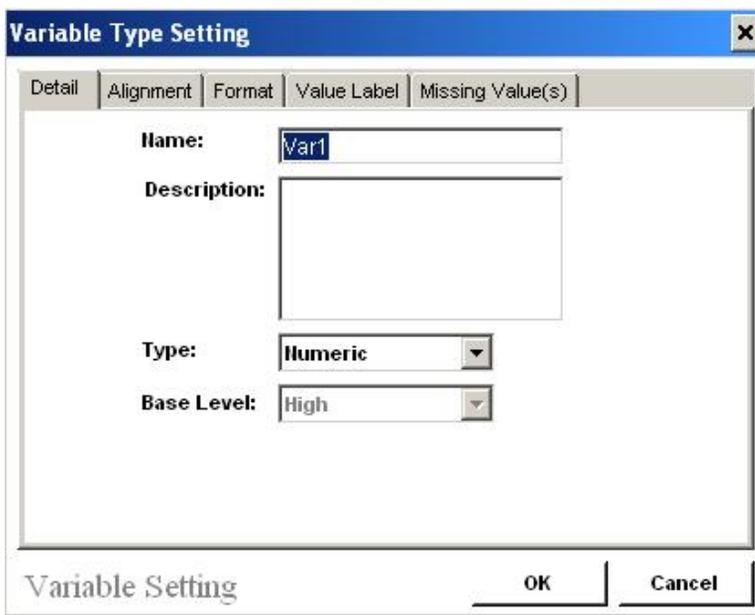
4.1.2 Creating Variables

To create a new Case Data set, invoke the menu **Home**. Click on the icon . Select **Case Data**. When you create a new case data set, all the columns are labeled **var**, indicating that new variables may be created in any of the columns. To create a new variable simply start entering data in a blank column. The column is given a default name Var1, Var2, etc. Alternatively, select any unused column, right click and select **Create Variable** from the menu that appears. The data editor will create all the variables with default names up to the column you are working on. To create a new

4 Data Editor

variable in the first unused column and to select its attributes, choose menu **Data Editor**. Click on the icon .

You will be presented with the dialog box shown below, in which you can select the variable name, variable type, alignment, format, value labels and missing values.



4.1.3 Variable Type Setting

You can change the default variable name and its type in this dialog box and click on the **OK** button. **East** will automatically add this new variable to the case data file. New variables are added immediately adjacent to the last existing variable in the case data set.

The **Variable Type Setting** dialog box contains five tabs: **Detail**, **Alignment**, **Format**, **Value Label**, and **Missing Value(s)**.

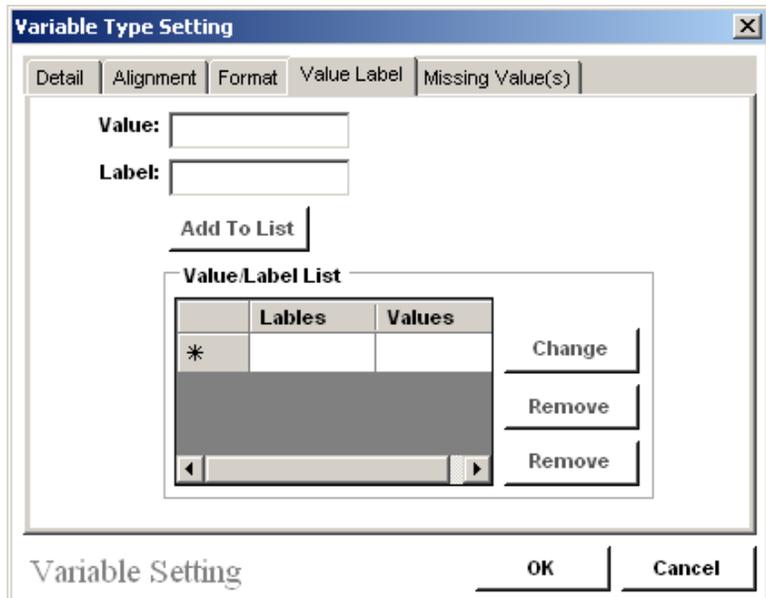
Detail

The **Detail** tab allows you to change the default variable name, add a description of the variable and select the type (**Numeric**, **String**, **Date**, **Binary**, **Categorical** or **Integer**). Note that depending on the **type** of the variable, different tabs and options become available in

the **Variable Type Settings**. For example the tab **Category Details** and option **Base Level** become available only if you select Variable **Type** as **Categorical**.

Value Label

The **Value label** tab is displayed below. Here, you can add labels for particular data values, change a selected label, or remove a selected label, or remove all value labels for the current variable.



Missing Value(s)

The **Missing Value (s)** tab is used for specifying which values are to be treated as missing. You have three choices: **Not Defined**, which means that no values will be treated as missing values; **Discrete value (s)**, which allows you to add particular values to the list of missing values; or **Range**, which lets you specify an entire range of numbers as missing values.

4.1.4 Editing Data

Besides changing the actual cell entries of a case data set you can:

- Add new Cases and Variables
- Insert or delete Cases and Variables

4 Data Editor

- Sort Cases

4.1.5 Filter Cases

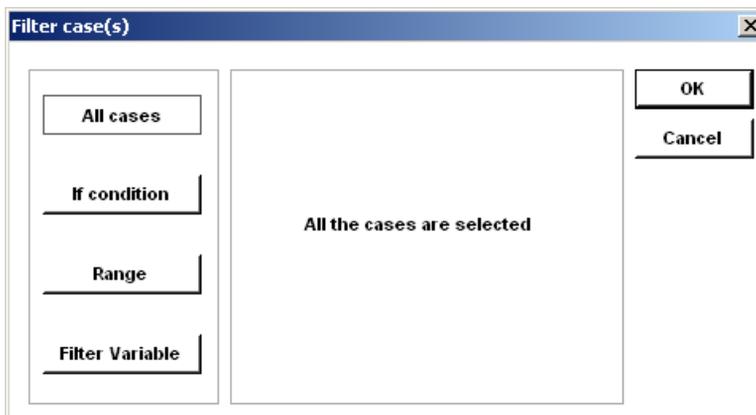
We illustrate the ability of **East** to filter cases with the help of the following example:

Step 1: Open the Data set

Open the data set **leukemia.cyd** by clicking on menu **Home**. Click on the icon  **Open**. Select **Data**. The data is stored in the **Samples** folder of the installation directory of **East**.

Step 2: Invoke the Filter Cases menu

Invoke the menu item **Data Editor**. Click on the icon  **Filter cases**. **East** will present you with a dialog box that allows you to use subsets of data in the Case Data editor. The dialog box will allow you to select **All cases**, those satisfying an **If condition**, falling in a **Range**, or using a **Filter Variable** as shown below.

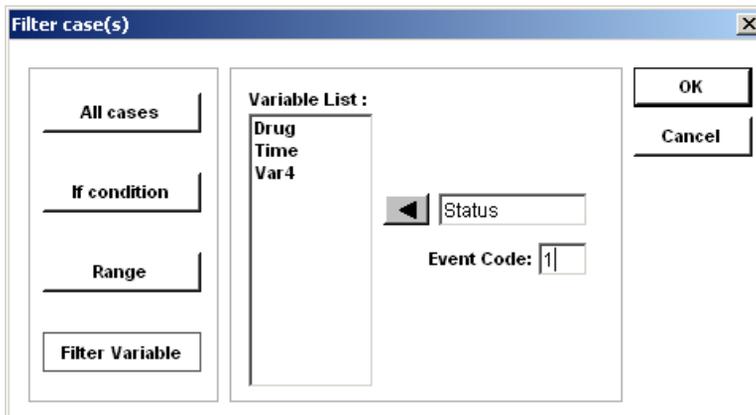


Step 3: Filter Variable option

Select the **Filter Variable** option. Select **Status** from the variable list and click on the black triangle, which will remove the variable **Status** from the variable list and add it to the empty box on the other side.

Suppose we want to filter the cases for which the **Status** variable has value 1. Insert

1 in the empty box next to **Event code**.



Step 4: Output

Click on **OK**. As shown in the following screenshot, **East** will grey out all the cases that have **Status** variable value 1. Now any analysis carried out on the data set uses only the filtered cases. In this way you, can carry out subgroup analyses if the

4 Data Editor

subgroups are identified by the values of a variable in the data set.

| Drug: 1 Value: Placebo | | | |
|------------------------|---------|------|--------|
| | Drug | Time | Status |
| 1 | Placebo | 1 | 1 |
| 2 | 6-MP | 10 | 1 |
| 3 | Placebo | 22 | 1 |
| 4 | 6-MP | 7 | 1 |
| 5 | Placebo | 3 | 1 |
| 6 | 6-MP | 32 | 0 |
| 7 | Placebo | 12 | 1 |
| 8 | 6-MP | 23 | 1 |
| 9 | Placebo | 8 | 1 |
| 10 | 6-MP | 22 | 1 |
| 11 | Placebo | 17 | 1 |
| 12 | 6-MP | 6 | 1 |
| 13 | Placebo | 2 | 1 |
| 14 | 6-MP | 16 | 1 |
| 15 | Placebo | 11 | 1 |
| 16 | 6-MP | 34 | 0 |
| 17 | Placebo | 8 | 1 |
| 18 | 6-MP | 32 | 0 |
| 19 | Placebo | 12 | 1 |
| 20 | 6-MP | 25 | 0 |
| 21 | Placebo | 2 | 1 |
| 22 | 6-MP | 11 | 0 |
| 23 | Placebo | 5 | 1 |
| 24 | 6-MP | 20 | 0 |
| 25 | Placebo | 4 | 1 |
| 26 | 6-MP | 19 | 0 |
| 27 | Placebo | 15 | 1 |
| 28 | 6-MP | 6 | 1 |
| 29 | Placebo | 8 | 1 |
| 30 | 6-MP | 17 | 0 |
| 31 | Placebo | 23 | 1 |
| 32 | 6-MP | 35 | 0 |
| 33 | Placebo | 5 | 1 |

4.2 Crossover Data

4.2.1 Data Editor

Capabilities for Crossover Data

4.2.2 Creating a New

Crossover Data Set

The **Data Editor** allows you to enter data for a 2×2 crossover trial with one record for each patient. You can use this crossover data editor to input individual patients' responses in 2×2 crossover trials. The response could be continuous (such as systolic blood pressure) or binary (such as the development of a tumor after injecting a carcinogenic agent). Only the continuous response type is currently supported in **East**.

4.2.1 Data Editor Capabilities for Crossover Data

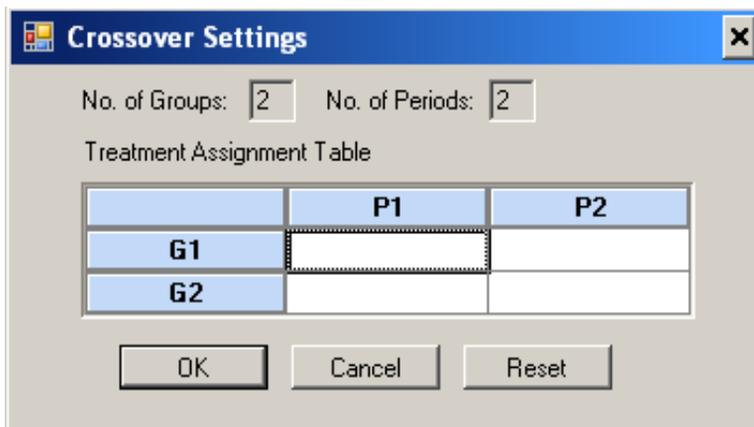
The Data Editor is used to create a new 2×2 Crossover Data file or to edit one that was previously saved. You can:

- Create and edit data with continuous response of individual patients.
- Edit period labels.
- Assign treatments to different groups and periods.
- Convert to case data.
- Convert case data into crossover data.
- List data to the log

4.2.2 Creating a New Crossover Data Set

To create a new crossover data set, invoke the menu **Home**. Click on icon  and from the drop down menu choose **Crossover data**.

You will be presented with a dialog box as shown below:



In the above dialog box, you see a 2×2 grid called **Treatment Assignment Table**. This grid is provided to assign the treatments to different groups and periods.

4 Data Editor

In this version of the software, you can analyze data for 2×2 crossover trials. Hence the number of groups and number of periods are always two. The rows specify the two groups labeled as **G1** and **G2**. The columns represent two periods of the crossover data labeled "P1" and "P2". If you'd like to change these labels, click inside the table cells. Type the treatment names associated with the corresponding group and period. Having entered the treatments, the crossover data editor settings dialog box will look as follows:

The screenshot shows a dialog box titled "Crossover Settings". At the top, there are two input fields: "No. of Groups:" with the value "2" and "No. of Periods:" with the value "2". Below these is a section titled "Treatment Assignment Table" containing a table with the following structure:

| | P1 | P2 |
|----|----|----|
| G1 | A | B |
| G2 | B | A |

At the bottom of the dialog box are three buttons: "OK", "Cancel", and "Reset".

Rules for editing these fields The row names **G1** and **G2** can be changed using a string consisting of a maximum of 8 characters from the set A-Z, 0-9, '.', '_' (underscore), starting with either a letter or a digit; blank spaces are not accepted as part of a name. The column names **P1** and **P2** can be changed the same way. Also note that the Group names as well as the Period names must be distinct. The letters are *not* case sensitive. Once you have assigned all the treatments, click on the button [OK](#).

This will open up the Patients' crossover data editor.

No. of Groups:

No. of Periods:

Treatment Assignment Table

| | | |
|----|----|----|
| | P1 | P2 |
| G1 | A | B |
| G2 | B | A |

| | Patient_Id | Group_Id | P1_Resp | P2_Resp | var |
|---|---|----------|---------|---------|-----|
| 1 | <input style="width: 100%;" type="text"/> | | | | |
| 2 | | | | | |
| 3 | | | | | |

This editor resembles the case data editor. Like the case data editor, this is a spreadsheet into which you can enter data directly. There are four pre-defined fields in this editor. The **PatientId** column must contain the Patients' identification number. The **GroupId** column will contain the group identification to which the patient belongs. The entry in this column should be one of the labels that you have entered as row names in the 2 × 2 grid earlier. The inputs in the next two columns are numeric and contain the responses of the patient in two periods respectively. The title of the next two columns is created by concatenating the word "Resp" to the period identifications that you have entered previously. For example, here in the setting dialog we have entered P1 and P2 as period identifiers and these two response columns are labeled as P1_Resp and P2_Resp. However, if the period values are starting with digits such as 1 and 2, then the period ids are prefixed by the letter P, and the heading of the next two columns would be P1_Resp and P2_Resp.

The variable names **PatientId**, **GroupId**, are fixed and cannot be edited in the data editor. If you use **Transform Variable** on **Group Id** and the result is either "G1" or "G2," then the value is displayed; otherwise, the value is shown as missing. You can also add covariates such as age and sex. All variable settings of the case data editor are applicable to these covariates. The button allows you to edit the **GroupId**, **PeriodId** or the treatment labels that you have edited earlier. If you make any changes, these changes will automatically be made in the data editor.

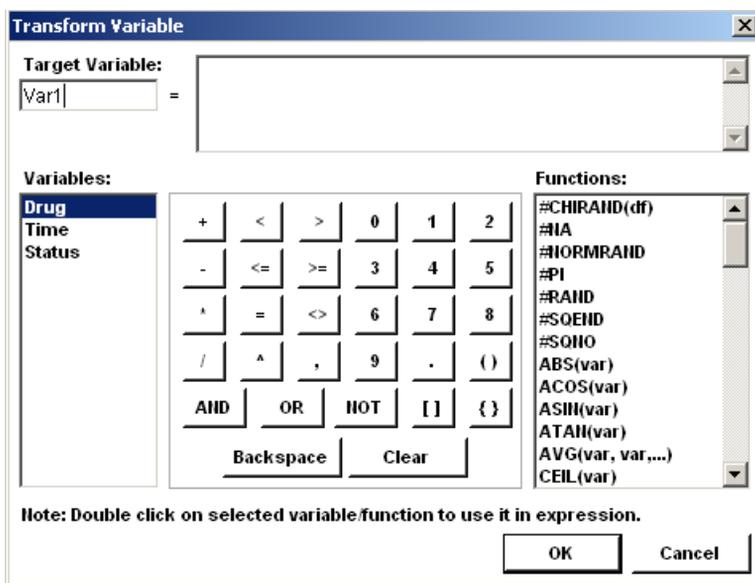
4.3 Data Transformation

You can transform an existing variable with the data transformation facility available in the **Data Editor** of **East** .

4 Data Editor

To transform any variable:

1. Select the menu **Data Editor**. Click on the icon . You will be presented with the expression builder dialog box screen. Here you can transform the values of the current variable using a combination of statistical, arithmetic, and logical operations.



The current variable name is the target variable on the left hand side of an equation with the form:

VAR =

Where, **VAR** is the variable name of the current variable. In order to create a new variable, type the variable name in the **target variable** field.

2. Complete the right hand side of the equation with any combination of allowable functions. To select a function, double-click on it. If the function that you select needs any extra parameters (typically variable names), this will be indicated by a **?** for each required parameter. Replace the **?** character with the desired parameter.
3. Select the **OK** button to fill in values for the current variable computed according to the expression that you have constructed.

The statistical, arithmetical, and logical functions that are available in the **Transform Variable** dialog box are given below:

4.4 Mathematical and Statistical Functions

The following is a list of mathematical and statistical functions available in **East** used for variable transformation.

ABS(X) Returns the absolute value of X.

Argument Range: $-1 \times 10^{25} \leq X \leq 1 \times 10^{25}$.

ACOS(X) Returns the arccosine of X.

Argument Range: $-1 \leq X \leq 1$.

ASIN(X) Returns the arcsine of X.

Argument Range: $-1 \leq X \leq 1$.

ATAN(X) Returns the arctangent of X.

Argument Range: $-1 \times 10^{25} \leq X \leq 1 \times 10^{25}$.

AVG(X_1, X_2, \dots) Returns the mean of (X_1, X_2, \dots).

Argument Range: $-1 \times 10^{25} \leq X \leq 1 \times 10^{25}$.

CEIL(X) Returns the ceiling, or smallest integer greater than or equal to X.

Argument Range: $-1 \times 10^{25} \leq X \leq 1 \times 10^{25}$.

CHIDIST(X,df) Returns the probability in the tail area to the left of X from the chi-squared distribution with $df > 0$ degrees of freedom.

Argument Range: $0 \leq X \leq 1 \times 10^{25}$.

CHIINV(X,df) Returns the Xth percentile value of the chi-squared distribution with $d > 0$ degrees of freedom, i.e., returns z such that $\Pr(Z \leq z) = X$.

Argument Range: $0.0001 \leq X \leq 0.9999$.

COS(X) Returns the cosine of X, where X is expressed in radians.

Argument Range: $-2.14 \times 10^9 \leq X \leq 2.14 \times 10^9$.

COSH(X) Returns the hyperbolic cosine of X.

Argument Range: $-87 \leq X \leq 87$.

CUMULATIVE(X) Given a column of X values this function returns a new column in which the entry in row j is the sum of entries in the first j rows of the original column.

EXP(X) Returns the exponential function evaluated at X.

Argument Range: $-87 \leq X \leq 87$.

FLOOR(X) Returns the floor, or largest integer less than or equal to X.

Argument Range: $-1 \times 10^{25} \leq X \leq 1 \times 10^{25}$.

INT(X) Returns the integer part of X.

Argument Range: $-1 \times 10^{25} \leq X \leq 1 \times 10^{25}$.

ISNA(X) Returns a value of 1 if X is a missing value 0 otherwise.

Argument Range: $-1 \times 10^{25} \leq X \leq 1 \times 10^{25}$. This function is useful. For example, set missing observations to average values.

X1 = IF (ISNA (X)=1, COLMEAN (X), X)

Another extremely useful task performed by the ISNA() function is to eliminate records from the data set in which there are missing values.

4 Data Editor

REJECTIF (ISNA (X) =1)

SELECTIF (ISNA (V1) +ISNA (V2) +ISNA (V3) =0)

LOG(X) Returns the logarithm of X to base 10.

Argument Range: $1 \times 10^{-25} \leq X \leq 1 \times 10^{25}$.

LN(X) Returns the logarithm of X to base e .

Argument Range: $1 \times 10^{-25} \leq X \leq 1 \times 10^{25}$.

MAX(X₁, X₂, ...) Returns the maximum value of (X₁, X₂, ...).

MIN(X₁, X₂, ...) Returns the minimum value of (X₁, X₂, ...).

MOD(X,Y) Returns the remainder of X divided by Y. The sign of this remainder is the same as that of X.

Argument Range: $-1 \times 10^{25} \leq X \leq 1 \times 10^{25}$.

NORMDIST(X) Returns the probability in the tail area to the left of X from the standardized normal distribution.

Argument Range: $-10 \leq X \leq 10$.

NORMINV(X) Returns the Xth percentile value of the standard normal distribution, i.e., returns z such that $\Pr(Z \leq z) = X$.

Argument Range: $0.001 \leq X \leq 0.999$.

ROUND(X,d) Returns a floating point number obtained by rounding X to d decimal digits. If d=0, X is rounded to the nearest integer.

Argument Range: $-1 \times 10^{25} \leq X \leq 1 \times 10^{25}$.

SIN(X) Returns the sine of X, where X is expressed in radians.

Argument Range: $-2.14 \times 10^9 \leq X \leq 2.14 \times 10^9$.

SINH(X) Returns the hyperbolic sine of X.

Argument Range: $-87 \leq X \leq 87$.

SQRT(X) Returns the square root of X.

Argument Range: $0 \leq X \leq 1 \times 10^{25}$.

TAN(X) Returns the tangent of X, where X is expressed in radians.

Argument Range: $-1 \times 10^{25} \leq X \leq 1 \times 10^{25}$; $X \neq (2n + 1)\frac{\pi}{2}$, n an integer.

TANH(X) Returns the hyperbolic tangent of X.

Argument Range: $-87 \leq X \leq 87$.

4.4.1 The IF Function

This function tests arithmetic or logical condition and returns one value if true, another value if false. The syntax is

IF (CONDITION, X, Y)

The function returns the value **X** if **CONDITION** is "true" and **Y** if **CONDITION** is "false". For example consider the following equation:

HIVPOS = IF (CD4>1, 1, -1)

The above equation defines a variable **HIVPOS** that assumes the value 1, if the variable **CD4** exceeds 1 and assumes the value -1 otherwise. Usually **CONDITION** is made up of two arithmetic expressions separated by a "comparison operator", e.g., **CD4>CD8**, **CD4+CD8=15*BLOOD**, etc. The following comparison operators are allowed:

=, >, <, >=, <=, <>

More generally, **CONDITION** can be constructed by combining two or more individual conditions with **AND**, **OR**, or **NOT** operators. For example consider the following expression

HIVPOS = IF ((CD4>1) !AND! (CD8>1), 1, -1)

The above expression means that **HIVPOS** will take on the value 1 if **both CD4>1** and **CD8>1**, and -1 otherwise. On the other hand consider the following expression:

HIVANY = IF ((CD4>1) !OR! (CD8>1), 1, -1)

The above expression means that **HIVANY** will take on the value 1 if **either CD4>1** or **CD8>1** and -1 otherwise.

4.4.2 The *SELECTIF* Function

This function provides a powerful way of selecting only those records that satisfy a specific arithmetic or logical condition. All other records are deleted from the current data set. The syntax is:

SELECTIF (CONDITION)

This function selects only those records for which **CONDITION** is "true" and excludes all other records from the current dataset. For example consider the following equation:

HIVPOS = SELECTIF (CD4>1)

The above condition retains records for which **CD4** exceeds 1. The same rules governing **CONDITION** for the **IF** function are applicable here as well.

Note that the column location of the cursor when **Transform Variable** was selected plays no role in the execution of this function.

4.4.3 The *RECODE* Function

This function recodes different ranges of a variable. It is extremely useful for creating a new variable consisting of discrete categories at pre-specified cut-points of the original variable. The syntax for **RECODE** has two forms — one for recoding a

4 Data Editor

categorical variable and one for recoding a continuous variable. In both cases, the variable being recoded must assume numerical values.

Recoding a Categorical Variable syntax is:

$$\mathbf{RECODE}(X, S_1 = c_1, S_2 = c_2, \dots, S_n = c_n, [\mathbf{else}]),$$

where X is the categorical variable (or arithmetic expression) being recoded, S_j represents a set of numbers in X , all being recoded to c_j , and the optional argument **[else]** is a default number to which all the numbers belonging to X , but excluded from the sets S_1, S_2, \dots, S_n , are recoded. If **[else]** is not specified as an argument of **RECODE**, then all the numbers excluded from the sets S_1, S_2, \dots, S_n are unchanged.

Notice that the argument $S_j = c_j$ in the **RECODE** function consists of a set of numbers S_j being recoded to a single number c_j . The usual mathematical convention is adopted of specifying a set of numbers within braces. Thus if set S_j consisted of m distinct numbers $s_{1j}, s_{2j}, \dots, s_{mj}$, it would be represented in the **RECODE** argument list as $\{s_{1j}, s_{2j}, \dots, s_{mj}\}$. For example

$$\mathbf{Y} = \mathbf{RECODE}(\mathbf{X}, \{1, 2, 3\}=1, \{7, 9\}=2, \{10\}=3)$$

will recode the categorical variable X into another categorical variable Y that assumes the value 1 for $X \in \{1, 2, 3\}$, 2 for $X \in \{7, 9\}$, and 3 for $X = 10$.

Other values of X , if any, remain unchanged. If you want those other values of X to be recoded to, e.g., -1, simply augment the argument list by including -1 at the end of the recode statement:

$$\mathbf{Y} = \mathbf{RECODE}(\mathbf{X}, \{1, 2, 3\}=1, \{7, 9\}=2, \{10\}=3, -1) .$$

Recoding a Continuous Variable syntax is:

$$\mathbf{RECODE}(X, I_1 = c_1, I_2 = c_2, \dots, I_n = c_n, [\mathbf{else}])$$

where X is the continuous variable (or arithmetic expression) being recoded, I_j represents an interval of numbers all being recoded to c_j , and the optional argument **[else]** is a default number to which all the numbers belonging to X , but excluded from the intervals I_1, I_2, \dots, I_n , are recoded. If **[else]** is not specified as an argument of **RECODE**, then all the numbers excluded from the intervals I_1, I_2, \dots, I_n are unchanged. Notice that the arguments of **RECODE** are intervals being recoded to individual numbers. The usual mathematical convention for specifying an interval I_j as open, semi-open, and closed is adopted.

Thus:

- An interval I_j of the form (u, v) is open and includes all numbers between u and v , but not the end points.

- An interval I_j of the form $[u, v]$ is closed and includes all numbers between u and v inclusive of the end points.
- An interval of the form $(u, v]$ is open on the left but closed on the right. It excludes u , includes v , and includes all the numbers in between.
- An interval of the form $[u, v)$ is closed on the left but open on the right. It includes u , excludes v , and includes all the numbers in between.

For example

```
Y = RECODE (X, (2.5, 5.7]=1, (5.7, 10.4]=2)
```

will recode the continuous variable X so that all numbers $2.5 < X \leq 5.7$ are replaced by 1, all numbers $5.7 < X \leq 10.4$ are replaced by 2, and all other values of X are unchanged. If you want all other values of X to also be recoded to say -1, append the -1 as the last argument of the equation:

```
Y = RECODE (X, (2.5, 5.7]=1, (5.7, 10.4]=2, -1) .
```

4.4.4 Column Functions

Column functions operate on an entire column of numbers and return a scalar quantity. The returned value is often used in arithmetic expressions. The following column functions are available. All of them are prefixed by the letters **COL**. The argument **var** of all these column functions must be a variable in the worksheet; arithmetic expressions are not permitted. This may require you to create an intermediate column of computed expressions before using a column function. Also note that missing values are ignored in computing these column functions.

COLMEAN(X) Returns the sample mean of X .

COLVAR(X) Returns the sample variance of X .

COLSTD(X) Returns the sample standard deviation of X .

COLSUM(X) Returns the sum of all the numbers in X .

COLMAX(X) Returns the maximum value of X .

COLMIN(X) Returns the minimum value of X .

COLRANGE(X) Returns the value of **COLMAX(X) - COLMIN(X)**.

COLCOUNT(X) Returns the number of elements in X .

You can use the values returned by these column functions in arithmetic expressions and as arguments of other functions. To do this, it is not necessary to know the actual value returned by the column function. However, if you want to know the value returned by any column function, you must define a new variable in the worksheet and fill its entire column with the value of the column function.

4.4.5 Random Numbers

4 Data Editor

You can fill an entire column of a worksheet with random numbers and constants. Suppose the cursor is in a cell of a variable named **RANDNUM**.

The expression

$$\mathbf{RANDNUM} = \mathbf{\#RAND}$$

will result in the variable **RANDNUM** being filled with a column of uniform random numbers in the range (0, 1).

Three random number functions or generators are available to you with the editors:

#RAND Generates uniform random numbers in the range (0, 1).

#NORMRAND Generates random numbers from the standard Normal Distribution.

#CHIRAND(X) Generates random numbers from the chi-squared distribution with X degrees of freedom.

You may of course use these three random number generators to generate random numbers from other distributions. For example, the equation

$$\mathbf{Y} = \mathbf{3+2*\#NORMRAND}$$

will generate random numbers from the normal distribution with mean 3 and standard deviation 2, in variable **Y**. Again, the equation

$$\mathbf{Z} = \mathbf{\#CHIRAND(5)}$$

will generate random numbers from the chi-squared distribution with 5 degrees of freedom.

4.4.6 Special functions

The following special functions are available for use in arithmetic expressions:

#PI This is the value of π .

#NA This is the missing value code. It can be used to detect if a value is missing, or to force a value to be treated as missing.

#SQNO This is the value of the current sequence number (**SQNO**) in the current data set.

#SQEND This is the largest value of the sequence number (**SQNO**) in the current data set.

Volume 2 *Continuous Endpoints*

| | |
|--|------------|
| <i>5 Introduction to Volume 2</i> | <i>73</i> |
| <i>6 Tutorial: Normal Endpoint</i> | <i>79</i> |
| <i>7 Normal Superiority One-Sample</i> | <i>91</i> |
| <i>8 Normal Noninferiority Paired-Sample</i> | <i>113</i> |
| <i>9 Normal Equivalence Paired-Sample</i> | <i>128</i> |
| <i>10 Normal Superiority Two-Sample</i> | <i>141</i> |
| <i>11 Nonparametric Superiority Two Sample</i> | <i>179</i> |
| <i>12 Normal Non-inferiority Two-Sample</i> | <i>185</i> |
| <i>13 Normal Equivalence Two-Sample</i> | <i>211</i> |
| <i>14 Normal: Many Means</i> | <i>232</i> |
| <i>15 Multiple Comparison Procedures for Continuous Data</i> | <i>240</i> |
| <i>16 Multiple Endpoints-Gatekeeping Procedures</i> | <i>265</i> |

| | |
|---|------------|
| <i>17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs</i> | 285 |
| <i>18 Two-Stage Multi-arm Designs using p-value combination</i> | 309 |
| <i>19 Normal Superiority Regression</i> | 332 |

5 *Introduction to Volume 2*

This volume describes the procedures for continuous endpoints (normal) applicable to one-sample, two-samples, many-samples and regression situations. All the three type of designs - superiority, non-inferiority and equivalence are discussed in detail.

Chapter 6 introduces you to East on the Architect platform, using an example clinical trial to test difference of means.

Chapter 7, 8 and 9 detail the design and interim monitoring in one-sample situation where it may be required to compare a new treatment to a well-established control, using a single sample. These chapters respectively cover superiority, non-inferiority and equivalence type of trials.

Chapter 10 details the design and interim monitoring in superiority two-sample situation where the superiority of a new treatment over the control treatment is tested comparing the group-dependent means of the outcome variables.

Chapter 11 details the design in the Wilcoxon-Mann-Whitney nonparametric test which is a commonly used test for the comparison of two distributions when the observations cannot be assumed to come from normal distributions. It is used when the distributions differ only in a location parameter and is especially useful when the distributions are not symmetric. For Wilcoxon-Mann-Whitney test, East supports single look superiority designs only.

Chapter 12 provides an account of the design and interim monitoring in non-inferiority two-sample situation where the goal is to establish that an experimental treatment is no worse than the standard treatment, rather than attempting to establish that it is superior. Non-inferiority trials are designed by specifying a non-inferiority margin. The amount by which the mean response on the experimental arm is worse than the mean response on the control arm must fall within this margin in order for the claim of non-inferiority to be sustained.

Chapter 13 narrates the details of the design and interim monitoring in equivalence two-sample situation where the goal is neither establishing superiority nor non-inferiority, but equivalence. When the goal is to show that two treatments are similar, it is necessary to develop procedures with the goal of establishing equivalence in mind. In Section 13.1, the problem of establishing the equivalence with respect to the difference of the means of two normal distributions using a parallel-group design is presented. The corresponding problem of establishing the equivalence with respect to

5 Introduction to Volume 2

the log ratio of means is presented in Section 13.2. For the crossover design, the problem of establishing the equivalence with respect to the difference of the means is presented in Section 13.3 and with respect to the log ratio of means in Section 13.4.

Chapter 16 details the clinical trials that are often designed to assess benefits of a new treatment compared to a control treatment with respect to multiple clinical endpoints which are divided into hierarchically ordered families. It discusses two methods - Section 16.2 discusses Serial Gatekeeping whereas section 16.3 discusses Parallel Gatekeeping.

Chapter 14 details the various tests available for comparing more than two continuous means in East. Sections 14.1, 14.2 and 14.3 discuss One Way ANOVA, One Way Repeated Measures ANOVA and Two Way ANOVA respectively.

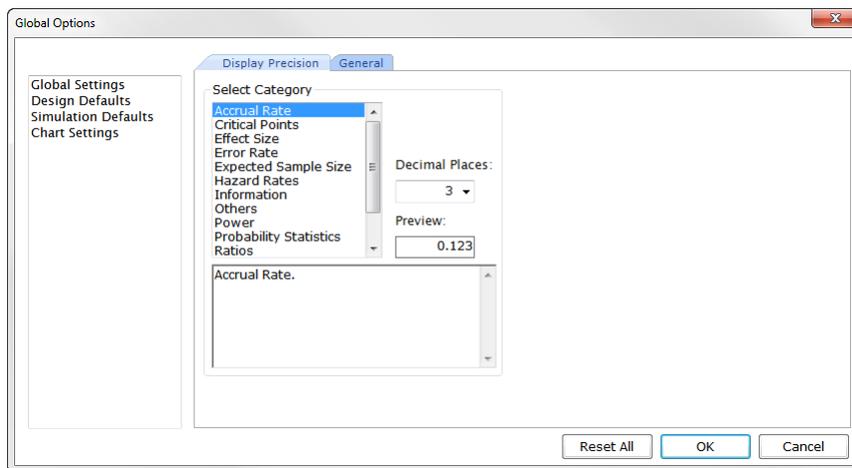
Chapter 15 details the Multiple Comparison Procedures (MCP) for continuous data. It is often the case that multiple objectives are to be addressed in one single trial. These objectives are formulated into a family of hypotheses. Multiple comparison (MC) procedures provides a guard against inflation of type I error while testing these multiple hypotheses. East supports several parametric and p-value based MC procedures. This chapter explains how to design a study using a chosen MC procedure that strongly maintains FWER.

Chapter 19 elaborates on the design and interim monitoring in superiority regression situation where linear regression models are used to examine the relationship between a response variable and one or more explanatory variables. This chapter discusses the design and interim monitoring of three types of linear regression models. Section 19.1 examines the problem of testing a single slope in a simple linear regression model involving one continuous covariate. Section 19.2 examines the problem of testing the equality of two slopes in a linear regression model with only one observation per subject. Finally Section 19.3 examines the problem of testing the equality of two slopes in a linear regression repeated measures model, applied to a longitudinal setting.

5.1 Settings



Click the **Global Options** icon in the **Home** menu to adjust default values in East 6.



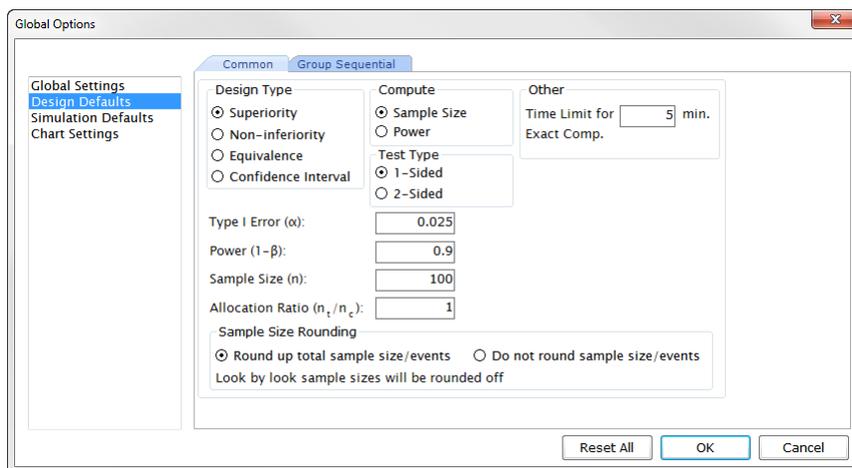
The options provided in the **Display Precision** tab are used to set the decimal places of numerical quantities. The settings indicated here will be applicable to all tests in East 6 under the **Design** and **Analysis** menus.

All these numerical quantities are grouped in different categories depending upon their usage. For example, all the average and expected sample sizes computed at simulation or design stage are grouped together under the category "Expected Sample Size". So to view any of these quantities with greater or lesser precision, select the corresponding category and change the decimal places to any value between 0 to 9.

The **General** tab has the provision of adjusting the paths for storing workbooks, files, and temporary files. These paths will remain throughout the current and future sessions even after East is closed. This is the place where we need to specify the installation directory of the R software in order to use the feature of R Integration in East 6.

5 Introduction to Volume 2

The **Design Defaults** is where the user can change the settings for trial design:

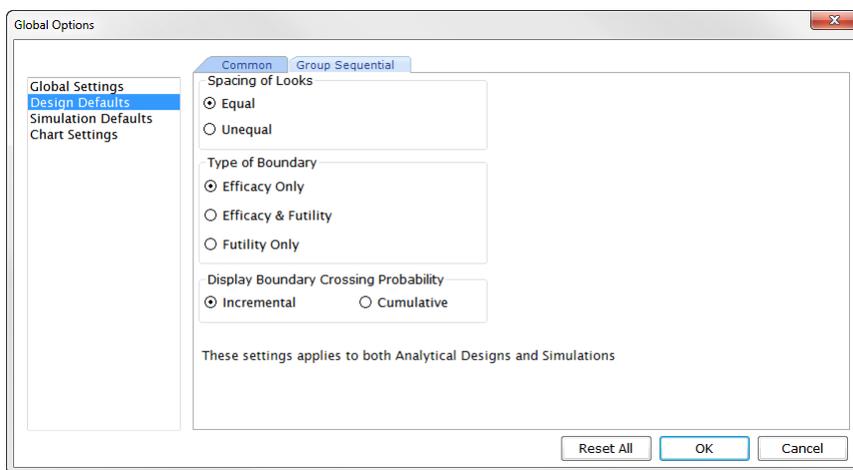


Under the **Common** tab, default values can be set for input design parameters.

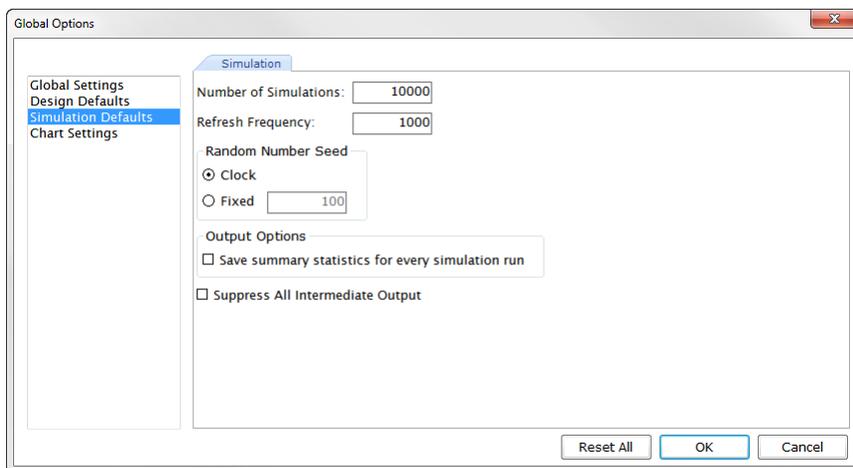
You can set up the default choices for the design type, computation type, test type and the default values for type-I error, power, sample size and allocation ratio. When a new design is invoked, the input window will show these default choices.

- **Time Limit for Exact Computation**
 This time limit is applicable only to exact designs and charts. Exact methods are computationally intensive and can easily consume several hours of computation time if the likely sample sizes are very large. You can set the maximum time available for any exact test in terms of minutes. If the time limit is reached, the test is terminated and no exact results are provided. Minimum and default value is 5 minutes.
- **Type I Error for MCP**
 If user has selected 2-sided test as default in global settings, then any MCP will use half of the alpha from settings as default since MCP is always a 1-sided test.
- **Sample Size Rounding**
 Notice that by default, East displays the integer sample size (events) by rounding up the actual number computed by the East algorithm. In this case, the look-by-look sample size is rounded off to the nearest integer. One can also see the original floating point sample size by selecting the option "Do not round sample size/events".

Under the **Group Sequential** tab, defaults are set for boundary information. When a new design is invoked, input fields will contain these specified defaults. We can also set the option to view the Boundary Crossing Probabilities in the detailed output. It can be either Incremental or Cumulative.



Simulation Defaults is where we can change the settings for simulation:



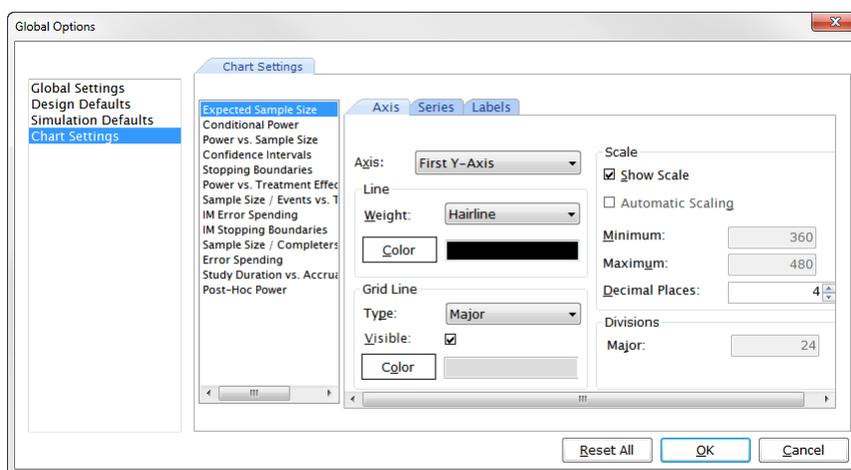
If the checkbox for "Save summary statistics for every simulation" is checked, then East simulations will by default save the per simulation summary data for all the

5 Introduction to Volume 2

simulations in the form of a case data.

If the checkbox for "Suppress All Intermediate Output" is checked, the intermediate simulation output window will be always suppressed and you will be directed to the **Output Preview** area.

The **Chart Settings** allows defaults to be set for the following quantities on East6 charts:



We suggest that you do not alter the defaults until you are quite familiar with the software.

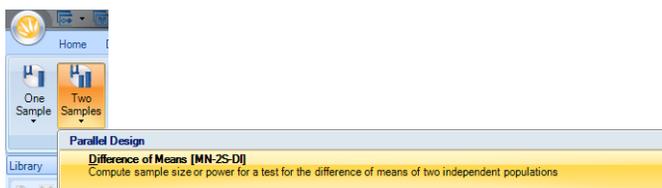
6 Tutorial: Normal Endpoint

This tutorial introduces you to East on the Architect platform, using an example clinical trial to test difference of means.

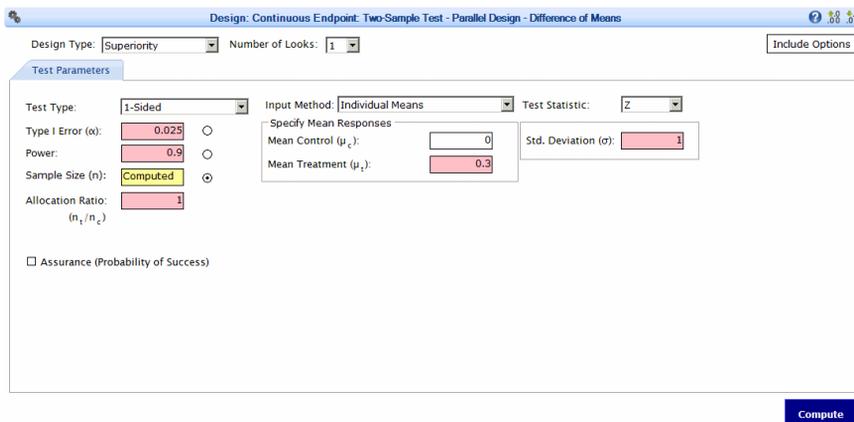
6.1 Fixed Sample Design

When you open East, by default, the Design tab in the ribbon will be active.

The items on this tab are grouped under the following categories of endpoints: Continuous, Discrete, Count, Survival, and General. Click **Continuous: Two Samples**, and then **Parallel Design: Difference of Means**.



The following input window will appear.



By default, the radio button for **Sample Size (n)** is selected, indicating that it is the variable to be computed. The default values shown for **Type I Error** and **Power** are 0.025 and 0.9. Keep the same for this design. Since the default inputs provide all of the necessary input information, you are ready to compute sample size by clicking the **Compute** button. The calculated result will appear in the **Output Preview** pane, as

6 Tutorial: Normal Endpoint

shown below.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | Input Method | δ | μ_c | Mean Treatment (Alt.) | σ | Test Statistic |
|------|-------------|--------------|-----------|--------------------|-------|-------|-------------|------------------|----------|---------|-----------------------|----------|----------------|
| Des1 | Superiority | 1 | 1-Sided | 0.025 | 0.9 | 1 | 467 | Individual Means | 0.3 | 0 | 0.3 | 1 | Z |

This single row of output contains relevant details of inputs and the computed result of total sample size (and total completers) of 467. Select this row, and click  to display a summary of the design details in the upper pane (known as **Output Summary**).

| Des 1 | |
|---|------------------|
| Mnemonic | MN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Input Method | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 0.3 |
| Mean Control (μ_c) | 0 |
| Mean Treatment (μ_t) | 0.3 |
| Std. Deviation (σ) | 1 |
| Test Statistic | Z |
| Sample Size | |
| Maximum | 467 |

The discussion so far gives you a quick feel of the software for computing sample size for a single look design. We will describe further features in an example for a group sequential design in the next section.

6.2 Group Sequential Design for a Normal Superiority Trial

6.2.1 Study Background

Drug X is a newly developed lipase inhibitor for obesity management that acts by inhibiting the absorption of dietary fats. The performance of this drug needs to be compared with an already marketed drug Y for the same condition. In a randomized,

double-blind, trial comparing the efficacy and safety of 1 year of treatment with X to Y (each at 120 mg for three times a day), obese adults are to be randomized to receive either X or Y combined with dietary intervention for a period of one year. The endpoint is weight loss (in pounds). You are to design a trial having 90% power to detect a mean difference of 9 lbs between X and Y, assuming 15 lbs and 6 lbs weight loss in each treatment arm, respectively, and a common standard deviation of 32 lbs. The design is required to be a 2-sided test at the 5% significance level.

From the design menu choose **Continuous: Two Samples**, and then **Parallel Design: Difference of Means**. Select **2-Sided** for **Test Type**, and enter **0.05** for **Type I Error**. Specify the **Mean Control** be **6**, the **Mean Treatment** to be **15**, and the common **Std. Deviation** to be **32**. Next, change the **Number of Looks** to be **5**. You will see a new tab, **Boundary**, added to the input dialog box.

Click the **Boundary** tab, and you will see the following screen. On this tab, you can choose whether to specify stopping boundaries for efficacy, or futility, or both. For this trial, choose efficacy boundaries only, and leave all other default values. We will implement the Lan-Demets (O’Brien-Fleming) spending function, with equally spaced

6 Tutorial: Normal Endpoint

looks.

Design Type: Superiority Number of Looks: 5

Test Parameters **Boundary**

Efficacy
 Boundary Family: Spending Functions
 Spending Function: Lan-DeMets
 Parameter: OF
 Type I Error (α): 0.05

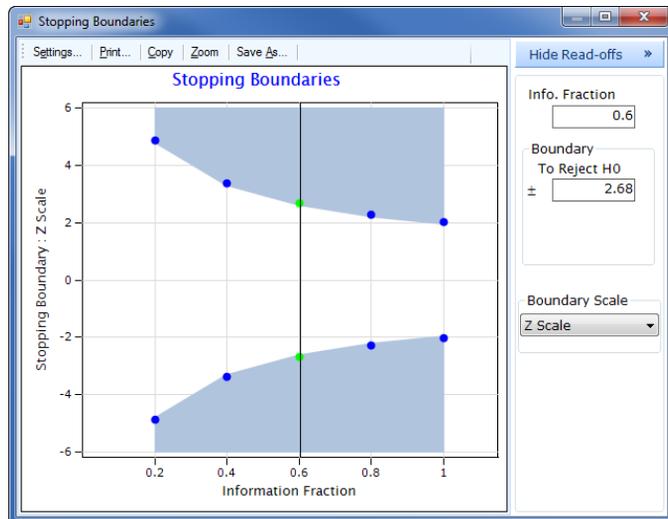
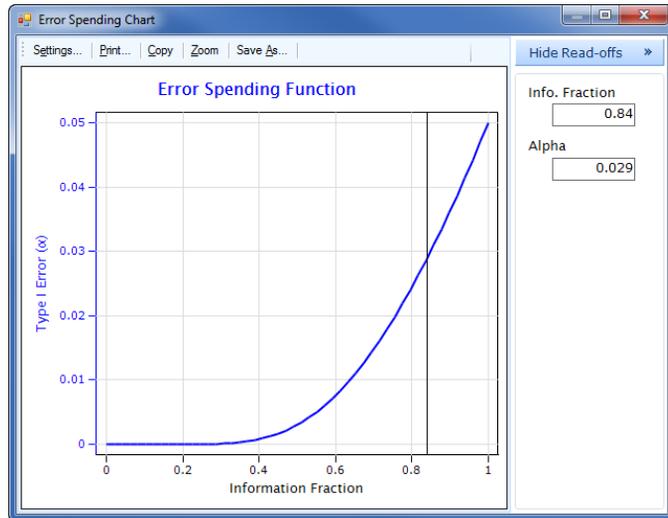
Futility
 Boundary Family: None

Spacing of Looks: Equal Unequal Efficacy Boundary: Z Scale  

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.200 | 0.000 | 4.877 | -4.877 |
| 2 | 0.400 | 0.001 | 3.357 | -3.357 |
| 3 | 0.600 | 0.008 | 2.680 | -2.680 |
| 4 | 0.800 | 0.024 | 2.290 | -2.290 |
| 5 | 1.000 | 0.050 | 2.031 | -2.031 |

On the **Boundary** tab near the Efficacy drop-down box, click on the icons  or

 , to generate the following charts.



Click **Compute**. East will show the results in the **Output Preview**.

The maximum combined sample size required under this design is 544. The expected

6 Tutorial: Normal Endpoint

sample sizes under H0 and H1 are 540 and 403, respectively. Click  in the **Output Preview** toolbar to save this design to Wbk1 in the **Library**. Double-click on Des1 to generate the following output.

Design: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means

| Test Parameters | |
|--------------------------------|------------------|
| Design ID | Des2 |
| Design Type | Superiority |
| Number of Looks | 5 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Test Statistic | Z |
| Input Method | Individual Means |
| Mean Control (μ_c) | 6 |
| Mean Treatment (μ_t) | 15 |
| $\delta = \mu_t - \mu_c$ | |
| Under H0 | 0 |
| Under H1 | 9 |
| Std. Deviation (σ) | 32 |
| Allocation Ratio (n_1/n_2) | 1 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|--------------------------------|-------------|---------------|---------|
| Sample Size (n) | | | |
| Maximum | 272 | 272 | 544 |
| Expected H1 | 201.775 | 201.474 | 403.249 |
| Expected H0 | 270.224 | 270.207 | 540.43 |
| Maximum Information (I): 0.133 | | | |

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | | Boundary Crossing Probability (Incremental) | | | |
|--------|--------------------------|-----------------|---------------------------|------------|--------|---|----------|----------|-----------|
| | | | | Efficacy Z | | Under H0 | | Under H1 | |
| | | | | Upper | Lower | Upper | Lower | Upper | Lower |
| 1 | 0.2 | 109 | 1.104E-6 | 4.872 | -4.872 | 5.519E-7 | 5.519E-7 | 3.32E-4 | 1.146E-10 |
| 2 | 0.401 | 218 | 7.981E-4 | 3.354 | -3.354 | 3.985E-4 | 3.985E-4 | 0.1 | 2.818E-8 |
| 3 | 0.599 | 326 | 0.008 | 2.682 | -2.682 | 0.003 | 0.003 | 0.344 | 8.483E-8 |
| 4 | 0.8 | 435 | 0.024 | 2.29 | -2.29 | 0.008 | 0.008 | 0.301 | 7.455E-8 |
| 5 | 1 | 544 | 0.05 | 2.031 | -2.031 | 0.013 | 0.013 | 0.154 | 4.003E-8 |

Once you have finished examining the output, close this window, and re-start East before continuing.

6.2.2 Creating multiple designs easily

In East, it is easy to create multiple designs by inputting multiple parameter values. In the trial described above, suppose we want to generate designs for all combinations of the following parameter values: **Power** = 0.8, 0.9, and **Difference in Means** = 8.5, 9, 9.5, 10. The number of such combinations is $2 \times 4 = 8$.

East can create all 8 designs by a single specification in the input dialog box. Enter the following values as shown below. Remember that the common **Std. Deviation** is 32. From the **Input Method**, select the **Difference of Means** option. The values of **Power** have been entered as a list of comma-separated values, while **Difference in**

Means has been entered as a colon-separated range of values: 8.5 to 10 in steps of 0.5.

Now click compute. East computes all 8 designs, and displays them in the **Output Preview** as shown below. Click to maximize the **Output Preview**.

Select the first 3 rows using the Ctrl key, and click to display a summary of the design details in the upper pane, known as the **Output Summary**.

| | Des1 | Des2 | Des3 |
|--|---------------------|---------------------|---------------------|
| Mnemonic | MN-25-DI | MN-25-DI | MN-25-DI |
| Test Parameters | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 5 | 5 | 5 |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.8 | 0.8 | 0.801 |
| Model Parameters | | | |
| Allocation Ratio (n ₁ /n ₂) | 1 | 1 | 1 |
| Input Method | Difference of Means | Difference of Means | Difference of Means |
| Diff. in Means ($\delta = \mu_1 - \mu_2$) | 8.5 | 9 | 9.5 |
| Std. Deviation (σ) | 32 | 32 | 32 |
| Test Statistic | Z | Z | Z |
| Boundary Parameters | | | |
| Spacing of Looks | Equal | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) |
| Sample Size | | | |
| Maximum | 456 | 407 | 366 |
| Expected Under H ₀ | 453.004 | 404.326 | 363.595 |
| Expected Under H ₁ | 366.549 | 327.093 | 294.023 |

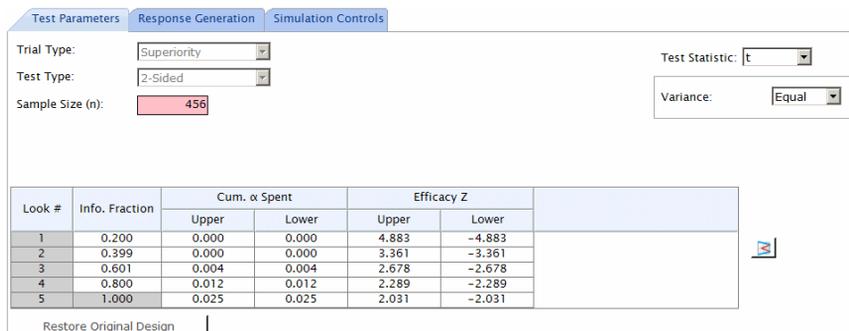
Select Des1 in the **Output Preview**, and click toolbar to save this design in the **Library**. We will use this design for simulation and interim monitoring, as described below. Now that you have saved Des1, delete all designs from the **Output Preview** before continuing, by selecting all designs with the Shift key, and clicking in the toolbar.

6.2.3 Simulation

Right-click Des1 in the **Library**, and select **Simulate**. Alternatively, you can select

6 Tutorial: Normal Endpoint

Des1 and click the  icon.



The screenshot shows the 'Simulation Controls' tab of a software interface. The parameters are set as follows:

- Trial Type: Superiority
- Test Type: 2-Sided
- Sample Size (n): 456
- Test Statistic: Z
- Variance: Equal

Below the parameters is a table with the following data:

| Look # | Info. Fraction | Cum. α Spent | | Efficacy Z | |
|--------|----------------|---------------------|-------|------------|--------|
| | | Upper | Lower | Upper | Lower |
| 1 | 0.200 | 0.000 | 0.000 | 4.883 | -4.883 |
| 2 | 0.399 | 0.000 | 0.000 | 3.361 | -3.361 |
| 3 | 0.601 | 0.004 | 0.004 | 2.678 | -2.678 |
| 4 | 0.800 | 0.012 | 0.012 | 2.289 | -2.289 |
| 5 | 1.000 | 0.025 | 0.025 | 2.031 | -2.031 |

At the bottom of the interface, there is a 'Restore Original Design' button.

We will carry out a simulation of Des1 to check whether it preserves the specified power. Click **Simulate**. East will execute by default 10000 simulations with the specified inputs. Close the intermediate window after examining the results. A row labeled as Sim1 will be added in the **Output Preview**.

Click the  icon to save this simulation to the **Library**. A simulation sub-node will be added under Des1 node. Double clicking on the Sim1 node, will display the

detailed simulation output in the work area.

Simulation: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means

| Simulation Parameters | |
|--------------------------------|------------------|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 5 |
| Test Type | 2-Sided |
| Sample Size (n) | 456 |
| Variance | Equal |
| Test Statistic | t |
| Avg. Power at Termination | 0.802 |
| Response Generation Parameters | |
| Generate Data Using | Individual Means |
| Mean Control (μ_c) | 0 |
| Mean Treatment (μ_t) | 8.5 |
| SD Control (σ_c) | 32 |
| SD Treatment (σ_t) | 32 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

⊖ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 91 |
| 2 | 182 |
| 3 | 274 |
| 4 | 365 |
| 5 | 456 |
| Average | 366.996 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries Efficacy | | Stopping For | | Total Simulations | |
|--------|-----------------|---------------------|--------|----------------|----------------|-------------------|---------|
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 91 | 4.883 | -4.883 | 2 | 0 | 2 | 0.020% |
| 2 | 182 | 3.361 | -3.361 | 609 | 0 | 609 | 6.090% |
| 3 | 274 | 2.678 | -2.678 | 2523 | 0 | 2523 | 25.230% |
| 4 | 365 | 2.289 | -2.289 | 2893 | 0 | 2893 | 28.930% |
| 5 | 456 | 2.031 | -2.031 | 1996 | 0 | 3973 | 39.730% |
| Total | | | | 8023 | 0 | 10000 | |
| % | | | | 80.230% | 0.000% | | |

Simulation Seed and Elapsed Time

Starting Seed: 725323
 Total Number of Simulations: 10000
 Elapsed Time: 00.01.05

In 80.23% of the simulated trials, the null hypothesis was rejected. This value is very close to the specified power of 80%. Note that your results may differ from the results displayed over here as the simulations would be run with different seed. The next section will explore interim monitoring with this design.

6 Tutorial: Normal Endpoint

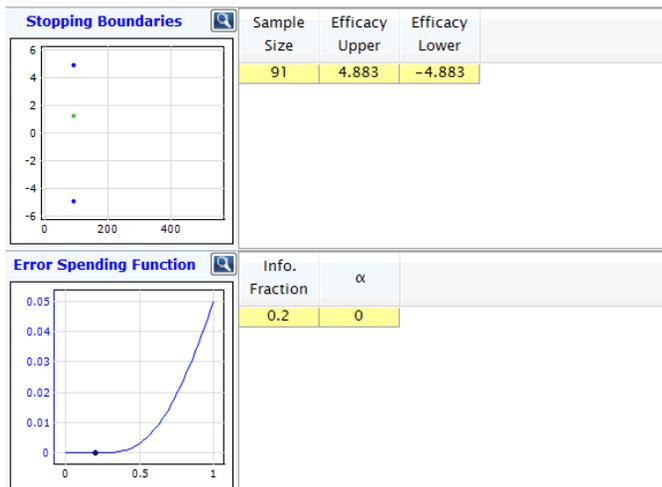
6.2.4 Interim Monitoring

Right-click Des1 in the **Library** and select **Interim Monitoring**. Click the **Enter Interim Data** to open the **Test Statistic Calculator**. Suppose that after 91 subjects, at the first look, you have observed a mean difference of 8.5, with a standard error of 6.709.

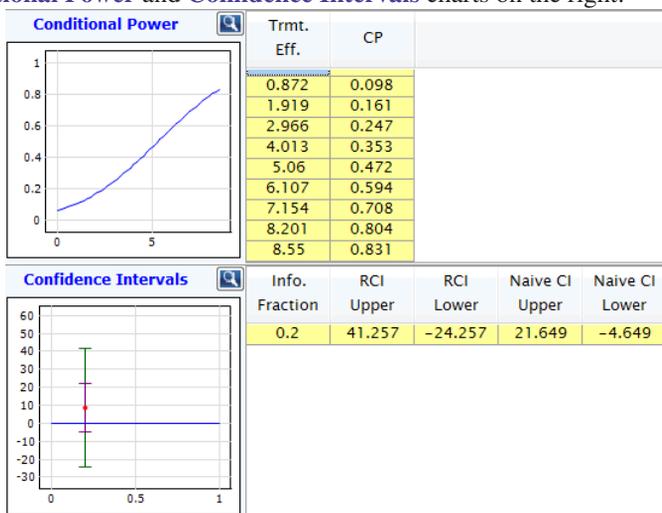
Click **OK** to update the IM Dashboard.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | | 95% RCI for δ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|--------|----------------------|---------|------------------|-------|------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.2 | 91 | 1.267 | 8.5 | 6.709 | 4.883 | -4.883 | 41.257 | -24.257 | 0.932 | 0.828 | 0.673 |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | |

The **Stopping Boundaries** and **Error Spending Function** charts on the left:



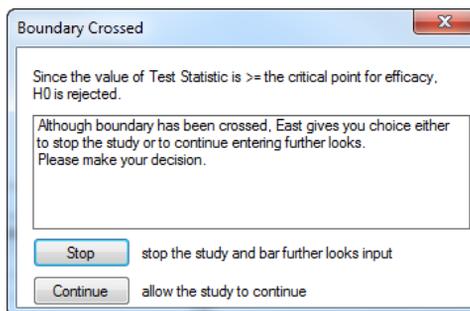
The Conditional Power and Confidence Intervals charts on the right:



Suppose that after 182 subjects, at the second look, you have observed a mean difference of 16, with a standard error of 4.744. Click **Recalc**, and then **OK** to update the IM Dashboard. In this case, a boundary has been crossed, and the following

6 Tutorial: Normal Endpoint

window appears.



Click **Stop** to complete the trial. The IM Dashboard will be updated accordingly, and a table for **Final Inference** will be displayed as shown below.

| Final Inference | |
|--|--------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.001 |
| Adj. Pt. Est. for δ | 16 |
| Adj. 95% CI for δ | |
| Upper Confidence Bound | 25.298 |
| Lower Confidence Bound | 6.702 |
| Post-Hoc Power | |

7 Normal Superiority One-Sample

To compare a new process or treatment to a well-established control, a single-sample study may suffice for preliminary information prior to a full-scale investigation. This single sample may either consist of a random sample of observations from a single treatment when the mean is to be compared to a specified constant or a random sample of paired differences or ratio between two treatments. The former is presented in Section (7.1) and the latter is discussed in Section (7.2) and Section (7.3).

7.1 Single Mean

7.1.1 Trial Design

7.1.2 Simulation

7.1.3 Interim Monitoring

7.1.4 Trial Design Using a t-Test (Single Look)

The problem of comparing the mean of the distribution of observations from a single random sample to a specified constant is considered. For example, when developing a new drug for treatment of a disease, there should be evidence of efficacy. For this single-sample problem, it is desired to compare the unknown mean μ to a fixed value μ_0 . The null hypothesis $H_0: \mu = \mu_0$ is tested against the two-sided alternative hypothesis $H_1: \mu \neq \mu_0$ or a one-sided alternative hypothesis $H_1: \mu < \mu_0$ or $H_1: \mu > \mu_0$. The power of the test is computed at a specified value of $\mu = \mu_1$ and standard deviation σ .

Let $\hat{\mu}_j$ denote the estimate of μ based on n_j observations, up to and including the j -th look, $j = 1, \dots, K$, with a maximum of K looks. The test statistic at the j -th look is based on the value specified by the null hypothesis, namely

$$Z_j = n_j^{1/2}(\hat{\mu}_j - \mu_0)/\hat{\sigma}_j, \tag{7.1}$$

where $\hat{\sigma}_j^2$ is the sample variance based on n_j observations.

7.1.1 Trial Design

Consider the situation where treatment for a certain infectious disorder is expected to result in a decrease in the length of hospital stay. Suppose that hospital records were reviewed and it was determined that, based on this historical data, the average hospital stay is approximately 7 days. It is hoped that the new treatment can decrease this to less than 6 days. It is assumed that the standard deviation is $\sigma = 2.5$ days. The null hypothesis $H_0: \mu = 7 (= \mu_0)$ is tested against the alternative hypothesis $H_1: \mu < 7$.

First, click **Continuous: One Sample** on the **Design** tab and then click **Single Arm Design: Single Mean**.

This will launch a new input window.

Single-Look Design

7 Normal Superiority One-Sample

We want to determine the sample size required to have power of 90% when $\mu = 6 (= \mu_1)$, using a test with a one-sided type-1 error rate of 0.05. Choose **Test Type** as **1-Sided**. Specify **Mean Response under Null (μ_0)** as 7, **Mean Response under Alt. (μ_1)** as 6 and **Std. Deviation (σ)** as 2.5. The upper pane should appear as below:

Click **Compute**. This will calculate the sample size for this design and the output is shown as a row in the **Output Preview**. The computed sample size is 54 subjects.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | Sample Size | σ | μ_0 | μ_1 | Test Statistic |
|------|-------------|--------------|-----------|--------------------|-------|-------------|----------|---------|---------|----------------|
| Des1 | Superiority | 1 | 1-Sided | 0.05 | 0.902 | 54 | 2.5 | 7 | 6 | Z |

This design has default name Des 1. Select this design by clicking anywhere along the row and click  in the **Output Preview** toolbar. Some of the design details will

be displayed in the upper pane, labeled as **Output Summary**.

| Output Summary | |
|--------------------------------------|--------------|
| | Des 1 |
| Mnemonic | MN-1S-SM |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.902 |
| Model Parameters | |
| Std. Deviation (σ) | 2.5 |
| Mean Response under Null (μ_0) | 7 |
| Mean Response under Alt. (μ_1) | 6 |
| Test Statistic | Z |
| Sample Size | |
| Maximum | 54 |

In the **Output Preview** toolbar select Des 1, click  to save this design to Wbk1 in the **Library**.

Five-Look Design

To allow the opportunity to stop early and proceed with a full-scale plan, five equally-spaced analyses are planned, using the Lan-DeMets (O'Brien-Fleming) stopping boundary. Create a new design by right-clicking Des 1 in the **Library**, and selecting **Edit Design**. In the Input, change the **Number of Looks** from 1 to 5, to generate a study with four interim looks and a final analysis. A new tab for **Boundary Info** should appear. Click this tab to reveal the stopping boundary parameters. By default, the **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The left side contains details for the **Efficacy** boundary, and the right side contains details for the **Futility** boundary. By default, there is an efficacy boundary (to reject H_0) selected, but no futility boundary (to reject H_1). The **Boundary Family** specified is of the **Spending Functions** type. The default **Spending Function** is the **Lan-DeMets** (Lan & DeMets, 1983), with **Parameter** as **OF** (O'Brien-Fleming), which generates boundaries that are very similar, though not identical, to the classical stopping boundaries of O'Brien and Fleming (1979). For a detailed description of the different spending functions and stopping boundaries available in East refer to Chapter 62. The cumulative alpha spent and the boundary values are displayed below.

7 Normal Superiority One-Sample

Design Type: **Superiority** Number of Looks: **5** Include Options

Design Parameters **Boundary Info**

Efficacy
 Boundary Family: **Spending Functions**
 Spending Function: **Lan-DeMets**
 Parameter: **OF**
 Type I Error (α): 0.05

Futility
 Boundary Family: **None**

Spacing of Looks: Equal Unequal Efficacy Boundary: **Z Scale**

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.200 | 0.000 | -4.229 |
| 2 | 0.400 | 0.002 | -2.888 |
| 3 | 0.600 | 0.011 | -2.298 |
| 4 | 0.800 | 0.028 | -1.962 |
| 5 | 1.000 | 0.050 | -1.740 |

Compute

Click **Compute**. The maximum and expected sample sizes are highlighted in yellow in the **Output Preview**. Save this design in the current workbook by selecting the corresponding row in the **Output Preview** and clicking  on the **Output Preview** toolbar. To compare Des 1 and Des 2, select both rows in **Output Preview** using the Ctrl key and click  in the **Output Preview** toolbar. This will display both designs in the **Output Summary** pane.

| | Des 1 | Des 2 |
|--------------------------------------|-------------|-------------|
| Mnemonic | MN-1S-SM | MN-1S-SM |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 5 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.902 | 0.903 |
| Model Parameters | | |
| Std. Deviation (σ) | 2.5 | 2.5 |
| Mean Response under Null (μ_0) | 7 | 7 |
| Mean Response under Alt. (μ_1) | 6 | 6 |
| Test Statistic | Z | Z |
| Boundary Parameters | | |
| Spacing of Looks | | Equal |
| Efficacy Boundary | | LD (OF) |
| Sample Size | | |
| Maximum | 54 | 56 |
| Expected Under H0 | | 55.531 |
| Expected Under H1 | | 39.897 |

Des 2 results in a maximum of 56 subjects in order to attain 90% power, with an expected sample size of 40 under the alternative hypothesis. In order to see the stopping probabilities, double-click Des 2 in the **Library**.

☺ **Stopping Boundaries: Look by Look**

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|------------|---|----------|
| | | | | | Under H0 | Under H1 |
| | | | | Efficacy Z | Efficacy | Efficacy |
| 1 | 0.196 | 11 | 9.767E-6 | -4.27 | 9.767E-6 | 0.002 |
| 2 | 0.393 | 22 | 0.002 | -2.918 | 0.002 | 0.147 |
| 3 | 0.607 | 34 | 0.012 | -2.279 | 0.01 | 0.376 |
| 4 | 0.804 | 45 | 0.029 | -1.958 | 0.017 | 0.25 |
| 5 | 1 | 56 | 0.05 | -1.741 | 0.021 | 0.128 |

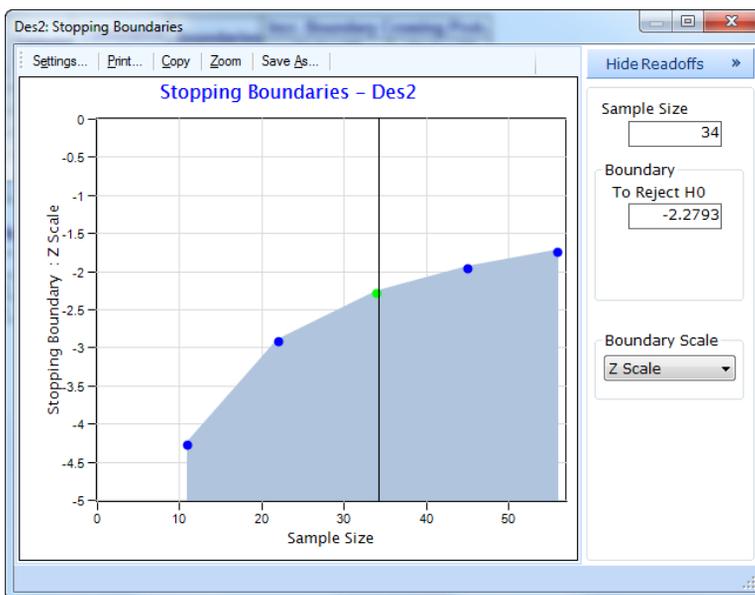
The clear advantage of this sequential design resides in the relatively high cumulative probability of stopping by the third look if the alternative is true, with a sample size of 34 patients, which is well below the requirements for a fixed sample study (54 patients). Close the Output window before continuing.

Examining stopping boundaries and spending functions

You can plot the boundary values of Des 2 by clicking  on the **Library** toolbar,

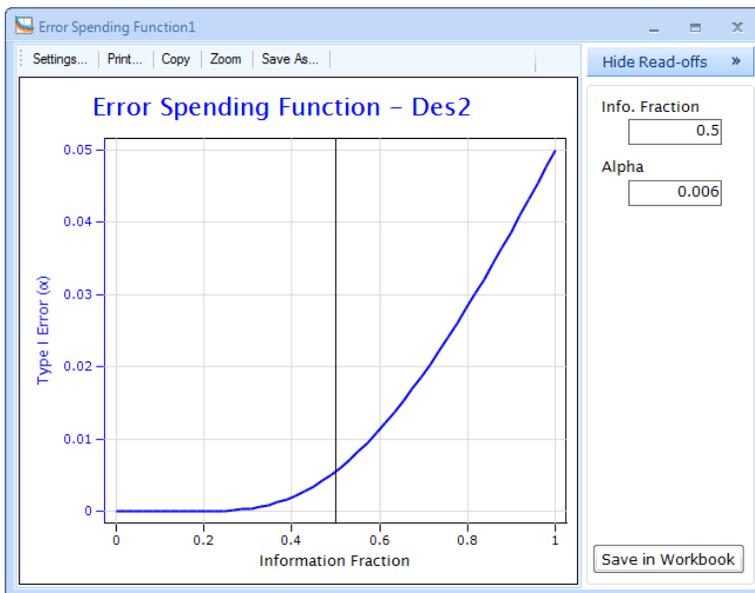
7 Normal Superiority One-Sample

and then clicking **Stopping Boundaries**. The following chart will appear:



You can choose different boundary scales from the drop down box located in the right hand side. The available boundary scales are Z scale, Score Scale, μ/σ Scale and p -value scale. To plot the error spending function for Des 2, select Des 2 in the **Library**, click  in the toolbar, and then click **Error Spending**. The following

chart will appear:



The above spending function is according to Lan and DeMets (1983) with O’Brien-Fleming flavor and for one-sided tests has the following functional form:

$$\alpha(t) = 2 - 2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{t}}\right)$$

Observe that very little of the total type-1 error is spent early on, but more is spent rapidly as the information fraction increases, and reaches 0.05 at an information fraction of 1. Feel free to try other plots by clicking  in the Library toolbar. Close all charts before continuing.

7.1.2 Simulation

Suppose we want to see the advantages of performing the interim analyses, as it relates to the chance of stopping prior to the final analysis. This examination can be conducted using simulation. Select Des 2 in the **Library**, and click  in the toolbar. Alternatively, right-click on Des 2 and select **Simulate**. A new Simulation window will appear. For example, suppose you wish to determine how quickly this trial could be

7 Normal Superiority One-Sample

terminated if the treatment difference was much greater than expected. For example, under the alternative hypothesis, $\mu = 4.5$. Click on the **Response Generation Info** tab, and specify: **Mean Response**(μ) = 4.5 and **Std. Deviation** (σ) = 2.5.

Test Parameters
Response Generation
Simulation Controls

Specify Mean Response

Mean Response (μ):

Std. Deviation (σ):

Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1.

Select Sim 1 in the **Output Preview** and click . Now double-click on Sim 1 in the **Library**. The simulation output details will be displayed in the upper pane.

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | | Count | % |
| | | Lower | Efficacy | | |
| 1 | 11 | -4.27 | Efficacy | 2606 | 26.060% |
| 2 | 22 | -2.918 | Efficacy | 6886 | 68.860% |
| 3 | 34 | -2.279 | Efficacy | 504 | 5.040% |
| 4 | 45 | -1.958 | Efficacy | 4 | 0.040% |
| 5 | 56 | -1.741 | Efficacy | 0 | 0.000% |
| Total | | | Efficacy | 10000 | |
| % | | | Efficacy | 100.000% | |

Observe that 100% simulated trials rejected the null hypothesis, and about 26% of these simulations were able to reject the null at the first look after enrolling only 11 subjects. Your numbers might differ slightly due to a different starting seed.

7.1.3 Interim Monitoring

Suppose that the trial has commenced and Des 2 was implemented. Right-click Des 2 in the **Library**, and select **Interim Monitoring**.

Although we specified that there will be five equally spaced interim looks, the Lan-DeMets methodology implemented in East allows you to alter the number and spacing of these looks. Accordingly, suppose that an interim look was taken after enrolling 20 subjects and the sample mean, based on these 20 subjects, was 5.1 with a standard error of 0.592. Since $\mu_0 = 7$, based on equation (7.1) the value of the test statistic at the first look would be $Z_1 = (5.1 - 7)/0.592$ or -3.209.

Click **Enter Interim Data** on the toolbar. In the **Test Statistic Calculator**, enter the following values, and click **Recalc** and then **OK**.

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Cumulative Sample Size: 20

Input for Normal end point

Estimate of μ : 5.1

Standard Error of Estimate of μ : 0.592

Output

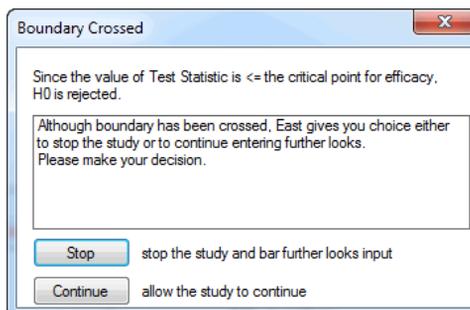
Estimate of $\mu - \mu_0$: -1.9

Test Statistic: -3.209

Recalc OK Cancel

Since the stopping boundary is crossed, the following dialog box appears.

7 Normal Superiority One-Sample



Click **Stop** to take you back to the interim monitoring dashboard. For final inference, East will display the following summary information on the dashboard.

| Final Inference | |
|---|-------|
| Final Outputs at Look # | 1 |
| Adj. p-value | 0.001 |
| Adj. Pt. Est. for μ | 5.1 |
| Adj. 90% CI for μ | |
| Upper Confidence Bound | 6.074 |
| Lower Confidence Bound | 4.126 |
| Post-Hoc Power | |

7.1.4 Trial Design Using a t-Test (Single Look)

The sample size obtained to correctly power Des 1 in Section (7.1.1) relied on using a Wald-type statistic for the hypothesis test, given by equation (7.1). Due to the assumption of normal distribution for the test statistic, we have ignored the fact that the variance σ is estimated from the sample. For large sample sizes this approximation is acceptable. However, in small samples with unknown standard deviation the test statistic

$$Z = n^{1/2}(\hat{\mu} - \mu_0)/\hat{\sigma}, \quad (7.2)$$

is distributed with student's t distribution with $(n - 1)$ degrees of freedom. Here, $\hat{\sigma}^2$ denotes the sample variance based on n observations.

Consider the example in Section 7.1.1 where we would like to test the null hypothesis that the average hospital stay is 7 days, $H_0: \mu = 7 (= \mu_0)$, against the alternative hypothesis that is less than 7 days, $H_1: \mu < 7$. We will now design the same trial in a different manner, using the t distribution for the test statistic.

Right-click Des 1 in the **Library**, and select **Edit Design**. In the input window, change

the **Test Stat.** from **z** to **t**. The entries for the other fields need not be changed.

Click **Compute**. East will add an additional row to the **Output Preview** labeled as Des 3. The required sample size is 55. Select the rows corresponding to Des 1 and Des 3 and click . This will display Des 1 and Des 3 in the **Output Summary**.

| | Des 1 | Des 3 |
|--------------------------------------|-------------|-------------|
| Mnemonic | MN-1S-SM | MN-1S-SM |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.902 | 0.9 |
| Model Parameters | | |
| Std. Deviation (σ) | 2.5 | 2.5 |
| Mean Response under Null (μ_0) | 7 | 7 |
| Mean Response under Alt. (μ_1) | 6 | 6 |
| Test Statistic | Z | t |
| Sample Size | | |
| Maximum | 54 | 55 |

Des 3, which uses the t distribution, requires that we commit a combined total of 55 patients to the study, just one more compared to Des 1, which uses the normal distribution. The extra patient is needed to compensate for the extra variability due to estimation of the $\text{var}[\hat{\delta}]$.

7.2 Mean of Paired Differences

7.2.1 Trial Design

7.2.2 Simulation

7.2.3 Interim Monitoring

7.2.4 Trial Design Using a t-Test (Single Look)

The paired t-test is used to compare the means of two normal distributions when each observation in the random sample from one distribution is matched with a unique observation from the other distribution. Let μ_c and μ_t denote the two means to be compared and let σ^2 denote the variance of the differences.

The null hypothesis $H_0: \mu_c = \mu_t$ is tested against the two-sided alternative hypothesis $H_1: \mu_c \neq \mu_t$ or a one-sided alternative hypothesis $H_1: \mu_c < \mu_t$ or $H_1: \mu_c > \mu_t$. Let $\delta = \mu_t - \mu_c$. The null hypothesis can be expressed as $H_0: \delta = 0$ and the alternative can be expressed as $H_1: \delta \neq 0$, $H_1: \delta > 0$, or $H_1: \delta < 0$. The power of the test is computed at specified values of μ_c, μ_t , and σ .

Let $\hat{\mu}_{cj}$ and $\hat{\mu}_{tj}$ denote the estimates of μ_c and μ_t based on n_j observations, up to and including j -th look, $j = 1, \dots, K$ where a maximum of K looks are to be made. The estimate of the difference at the j -th look is

$$\hat{\delta}_j = \hat{\mu}_{tj} - \hat{\mu}_{cj}$$

7 Normal Superiority One-Sample

and the test statistic at the j -th look is

$$Z_j = n_j^{1/2} \hat{\delta}_j / \hat{\sigma}_j, \tag{7.3}$$

where $\hat{\sigma}_j^2$ is the sample variance of n_j paired differences.

7.2.1 Trial Design

Consider the situation where subjects are treated once with placebo after pain is experimentally induced, and later treated with a new analgesic after pain is induced a second time. Pain is reported by the subjects using a 10 cm visual analog scale (0=“no pain”, . . . , 10=“extreme pain”). After treatment with placebo, the average is expected to be 6 cm. After treatment with the analgesic, the average is expected to be 4 cm. It is assumed that the common standard deviation is $\sigma = 5$ cm. The null hypothesis $H_0: \delta = 0$ is tested against the alternative hypothesis $H_1: \delta < 0$.

Start East afresh. First, **Continuous: One Sample** on the **Design** tab, and then click **Paired Design: Mean of Paired Differences**

This will launch a new input window.

Single-Look Design

We want to determine the sample size required to have power of 90% when $\mu_c = 6$ and $\mu_t = 4$, using a test with a one-sided type-1 error rate of 0.05. Select **Test Type** as **1-Sided, Individual Means** for **Input Method**, and specify the **Mean Control** (μ_c) as 6 and **Mean Treatment** (μ_t) as 4. Enter **Std. Dev. of Paired Difference** (σ_D) as 5. The upper pane should appear as below:

Design: Continuous Endpoint: One-Sample Test - Paired Design - Mean of Paired Differences

Design Type: Superiority Number of Looks: 1

Test Parameters

Test Type: 1-Sided Input Method: Individual Means Test Statistic: Z

Type I Error (α): 0.05

Power: 0.9

Sample Size (n): Computed

Specify Mean Responses

Mean Control (μ_c): 6

Mean Treatment (μ_t): 4

Std. Dev. of Paired Difference (σ_D): 5

Assurance (Probability of Success)

Compute

Click **Compute**. This will calculate the sample size for this design and the output is shown as a row in the **Output Preview**. The computed sample size is 54 subjects.

| ▲ | ID | Design Type | No. of Looks | Test Type | Specified α | Power | Sample Size | Test Statistic | Input Method | μ c | Mean Treatment (Alt.) | δ 1 | σ D |
|---|------|-------------|--------------|-----------|--------------------|-------|-------------|----------------|------------------|---------|-----------------------|------------|------------|
| ■ | Des1 | Superiority | 1 | 1-Sided | 0.05 | 0.902 | 54 | Z | Individual Means | 6 | 4 | -2 | 5 |

This design has default name Des 1. Select this design by clicking anywhere along the row in the **Output Preview** and click . Some of the design details will be displayed in the upper pane, labeled as **Output Summary**.

| Des 1 | |
|-------------------------------------|------------------|
| Mnemonic | MN-1S-PDI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.902 |
| Model Parameters | |
| Test Statistic | Z |
| Input Method | Individual Means |
| Mean Control (μ c) | 6 |
| Mean Treatment (μ t) | 4 |
| Diff. of Means (μ t - μ c) | -2 |
| Std. Deviation (σ D) | 5 |
| Sample Size | |
| Maximum | 54 |

In the **Output Preview** toolbar select Des 1, click  to save this design to Wbk1 in the **Library**.

Three-Look Design

For the above study, suppose we wish to take up to two equally spaced interim looks and one final look as we accrue data, using the Lan-DeMets (O'Brien-Fleming) stopping boundary. Create a new design by right-clicking Des 1 in the **Library**, and **Edit Design**. In the Input, change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis.

Click **Compute**. The maximum and expected sample sizes are highlighted in yellow in the **Output Preview**. Save this design in the current workbook by selecting the corresponding row in **Output Preview** and clicking  on the **Output Preview** toolbar. To compare Des 1 and Des 2, select both rows in **Output Preview** using the

7 Normal Superiority One-Sample

Ctrl key and click  . Both designs will be displayed in the **Output Summary** pane.

| | Des 1 | Des 2 |
|------------------------------------|------------------|------------------|
| Mnemonic | MN-1S-PDI | MN-1S-PDI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.902 | 0.902 |
| Model Parameters | | |
| Test Statistic | Z | Z |
| Input Method | Individual Means | Individual Means |
| Mean Control (μ_C) | 6 | 6 |
| Mean Treatment (μ_T) | 4 | 4 |
| Diff. of Means ($\mu_T - \mu_C$) | -2 | -2 |
| Std. Deviation (σ_D) | 5 | 5 |
| Boundary Parameters | | |
| Efficacy Boundary | | LD (OF) |
| Spacing of Looks | | Equal |
| Sample Size | | |
| Maximum | 54 | 55 |
| Expected Under H0 | | 54.685 |
| Expected Under H1 | | 42.646 |

Des 2 results in a maximum of 55 subjects in order to attain 90% power, with an expected sample size of 43 under the alternative hypothesis. In the **Output Preview** toolbar select Des 2, click  to save this design to Wbk1 in the **Library**. In order to see the stopping probabilities, double-click Des 2 in the **Library**.

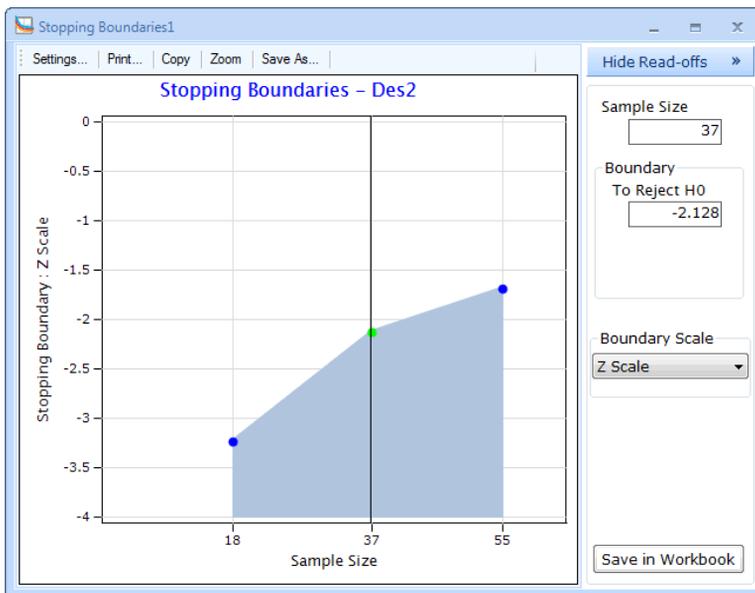
Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|------------|---|----------|
| | | | | | Under H0 | Under H1 |
| | | | | Efficacy Z | Efficacy | Efficacy |
| 1 | 0.327 | 18 | 6.124E-4 | -3.233 | 6.124E-4 | 0.062 |
| 2 | 0.673 | 37 | 0.017 | -2.128 | 0.016 | 0.558 |
| 3 | 1 | 55 | 0.05 | -1.696 | 0.033 | 0.282 |

The clear advantage of this sequential design resides in the high cumulative probability of stopping by the third look if the alternative is true, with a sample size of 37 patients, which is well below the requirements for a fixed sample study (54 patients). Close the Output window before continuing.

Select Des 2 and click  on the Library toolbar. You can select one of many

plots, including one for **Stopping Boundaries**:



Close this chart before continuing.

7.2.2 Simulation

Select Des 2 in the **Library**, and click  in the toolbar. Click on the **Response Generation Info** tab, and make sure **Mean Treatment**(μ_t) = 4, **Mean Control**(μ_c) = 6 and **Std. Deviation** (σ) = 5. Click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1.

Select Sim 1 in the **Output Preview** and click  . Now double-click on Sim 1 in

7 Normal Superiority One-Sample

the **Library**. The simulation output details will be displayed.

Simulation: Continuous Endpoint: One-Sample Test - Paired Design - Mean of Paired Differences

| Simulation Parameters | |
|--------------------------------|-------------|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 55 |
| Test Statistic | t |
| Avg. Power at Termination | 0.895 |
| Response Generation Parameters | |
| Mean Control (μ_c) | 6 |
| Mean Treatment (μ_t) | 4 |
| Std. Deviation (σ_p) | 5 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 18 |
| 2 | 37 |
| 3 | 55 |
| Average | 41.981 |

Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | | Count | % |
| | | Lower | Efficacy | | |
| 1 | 18 | -3.233 | 950 | 950 | 9.500% |
| 2 | 37 | -2.128 | 5280 | 5280 | 52.800% |
| 3 | 55 | -1.696 | 2719 | 3770 | 37.700% |
| Total | | | 8949 | 10000 | |
| % | | | 89.490% | | |

Overall, close to 90% of simulations have rejected H_0 . The numbers on your screen might differ slightly due to a different seed.

7.2.3 Interim Monitoring

For an ongoing study we evaluate the test statistic at an interim stage to see whether we have enough evidence to reject H_0 . Right-click Des 2 in the Library, and select **Interim Monitoring**.

Although the design specified that there be three equally spaced interim looks, the Lan-DeMets methodology implemented in East allows you to alter the number and spacing of these looks. Suppose that an interim look was taken after enrolling 18 subjects and the sample mean, based on these subjects, was -2.2 with a standard error of 1.4. Then based on equation (7.3), the value of the test statistic at first look would be $Z_1 = (-2.2)/1.4$ or -1.571.

Click **Enter Interim Data** on the toolbar. In the **Test Statistic Calculator**, enter the

following values, and click **Recalc** and then **OK**.

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Cumulative Sample Size:

Input for Normal end point:

Estimate of δ :
 δ = mean of paired difference

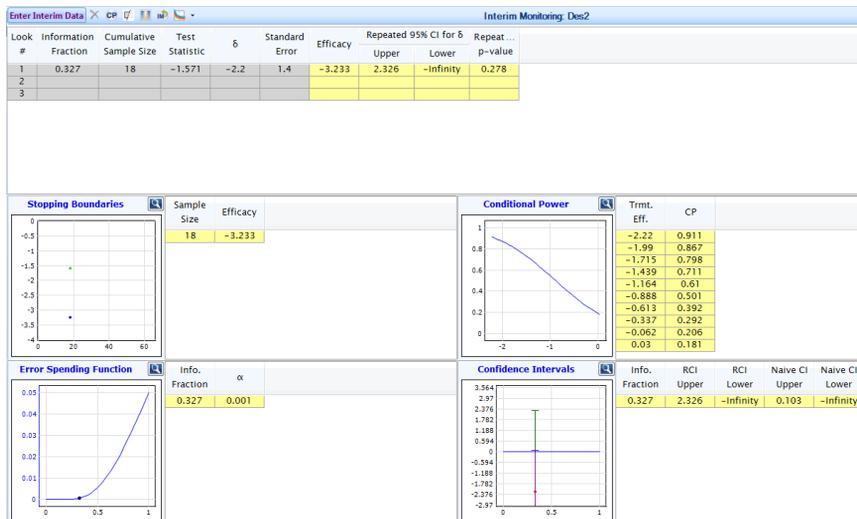
Standard Error of Estimate of δ :

Output:

Test Statistic:

Recalc OK Cancel

The dashboard will be updated accordingly.



As the observed value -1.571 has not crossed the critical boundary value of -3.233, the trial continues. Now, 18 additional subjects are enrolled, and a second interim analysis with 36 subjects is conducted. Suppose that the observed difference is -2.3 with

7 Normal Superiority One-Sample

standard error as 0.8. Select the Look 2 row and click **Enter Interim Data**. Enter these values, and click **Recalc**, and then **OK**.

Test Statistic Calculator

Editing look #2

Set Current Look as Last

Cumulative Sample Size:

Input for Normal end point

Estimate of δ :

δ = mean of paired difference

Standard Error of Estimate of δ :

Output

Test Statistic:

Recalc OK Cancel

Since the stopping boundary is crossed, the following dialog box appears. Click on **Stop**.

Boundary Crossed

Since the value of Test Statistic is \leq the critical point for efficacy, H_0 is rejected.

Although boundary has been crossed, East gives you choice either to stop the study or to continue entering further looks. Please make your decision.

Stop stop the study and bar further looks input

Continue allow the study to continue

For final inference, East will display the following summary information on the dashboard.

| Final Inference | |
|--|--------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.002 |
| Adj. Pt. Est. for δ | -2.287 |
| Adj. 90% CI for δ | |
| Upper Confidence Bound | -0.959 |
| Lower Confidence Bound | -3.607 |
| Post-Hoc Power | |

7.2.4 Trial Design Using a t-Test (Single Look)

The sample size obtained to correctly power the trial in Section (7.2.1) relied on using a Wald-type statistic for the hypothesis test, given by equation (7.3). However, we neglected the fact that the variance σ is estimated by assuming that the test statistic follows a standard normal distribution. For large sample sizes, asymptotic theory supports this approximation. In a single-look design, this test statistic is calculated as

$$Z = n^{1/2} \hat{\delta} / \hat{\sigma}, \quad (7.4)$$

where $\hat{\sigma}^2$ is the sample variance based on n observed paired differences. In the following calculations we take into consideration that Z follows a Student's t-distribution with $(n - 1)$ degrees of freedom.

Consider the example in Section 7.2.1 where we would like to test the null hypothesis that the analgesic does not reduce pain, $H_0: \delta = 0$, against the alternative hypothesis that the new analgesic works to reduce pain, $H_1: \delta < 0$. We will design this same trial using the t distribution for the test statistic.

Right-click Des 1 from the **Library**, and select **Edit Design**. Change the **Test Stat.** from **Z** to **t**. The entries for the other fields need not be changed, and click **Compute**. East will add an additional row to the **Output Preview** labeled as Des 3. Select the rows corresponding to Des 1 and Des 3. This will display Des 1 and Des 3 in the

7 Normal Superiority One-Sample

Output Summary.

| | Des1 | Des3 |
|------------------------------------|------------------|------------------|
| Mnemonic | MN-1S-PDI | MN-1S-PDI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.902 | 0.9 |
| Model Parameters | | |
| Test Statistic | Z | t |
| Input Method | Individual Means | Individual Means |
| Mean Control (μ_c) | 6 | 6 |
| Mean Treatment (μ_t) | 4 | 4 |
| Diff. of Means ($\mu_t - \mu_c$) | -2 | -2 |
| Std. Deviation (σ_D) | 5 | 5 |
| Sample Size | | |
| Maximum | 54 | 55 |

Using the t distribution, we need one extra subject to compensate for the extra variability due to estimation of the $\text{var}[\hat{\delta}]$.

7.3 Ratio of Paired Means

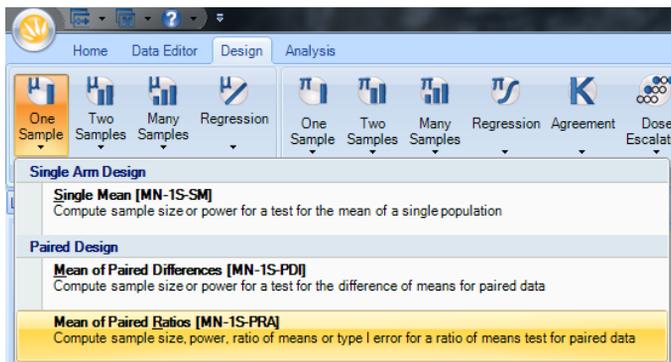
The test for ratio of paired difference is used to compare the means of two log normal distributions when each observation in the random sample from one distribution is matched with a unique observation from the other distribution. Let μ_c and μ_t denote the two means to be compared and let σ_c^2 and σ_t^2 be the respective variances.

The null hypothesis $H_0: \mu_c/\mu_t = 1$ is tested against the two-sided alternative hypothesis $H_1: \mu_c/\mu_t \neq 1$ or a one-sided alternative hypothesis $H_1: \mu_c/\mu_t < 1$ or $H_1: \mu_c/\mu_t > 1$. Let $\rho = \mu_t/\mu_c$. Then the null hypothesis can be expressed as $H_0: \rho = 1$ and the alternative can be expressed as $H_1: \rho \neq 1$, $H_1: \rho > 1$, or $H_1: \rho < 1$. The power of the test is computed at specified values of μ_c , μ_t , and σ . We assume that $\sigma_t/\mu_t = \sigma_c/\mu_c$ i.e., the coefficient of variation (CV) is the same under both control and treatment.

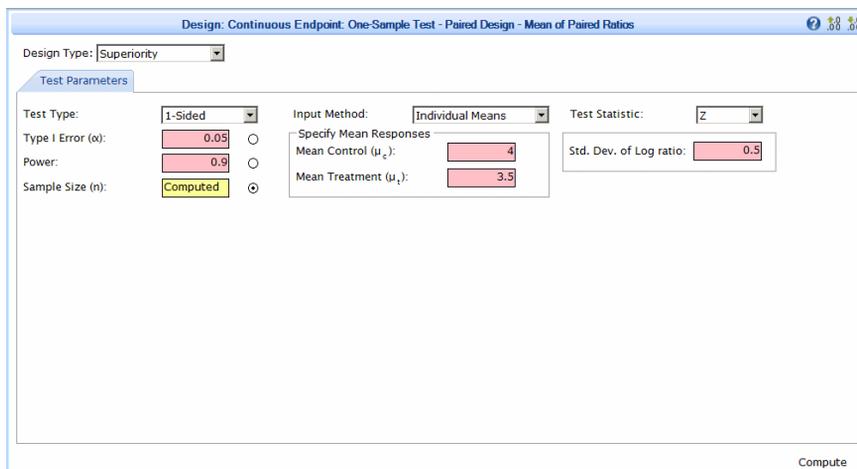
7.3.1 Trial Design

Start East afresh. Click **Continuous: One Sample** on the **Design** tab, and then click

Paired Design: Mean of Paired Ratios as shown below.



This will launch a new window. The upper pane of this window displays several fields with default values. Select **Test Type** as **1-Sided**, and **Individual Means** for **Input Method**. Specify the **Mean Control** (μ_c) as 4 and **Mean Treatment** (μ_t) as 3.5. Enter **Std. Dev. of Log ratio** as 0.5. The upper pane should appear as below:



Click **Compute**. This will calculate the sample size for this design and the output is shown as a row in the **Output Preview**. The computed sample size is 121 subjects (or pairs of observations).

This design has default name Des 1. In the **Output Preview** toolbar select Des 1, click

7 Normal Superiority One-Sample



to save this design to Wbk1 in the **Library**.

7.3.2 Trial Design Using a t-test

Right-click Des 1 in the **Library** and select **Edit Design**. In the input window, change the **Test Stat.** from **Z** to **t**.

Click **Compute**. East will add an additional row to the **Output Preview** labeled as Des 2. Select the rows corresponding to Des 1 and Des 2 using the Ctrl key and click



. This will display Des 1 and Des 2 in the **Output Summary**.

| | Des 1 | Des 2 |
|----------------------------|------------------|------------------|
| Mnemonic | MN-1S-PRA | MN-1S-PRA |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.902 | 0.901 |
| Model Parameters | | |
| Mean Control (μ_c) | 4 | 4 |
| Mean Treatment (μ_t) | 3.5 | 3.5 |
| Std.Dev. of Log Ratio | 0.5 | 0.5 |
| Input Method | Individual Means | Individual Means |
| Test Statistic | Z | t |
| Sample Size | | |
| Maximum | 121 | 122 |

Des 2 uses the t distribution and requires that we commit a combined total of 122 patients to the study, one more compared to Des 1, which uses a normal distribution.

8 Normal Noninferiority Paired-Sample

Two common applications of the paired sample design include: (1) comparison of two treatments where patients are matched on demographic and baseline characteristics, and (2) two observations made from the same patient under different experimental conditions. The type of endpoint for paired noninferiority design could be difference of means or ratio of means. The former is presented in Section 8.1 and the latter is discussed in Section 8.2. For paired sample noninferiority trials, East can be used only when no interim look is planned.

8.1 Mean of Paired Differences

8.1.1 Trial Design

8.1.2 Trial Design Using a t-Test (Single Look)

8.1.3 Simulation

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of outcome variable, X, with means μ_t and μ_c , respectively, and with a standard deviation of paired difference as σ_D^2 . Here, the null hypothesis $H_0: \mu_t - \mu_c \leq \delta_0$ is tested against the one-sided alternative hypothesis $H_1: \mu_t - \mu_c > \delta_0$. Here δ_0 denotes the noninferiority margin and $\delta_0 < 0$. Let $\delta = \mu_t - \mu_c$. Then the null hypothesis can be expressed as $H_0: \delta \leq \delta_0$ and the alternative can be expressed as $H_1: \delta > \delta_0$.

Here we assume that the each paired observation on X from T and C are distributed according to a bivariate normal distribution with means as (μ_t, μ_c) , variances as (σ_t^2, σ_c^2) and correlation coefficient as ρ . Let us have N such paired observations from T and C and $\hat{\mu}_c$ and $\hat{\mu}_t$ denote the estimates of μ_c and μ_t based on these N pairs. Therefore, the estimate of the difference is $\hat{\delta} = \hat{\mu}_t - \hat{\mu}_c$. Denoting the standard error of $\hat{\delta}$ by $se(\hat{\delta})$, the test statistic can be defined as

$$Z = \frac{\hat{\delta} - \delta_0}{se(\hat{\delta})} \tag{8.1}$$

The test statistic Z is distributed as a t distribution with $(N - 1)$ degrees of freedom. For large samples, the t-distribution can be approximated by the standard normal distribution. The power of the test is computed at specified values of μ_c, μ_t , and σ_D . East allows you to analyze using both normal and t distribution.

The advantage of the paired sample noninferiority design compared to the two independent sample noninferiority design lies in the smaller $se(\hat{\delta})$ in former case. The paired sample design is more powerful than the two independent sample design: to achieve the same level of power, the paired sample design requires fewer subjects.

8.1.1 Trial Design

Iezzi et. al. (2011) investigated the possibility of reducing radiation dose exposure

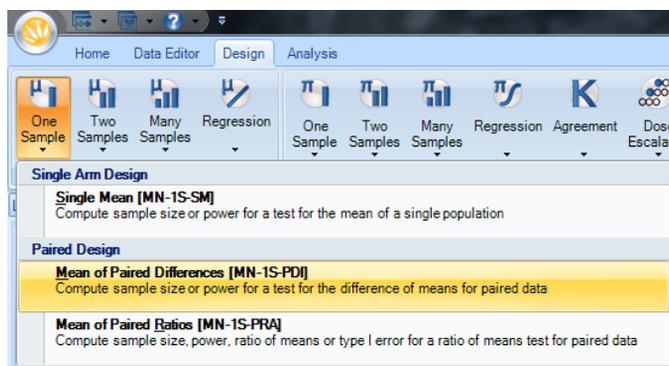
8 Normal Noninferiority Paired-Sample

while maintaining the image quality in a prospective, single center, intra-individual study. In this study, patients underwent two consecutive multidetector computed tomography angiography (MDCTA) scans 6 months apart, one with a standard acquisition protocol (C) and another using a low dose protocol (T). Image quality was rated as an ordinal number using a rating scale ranging from 1 to 5. Let μ_c and μ_t denote the average rating of image quality for standard acquisition and low dose protocol, respectively, and $\delta = \mu_t - \mu_c$ be the difference between two means. Based on the 30 samples included in the study, μ_c and μ_t were estimated as 3.67 and 3.12, respectively. The noninferiority margin for image quality considered was -1 . Accordingly, we will design the study to test

$$H_0 : \delta \leq -1 \quad \text{against} \quad H_1 : \delta > -1$$

The standard deviation of paired difference was estimated as 0.683. We want to design a study with 90% power at $\mu_c = 3.67$ and $\mu_t = 3.12$ and that maintains overall one-sided type I error of 0.025.

First, click **Continuous: One Sample** on the **Design** tab and then click **Paired Design: Mean of Paired Differences** as shown below.



This will launch a new window. Select **Noninferiority** for **Design Type**, and **Individual Means** for **Input Method**. Specify the **Mean Control** (μ_c) as 3.67, **Mean Treatment** (μ_t) as 3.12, and the **Std. Dev. of Paired Difference** (σ_D) as 0.683. Finally, enter -1 for the **Noninferiority Margin** (δ_0). Leave all other entries with their

default values. The upper pane should appear as below:

Click **Compute**. This will calculate the sample size for this design and the output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed sample size (25 subjects) is highlighted.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | Sample Size | Test Statistic | Input Method | μ_c | Mean Treatment (Alt.) | δ_1 | δ_0 | σ_D |
|------|----------------|--------------|-----------|--------------------|-------|-------------|----------------|------------------|---------|-----------------------|------------|------------|------------|
| Des1 | Noninferiority | 1 | 1-Sided | 0.025 | 0.909 | 25 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -1 | 0.683 |

This design has default name Des 1. You can select this design by clicking anywhere along the row in the **Output Preview**. Select this design and click  in the **Output Preview** toolbar. Some of the design details will be displayed in the upper

8 Normal Noninferiority Paired-Sample

pane, labeled as **Output Summary**.

| Des 1 | |
|--------------------------------------|------------------|
| Mnemonic | MN-1S-PDI |
| Test Parameters | |
| Design Type | Noninferiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.909 |
| Model Parameters | |
| Test Statistic | Z |
| Input Method | Individual Means |
| Mean Control (μ_c) | 3.67 |
| Mean Treatment (μ_t) | 3.12 |
| Diff. of Means ($\mu_t - \mu_c$) | -0.55 |
| Noninferiority Margin (δ_0) | -1 |
| Std. Deviation (σ_D) | 0.683 |
| Sample Size | |
| Maximum | 25 |

A total of 25 subjects must be enrolled in order to achieve the desired 90% power under the alternative hypothesis. In the **Output Preview** select Des 1 and click  in the toolbar to save this design to Wbk1 in the **Library**.

The noninferiority margin of -1 considered above is the minimal margin. Since the observed difference is only little less than -0.5 we would like to calculate sample size for a range of noninferiority margins, say, -0.6 , -0.7 , -0.8 , -0.9 and -1 . This can be done easily in East. First select Des 1 in the **Library**, and click  on the **Library** toolbar. In the Input, change the **Noninferiority Margin** (δ_0) $-0.6 : -1 : -0.1$.

Specify Null Hypothesis

Noninferiority Margin (δ_0):

Click **Compute** to generate sample sizes for different noninferiority margins. This will add 5 new rows to the **Output Preview**. There will be a single row for each of the

noninferiority margins.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | Sample Size | Test Statistic | Input Method | μ | Mean Treatment (Alt.) | δ_1 | δ_0 | σ_D |
|------|----------------|--------------|-----------|--------------------|-------|-------------|----------------|------------------|-------|-----------------------|------------|------------|------------|
| Des1 | Noninferiority | 1 | 1-Sided | 0.025 | 0.909 | 25 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -1 | 0.683 |
| Des2 | Noninferiority | 1 | 1-Sided | 0.025 | 0.9 | 1961 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -0.6 | 0.683 |
| Des3 | Noninferiority | 1 | 1-Sided | 0.025 | 0.9 | 218 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -0.7 | 0.683 |
| Des4 | Noninferiority | 1 | 1-Sided | 0.025 | 0.902 | 79 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -0.8 | 0.683 |
| Des5 | Noninferiority | 1 | 1-Sided | 0.025 | 0.907 | 41 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -0.9 | 0.683 |
| Des6 | Noninferiority | 1 | 1-Sided | 0.025 | 0.909 | 25 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -1 | 0.683 |

The computed sample sizes are 1961, 218, 79, 41 and 25 with noninferiority margins -0.60 , -0.7 , -0.8 , -0.9 and -1 , respectively. To compare all 5 designs, select last 5 rows in **Output Preview**, and click . The 5 designs will be displayed in the **Output Summary** pane.

| | Des2 | Des3 | Des4 | Des5 | Des6 |
|--------------------------------------|------------------|------------------|------------------|------------------|------------------|
| Mnemonic | MN-1S-PDI | MN-1S-PDI | MN-1S-PDI | MN-1S-PDI | MN-1S-PDI |
| Test Parameters | | | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 1 | 1 | 1 | 1 |
| Test Type | 1-Sided | 1-Sided | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 |
| Power | 0.9 | 0.9 | 0.902 | 0.907 | 0.909 |
| Model Parameters | | | | | |
| Test Statistic | Z | Z | Z | Z | Z |
| Input Method | Individual Means |
| Mean Control (μ) | 3.67 | 3.67 | 3.67 | 3.67 | 3.67 |
| Mean Treatment (μ) | 3.12 | 3.12 | 3.12 | 3.12 | 3.12 |
| Diff. of Means ($\mu - \mu$) | -0.55 | -0.55 | -0.55 | -0.55 | -0.55 |
| Noninferiority Margin (δ_0) | -0.6 | -0.7 | -0.8 | -0.9 | -1 |
| Std. Deviation (σ_D) | 0.683 | 0.683 | 0.683 | 0.683 | 0.683 |
| Sample Size | | | | | |
| Maximum | 1961 | 218 | 79 | 41 | 25 |

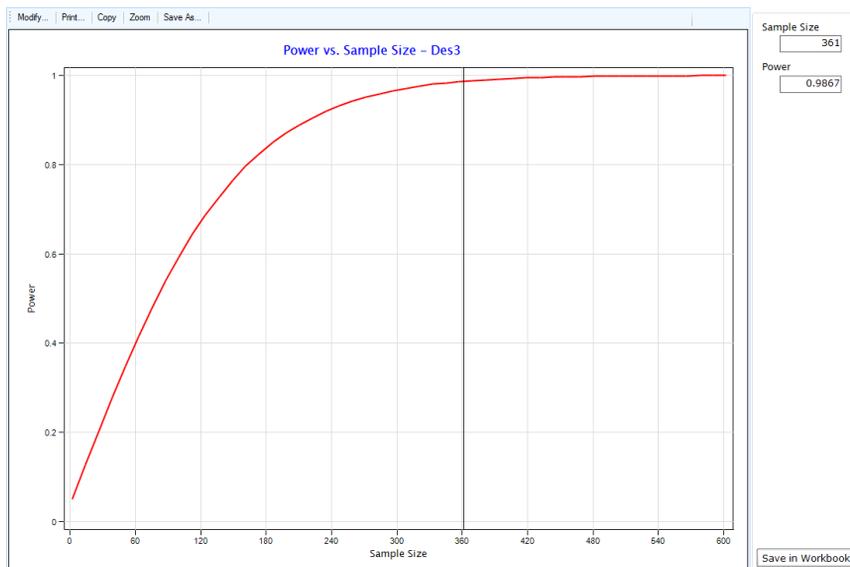
Suppose we have decided to go with Des 3 to test the noninferiority hypothesis with noninferiority margin of -0.7 . This requires a total sample size of 218 to achieve 90% power. Select Des 3 in the **Output Preview** and click in the toolbar to save this design to Wbk1 in the **Library**. Before we proceed we would like to delete all designs from the **Output Preview**. Select all rows and then either click in the toolbar, or click **Delete** after right click. To delete the designs from the workbook in **Library** select the corresponding designs individually (one at a time) and then click **Delete** after right click. You can try deleting Des 1 from the **Library**.

Plotting

With Des 3 selected in the **Library**, click on the **Library** toolbar, and then

8 Normal Noninferiority Paired-Sample

click **Power vs Sample Size**. The resulting power curve for this design will appear.



You can move the vertical bar along the X axis. To find out power at any sample size, move the vertical bar to that sample size and the numerical value of sample size and power will be displayed on the right of the plot. You can export this chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...**. Close this chart before continuing. In a similar fashion one can see power vs delta plot by clicking  and then **Power vs Treatment Effect**.

You can obtain the tables associated with these plot by clicking , and then clicking the appropriate table. Close the plots before continuing.

8.1.2 Trial Design Using a t-Test (Single Look)

The sample size obtained to correctly power Des 3 relied on using a Wald-type statistic for the hypothesis test. Due to the assumption of a normal distribution for the test statistic, we have ignored the fact that the variance σ is estimated from the sample. For large sample sizes, this approximation is acceptable. However, in small samples with unknown standard deviation, the test statistic

$$Z = (\hat{\delta} - \delta_0) / se(\hat{\sigma})$$

is distributed as Student's t distribution with $(n - 1)$ degrees of freedom where n is the

number of paired observations.

Select Des 3 from the **Library**, and click . This will take you to the input window. Now change the **Test Statistic** from **z** to **t**. The entries for the other fields need not be changed.

Click **Compute**. East will add an additional row to the **Output Preview**. The required sample size is 220. This design uses the t distribution and it requires us to commit a combined total of 220 patients to the study, two more compared to Des 3 which uses the normal distribution. The extra couple of patients are needed to compensate for the extra variability due to estimation of the var[$\hat{\delta}$].

8.1.3 Simulation

Select Des 3 in the **Library**, and click  in the toolbar. Alternatively, right-click on Des 3 and select **Simulate**. A new Simulation window will appear. Click on the **Response Generation Info** tab, and specify: **Mean control** = 3.67, **Mean Treatment** = 3.12, and **Std. Deviation of Paired Difference** (σ_D) = 0.683.

| Simulation Parameters | Response Generation Info | Simulation Control Info |
|-----------------------------|-----------------------------------|---|
| Specify Mean Responses | | |
| Mean Control (μ_c): | <input type="text" value="3.67"/> | Std. Dev. of Paired Difference (σ_D): <input type="text" value="0.683"/> |
| Mean Treatment (μ_t): | <input type="text" value="3.12"/> | |

Leave all default values, and click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1.

Select Sim 1 in the **Output Preview** and click . Double-click Sim 1 in the **Library**, and the simulation output details will be displayed in the right pane under the

8 Normal Noninferiority Paired-Sample

Simulation tab.

Simulation: Continuous Endpoint: One-Sample Test - Paired Design - Mean of Paired Differences

| Simulation Parameters | |
|--------------------------------|----------------|
| Simulation ID | Sim1 |
| Design Type | Noninferiority |
| Number of Looks | 1 |
| Test Type | 1-Sided |
| Sample Size (n) | 218 |
| Noninf. Margin (δ_c) | -0.7 |
| Test Statistic | t |
| Avg. Power at Termination | 0.899 |
| Response Generation Parameters | |
| Mean Control (μ_c) | 3.67 |
| Mean Treatment (μ_t) | 3.12 |
| Std. Deviation (σ_D) | 0.683 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 218 |
| Average | 218 |

Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|----------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 218 | Upper | 8986 | 10000 | 100.000% |
| Total | | | 8986 | 10000 | |
| % | | | 89.860% | | |

Notice that the percentage of rejections out of 10000 simulated trials is consistent with the design power of 90%. The exact result of the simulations may differ slightly, depending on the seed.

Now we wish to simulate from a point that belongs to H_0 to check whether the chosen design maintains type I error of 5%. Right-click Sim 1 in the **Library** and select **Edit Simulation**. Go to the **Response Generation Info** tab in the upper pane and specify: **Mean control = 3.67**, **Mean Treatment = 2.97**, and **Std. Deviation of Paired Difference (σ_D) = 0.683**.

| Simulation Parameters | Response Generation Info | Simulation Control Info |
|-----------------------------|-----------------------------------|---|
| Specify Mean Responses | | |
| Mean Control (μ_c): | <input type="text" value="3.67"/> | Std. Dev. of Paired Difference (σ_D): <input type="text" value="0.683"/> |
| Mean Treatment (μ_t): | <input type="text" value="2.97"/> | |

Click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 2. Select Sim 2 in the **Output Preview** and click . Now double-click on Sim 2 in the **Library**. The simulation output

details will be displayed.

Simulation: Continuous Endpoint: One-Sample Test - Paired Design - Mean of Paired Differences

| Simulation Parameters | |
|--------------------------------|----------------|
| Simulation ID | Sim2 |
| Design Type | Noninferiority |
| Number of Looks | 1 |
| Test Type | 1-Sided |
| Sample Size (n) | 218 |
| Noninf. Margin (δ_y) | -0.7 |
| Test Statistic | t |
| Avg. Power at Termination | 0.023 |
| Response Generation Parameters | |
| Mean Control (μ_c) | 3.67 |
| Mean Treatment (μ_t) | 2.97 |
| Std. Deviation (σ_c) | 0.683 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 10000 |

⊖ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 218 |
| Average | 218 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|----------|
| | | Efficacy | Stopping For | Count | % |
| 1 | 218 | 1.96 | Efficacy | 10000 | 100.000% |
| Total | | | 233 | 10000 | |
| % | | | 2.330% | | |

The upper efficacy stopping boundary was crossed close to the specified type I error of 2.5%. The exact result of the simulations may differ slightly, depending on the seed.

8.2 Ratio of Paired Means

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of outcome variable, X, with means μ_t and μ_c , respectively, and let σ_t^2 and σ_c^2 denote the respective variances. The null hypothesis $H_0: \mu_t/\mu_c \leq \rho_0$ is tested against the one-sided alternative hypothesis $H_1: \mu_t/\mu_c > \rho_0$. Here, ρ_0 denotes the noninferiority margin and $\rho_0 < 1$. Let $\rho = \mu_t/\mu_c$. Then the null hypothesis can be expressed as $H_0: \rho \leq \rho_0$ and the alternative can be expressed as $H_1: \rho > \rho_0$.

Let us have N such paired observations from T and C and (X_{it}, X_{ic}) denotes the i th pair of observations ($i = 1, \dots, N$). Then $\log X_{it} - \log X_{ic} = \log (X_{it}/X_{ic})$ denotes the logarithm of ratio of means for i th subject. We assume that the paired log-transformed observations on X from T and C, $(\log X_{it}, \log X_{ic})$ are bivariate normally distributed with common parameters. In other words, (X_{it}, X_{ic}) is distributed as bivariate log-normal distribution.

Denote $\log X_{it}$ by y_{it} , $\log X_{ic}$ by y_{ic} , and the corresponding difference by $\delta_{yi} = y_{it} - y_{ic}$. Assume that $\hat{\delta}_y$ denotes the sample mean for these paired differences with estimated standard error $se(\hat{\delta}_y)$. The test statistic can be defined as

$$Z = \frac{\hat{\delta}_y - \log \rho_0}{se(\hat{\delta}_y)}, \tag{8.2}$$

8 Normal Noninferiority Paired-Sample

The test statistic Z is distributed as a t distribution with $(N - 1)$ degrees of freedom. For large samples, the t-distribution can be approximated by the standard normal distribution. East allows you to analyze using both normal and t distribution. The power of the test is computed at specified values of μ_c , μ_t , and σ .

8.2.1 Trial Design

We will use the same example cited in the previous section, but will transform the difference hypothesis into the ratio hypothesis. Let μ_c and μ_t denote the average rating of image quality for standard acquisition and low dose protocol, estimated as 3.67 and 3.12, respectively. Let $\rho = \mu_t/\mu_c$ be the ratio between two means. Considering a noninferiority margin of -0.7 for the test of difference, we can rewrite the hypothesis mentioned in previous section as

$$H_0 : \rho \leq 0.81 \quad \text{against} \quad H_1 : \rho > 0.81$$

We are considering a noninferiority margin of $0.81 (= \rho_0)$. For illustration we will assume the standard deviation of log ratio as 0.20. As before, we want to design a study with 90% power at $\mu_c = 3.67$ and $\mu_t = 3.12$, and maintains overall one-sided type I error of 0.025.

Start East afresh. Click **Continuous: One Sample** on the **Design** tab and then click **Paired Design: Mean of Paired Ratios**.

This will launch a new window. The upper pane of this window displays several fields with default values. Select **Noninferiority** for **Design Type**, and **Individual Means** for **Input Method**. Specify the **Mean Control** (μ_c) as 3.67, **Mean Treatment** (μ_t) as 3.12, and **Noninferiority margin** (ρ_0) as 0.81. Enter 0.20 for **Std. Dev. of Log Ratio**, and 0.025 for **Type I Error** (α). The upper pane now should appear as below:

Click **Compute**. This will calculate the sample size for this design and the output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed sample size (180 subjects) is highlighted in yellow.

| ▲ | ID | Design Type | Test Type | Specified α | Power | Input Method | Sample Size | Test Statistic | μ c | Mean Treatment (Alt.) | ρ 0 | Std Dev Log Ratio |
|---|------|----------------|-----------|--------------------|-------|------------------|-------------|----------------|---------|-----------------------|----------|-------------------|
| ■ | Des1 | Noninferiority | 1-Sided | 0.025 | 0.9 | Individual Means | 180 | Z | 3.67 | 3.12 | 0.81 | 0.2 |

This design has default name Des 1. You can select this design by clicking anywhere along the row in the **Output Preview**. Select this design and click  in the **Output Preview** toolbar. Some of the design details will be displayed in the upper pane, labeled as **Output Summary**.

| Des 1 | |
|-----------------------------------|------------------|
| Mnemonic | MN-15-PRA |
| Test Parameters | |
| Design Type | Noninferiority |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Mean Control (μ c) | 3.67 |
| Mean Treatment (μ t) | 3.12 |
| Noninferiority Margin (ρ 0) | 0.81 |
| Std.Dev. of Log Ratio | 0.2 |
| Input Method | Individual Means |
| Test Statistic | Z |
| Sample Size | |
| Maximum | 180 |

A total of 180 subjects must be enrolled in order to achieve the desired 90% power under the alternative hypothesis. In the **Output Preview** select Des 1 and click  in the toolbar to save this design to Wbk1 in the **Library**.

Suppose you think enrolling 180 subjects is too much for your organization and you can go up to only 130 subjects. You want to evaluate the power of your study at sample size 130 but with the design parameters remain unaltered. In order to compute power with 130 subjects, first select the Des 1 in the **Library**, and click  on the **Library** toolbar. In the Input dialog box, first select the radiobutton for **Power**, and

8 Normal Noninferiority Paired-Sample

then enter 130 for **Sample Size**.

Design Type: Noninferiority

Design Parameters

Test Type: 1-Sided

Type I Error (α): 0.025

Power: Computed

Sample Size (n): 130

Now click **Compute**. This will add another row labeled as Des 2 in **Output Preview** with computed power highlighted in yellow. The design attains a power of 78.7%.

Now select both the rows in **Output Preview** by pressing the Ctrl key, and click  in the **Output Preview** toolbar to see a summary of both designs in the **Output Summary**.

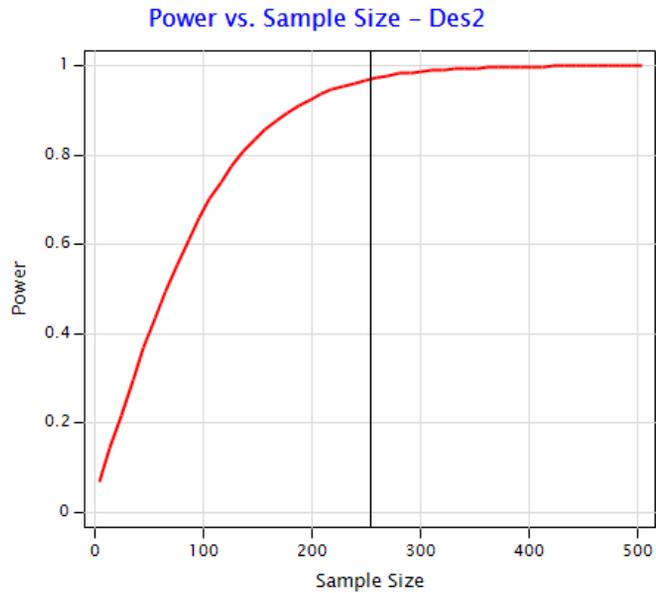
| | Des 1 | Des 2 |
|---------------------------------|------------------|------------------|
| Mnemonic | MN-1S-PRA | MN-1S-PRA |
| Test Parameters | | |
| Design Type | Noninferiority | Noninferiority |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.9 | 0.787 |
| Model Parameters | | |
| Mean Control (μ_c) | 3.67 | 3.67 |
| Mean Treatment (μ_t) | 3.12 | 3.12 |
| Noninferiority Margin (p_0) | 0.81 | 0.81 |
| Std.Dev. of Log Ratio | 0.2 | 0.2 |
| Input Method | Individual Means | Individual Means |
| Test Statistic | Z | Z |
| Sample Size | | |
| Maximum | 180 | 130 |

In the **Output Preview** select Des 2 and click  in the toolbar to save this design to Wbk1 in the **Library**.

Plotting

With Des 2 selected in the **Library**, click  on the **Library** toolbar, and then click **Power vs Sample Size**. The resulting power curve for this design will appear.

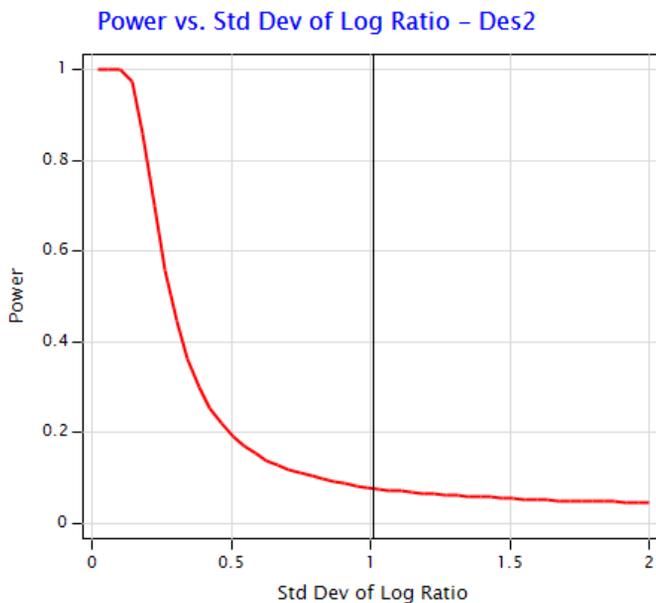
You can move the vertical bar along the X axis.



Suppose you would like to explore the relationship between power and standard deviation. In order to visualize this relationship, select Des 2 in the **Library**, click  on the **Library** toolbar, and then click **General (User Defined Plot)**. Select **Std Dev**

8 Normal Noninferiority Paired-Sample

of **Log Ratio** for **X-Axis**. This will display the power vs. standard deviation plot.



Close the plot window before you continue.

8.2.2 Simulation

Select Des 2 in the **Library**, and click  in the toolbar. Alternatively, right-click on Des 2 and select **Simulate**. A new Simulation window will appear. Click on the **Response Generation Info** tab, and specify: **Mean control** = 3.67, **Mean Treatment** = 3.12, and **Std Dev of Log Ratio** = 0.2.

| Simulation Parameters | Response Generation Info | Simulation Control Info |
|----------------------------|-----------------------------------|---|
| Mean Control(μ_c): | <input type="text" value="3.67"/> | Std Dev. of Log Ratio: <input type="text" value="0.2"/> |
| Mean Treatment(μ_t): | <input type="text" value="3.12"/> | |

Click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1.

Select Sim 1 in the **Output Preview** and click  . Now double-click on Sim 1 in the **Library**. The simulation output details will be displayed.

Simulation: Continuous Endpoint: One-Sample Test - Paired Design - Mean of Paired Ratios

| Simulation Parameters | |
|---|----------------|
| Simulation ID | Sim1 |
| Trial Type | Noninferiority |
| Test Type | 1-Sided |
| Sample Size (n) | 130 |
| Test Statistic | t |
| Noninferiority Margin (ρ_0) | 0.81 |
| Response Generation Parameters | |
| Mean Response under Control (μ_c) | 3.67 |
| Mean Response under Treatment (μ_t) | 3.12 |
| Simulation Std. Dev. of Log Ratio | 0.2 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Simulation Boundaries

Critical Point: 1.96

Overall Simulation Results

| | Upper H0 | Lower H0 |
|-------------------|----------|----------|
| No. of Rejections | 7832 | NA |
| % | 78.32 | NA |

Starting Seed: 6641254
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:05

9

Normal Equivalence Paired-Sample

Two common applications of the paired sample designs include: (1) comparison of two treatments where patients are matched on demographic and baseline characteristics, and (2) two observations made from the same patient under different experimental conditions. The type of endpoint for paired equivalence design may be a difference of means or ratio of means. The former is presented in Section 9.1 and the latter is discussed in Section 9.2.

9.1 Mean of Paired Differences

9.1.1 Trial Design

9.1.2 Simulation

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of a outcome variable, X, with means μ_t and μ_c , respectively, and with a standard deviation of paired difference as σ_D^2 . Here, the null hypothesis $H_0: \mu_t - \mu_c < \delta_L$ or $\mu_t - \mu_c > \delta_U$ is tested against the two-sided alternative hypothesis $H_1: \delta_L \leq \mu_t - \mu_c \leq \delta_U$. Here, δ_L and δ_U denote the equivalence limits. The two one-sided tests (TOST) procedure of Schuirmann (1987) is commonly used for this analysis.

Let $\delta = \mu_t - \mu_c$ denotes the true difference in the means. The null hypothesis $H_0: \delta \leq \delta_L$ or $\delta \geq \delta_U$ is tested against the two-sided alternative hypothesis $H_1: \delta_L < \delta < \delta_U$ at level α , using TOST procedure. Here, we perform the following two tests together:

- Test1: $H_{0L}: \delta \leq \delta_L$ against $H_{1L}: \delta > \delta_L$ at level α
- Test2: $H_{0U}: \delta \geq \delta_U$ against $H_{1U}: \delta < \delta_U$ at level α

H_0 is rejected in favor of H_1 at level α if and only if both H_{0L} and H_{0U} are rejected. Note that this is the same as rejecting H_0 in favor of H_1 at level α if the $(1 - 2\alpha)$ 100% confidence interval for δ is completely contained within the interval (δ_L, δ_U) .

Here we assume that the each paired observation on X from T and C are bivariate normally distributed with parameters $\mu_t, \mu_c, \sigma_t^2, \sigma_c^2$ and ρ . Let us have N such paired observations from T and C, and let $\hat{\mu}_c$ and $\hat{\mu}_t$ denote the estimates of μ_c and μ_t based on these N pairs. The estimate of the difference is $\hat{\delta} = \hat{\mu}_t - \hat{\mu}_c$. Denoting the standard error of $\hat{\delta}$ by $se(\hat{\delta})$, test statistics for Test1 and Test2 are defined as:

$$T_L = \frac{(\hat{\delta} - \delta_L)}{se(\hat{\delta})} \quad \text{and} \quad T_U = \frac{(\hat{\delta} - \delta_U)}{se(\hat{\delta})}$$

T_L and T_U are assumed to follow Student's t-distribution with $(N - 1)$ degrees of freedom under H_{0L} and H_{0U} , respectively. H_{0L} is rejected if $T_L \geq t_{1-\alpha, (N-1)}$, and H_{0U} is rejected if $T_U \leq t_{\alpha, (N-1)}$.

The null hypothesis H_0 is rejected in favor of H_1 if $T_L \geq t_{1-\alpha, (N-1)}$ and $T_U \leq t_{\alpha, (N-1)}$, or in terms of confidence intervals: Reject H_0 in favor of H_1 at level α if

$$\delta_L + t_{1-\alpha, (N-1)} se(\hat{\delta}) < \hat{\delta} < \delta_U + t_{\alpha, (N-1)} se(\hat{\delta}) \quad (9.1)$$

We see that decision rule (9.1) is the same as rejecting H_0 in favor of H_1 if the $(1 - 2\alpha)$ 100% confidence interval for δ is entirely contained within interval (δ_L, δ_U) .

The power or sample size of such a trial design is determined for a specified value of δ , say δ_1 , for a single-look study only. The choice $\delta_1 = 0$, i.e. $\mu_t = \mu_c$, is common. For a specified value of δ_1 , the power is given by

$$\Pr(\text{Reject } H_0) = 1 - \tau_\nu(t_{\alpha, \nu} | \Omega_1) + \tau_\nu(-t_{\alpha, \nu} | \Omega_2) \quad (9.2)$$

where $\nu = N - 1$ and Ω_1 and Ω_2 are non-centrality parameters given by $\Omega_1 = (\delta_1 - \delta_L) / se(\hat{\delta})$ and $\Omega_2 = (\delta_1 - \delta_U) / se(\hat{\delta})$, respectively. $t_{\alpha, \nu}$ denotes the upper $\alpha \times 100\%$ percentile from a Student's t distribution with ν degrees of freedom. $\tau_\nu(x | \Omega)$ denotes the distribution function of a non-central t distribution with ν degrees of freedom and non-centrality parameter Ω , evaluated at x .

Since the sample size N is not known ahead of time, we cannot characterize the bivariate t -distribution. Thus, solving for sample size must be performed iteratively by equating the formula (9.2) to the power $1 - \beta$.

The advantage of the paired sample equivalence design compared to the two sample equivalence design lies in the smaller $se(\hat{\delta})$ in former case. The paired sample equivalence design is more powerful than the two sample equivalence design: to achieve the same level of power, the paired sample equivalence design requires fewer subjects.

9.1.1 Trial Design

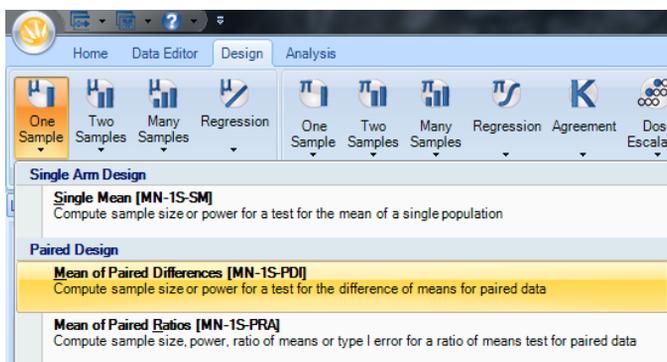
To ensure that comparable results can be achieved between two laboratories or methods, it is important to conduct cross-validation or comparability studies to establish statistical equivalence between the two laboratories or methods. Often, to establish equivalence between two laboratories, a paired sample design is employed. Feng et al. (2006) reported the data on 12 quality control (QC) samples. Each sample was analyzed first by Lab1 and then by Lab2. In this example we will consider Lab1 as the standard laboratory (C) and Lab2 is the one to be validated (T). Denote the mean concentrations from Lab1 and Lab2 by μ_c and μ_t , respectively. Considering an equivalence limit of $(-10, 10)$ we can state our hypotheses as:

$$H_0: \mu_t - \mu_c < -10 \text{ or } \mu_t - \mu_c > 10 \text{ against } H_1: -10 \leq \mu_t - \mu_c \leq 10$$

9 Normal Equivalence Paired-Sample

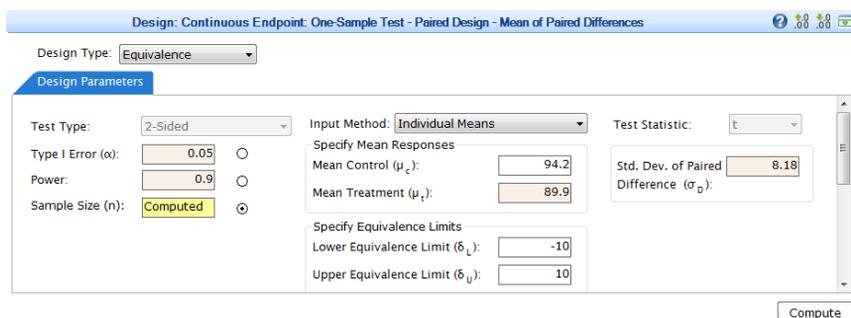
Based on the reported data μ_c and μ_t are estimated as 94.2 pg mL^{-1} and 89.9 pg mL^{-1} , respectively. The standard deviation of paired difference was estimated as 8.18. We want to design a study with 90% power at $\mu_c = 94.2$ and $\mu_t = 89.9$. We want to reject H_0 with type I error not exceeding 0.025.

First, click **Continuous: One Sample** on the **Design** tab, and then click **Paired Design: Mean of Paired Differences** as shown below.



This will launch a new window.

Since we are interested in testing an equivalence hypothesis select **Equivalence** for **Trial Type**, with an **Type I Error** of 0.025, and **Power** of 0.9. Select **Individual Means** for **Input Method**. Enter -10 for **Lower Equivalence Limit** (δ_L) and 10 for **Upper Equivalence Limit** (δ_U). Specify the **Mean Control** (μ_c) as 94.2, **Mean Treatment** (μ_t) as 89.9, and **Std. Dev. of Paired Difference** (σ_D) as 8.18. The upper pane should appear as below:



Click **Compute**. This will calculate the sample size for this design and the output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed sample size (20 samples) is highlighted in yellow.

| ▲ | ID | Design Type | Test Type | Specified α | Power | Sample Size | Test Statistic | Input Method | μ_c | Mean Treatment (Alt.) | δ_L | δ_U | δ_U | σ_D |
|---|------|-------------|-----------|--------------------|-------|-------------|----------------|------------------|---------|-----------------------|------------|------------|------------|------------|
| ■ | Des1 | Equivalence | 2-Sided | 0.05 | 0.913 | 20 | t | Individual Means | 94.2 | 89.9 | -4.3 | -10 | 10 | 8.18 |

This design has default name Des 1 and you can select this design by clicking anywhere along the row in the **Output Preview** and then clicking  in the **Output Preview** toolbar. Some of the design details will be displayed in the upper pane, labeled as **Output Summary**.

| Output Summary | |
|--|------------------|
| | Des 1 |
| Mnemonic | MN-15-PDI |
| Test Parameters | |
| Design Type | Equivalence |
| No. of Looks | 1 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.913 |
| Model Parameters | |
| Test Statistic | t |
| Input Method | Individual Means |
| Mean Control (μ_c) | 94.2 |
| Mean Treatment (μ_t) | 89.9 |
| Diff. of Means ($\mu_t - \mu_c$) | -4.3 |
| Equivalence Lower Limit (δ_L) | -10 |
| Equivalence Upper Limit (δ_U) | 10 |
| Std. Deviation (σ_D) | 8.18 |
| Sample Size | |
| Maximum | 20 |

A total of 20 samples is required to achieve the desired 90% power under the alternative hypothesis. In the **Output Preview** select Des 1 and click  in the toolbar to save this design to Wbk1 in the **Library**.

The equivalence limits of $(-10, 10)$ might be too narrow and therefore a wider equivalence interval $(-12.5, 12.5)$ could be considered. Select Des 1 in the **Library**, and click  on the **Library** toolbar. In the **Design Parameters** tab, change the entry for **Lower Equivalence Limit (δ_L)** and **Upper Equivalence Limit (δ_U)** to -12.5 and 12.5 , respectively, and click **Compute**.

This will add a new row in the **Output Preview** labeled as Des 2. In the **Output**

9 Normal Equivalence Paired-Sample

Preview select Des 2 and click  in the toolbar to save this design to Wbk1 in the **Library**. To compare the two designs, select both rows in **Output Preview** using the Ctrl key and click  in the **Output Preview** toolbar. This will display the two designs side by side in the **Output Summary** pane.

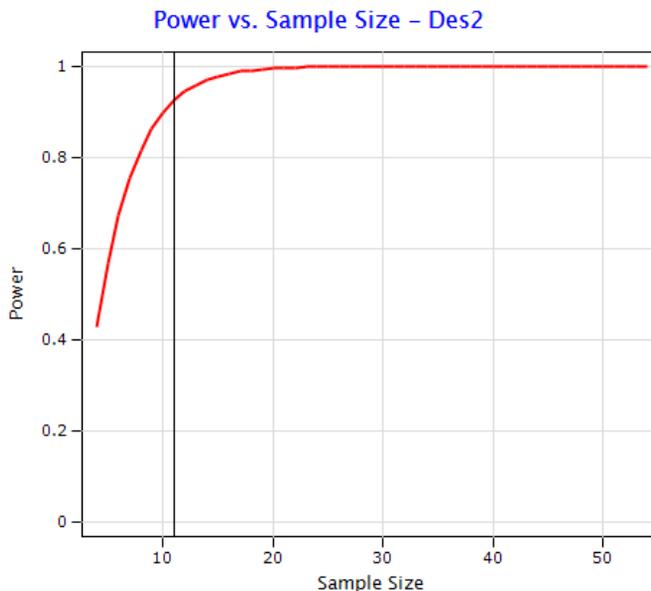
| | Des 1 | Des 2 |
|--|------------------|------------------|
| Mnemonic | MN-1S-PDI | MN-1S-PDI |
| Test Parameters | | |
| Design Type | Equivalence | Equivalence |
| No. of Looks | 1 | 1 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.913 | 0.926 |
| Model Parameters | | |
| Test Statistic | t | t |
| Input Method | Individual Means | Individual Means |
| Mean Control (μ_C) | 94.2 | 94.2 |
| Mean Treatment (μ_T) | 89.9 | 89.9 |
| Diff. of Means ($\mu_T - \mu_C$) | -4.3 | -4.3 |
| Equivalence Lower Limit (δ_L) | -10 | -12.5 |
| Equivalence Upper Limit (δ_U) | 10 | 12.5 |
| Std. Deviation (σ_D) | 8.18 | 8.18 |
| Sample Size | | |
| Maximum | 20 | 11 |

As we widen the equivalence limit from $(-10, 10)$ to $(-12.5, 12.5)$, the required sample size is reduced from 20 to 11.

Plotting

We would like to explore how power is related to the required sample size. Select Des 2 in the **Library**, click  on the **Library** toolbar, and then click **Power vs**

Sample Size. The resulting power curve for this design will appear.

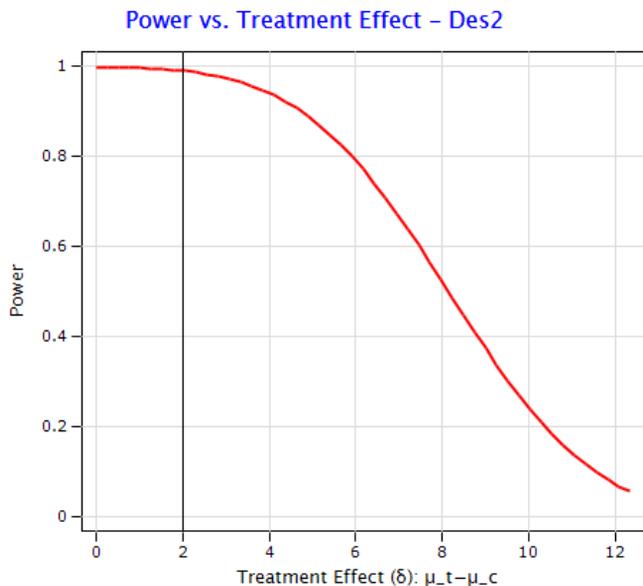


You can move the vertical bar along the X axis. To find out power at any sample size simply move the vertical bar to that sample size and the numerical value of sample size and power will be displayed on the right of the plot. You can export this chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...** Close this chart before continuing.

In a similar fashion one can see power vs delta plot by clicking  and then

9 Normal Equivalence Paired-Sample

Power vs Treatment Effect.



To produce tables associated with these plots, first click  in the toolbar and then select the appropriate table.

9.1.2 Simulation

Now we wish to simulate from Des 2 to verify whether the study truly maintains the power and type I error. Select Des 2 in the **Library**, and click  in the toolbar. Alternatively, right-click on Des 2 and select **Simulate**. Click on the **Response Generation Info** tab, and specify: **Mean control** = 94.2, **Mean Treatment** = 89.9, and **Std. Dev. of Paired Difference** (σ_D) = 8.18.

| Simulation Parameters | Response Generation Info | Simulation Control Info |
|-----------------------------|--|-----------------------------------|
| Specify Mean Responses | | |
| Mean Control (μ_c): | <input type="text" value="94.2"/> | |
| Mean Treatment (μ_t): | <input type="text" value="89.9"/> | |
| | Std. Dev. of Paired Difference (σ_D): | <input type="text" value="8.18"/> |

Click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1.

Select Sim 1 in the **Output Preview** and click  icon. Now double-click on Sim 1 in the **Library**. The simulation output details will be displayed.

Simulation: Continuous Endpoint: One-Sample Test - Paired Design - Mean of Paired Differences

| Simulation Parameters | |
|--|-------------|
| Simulation ID | Sim1 |
| Design Type | Equivalence |
| Number of Looks | 1 |
| Test Type | 2-Sided |
| Sample Size (n) | 11 |
| Test Statistic | t |
| Lower Equivalence Limit (δ_L) | -12.5 |
| Upper Equivalence Limit (δ_U) | 12.5 |
| Avg. Power at Termination | 0.928 |
| Response Generation Parameters | |
| Mean Control (μ_c) | 94.2 |
| Mean Treatment (μ_t) | 89.9 |
| Std. Deviation (σ_c) | 8.18 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 11 |
| Average | 11 |

Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|--------|----------------|----------------|-------------------|----------|
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 11 | 1.812 | -1.812 | 9275 | 0 | 10000 | 100.000% |
| Total | | | | 9275 | 0 | 10000 | |
| % | | | | 92.750% | 0.000% | | |

Notice that the simulated power is close to the attained power of 92.6% for Des 2. The exact result of the simulations may differ slightly, depending on the seed.

Now we wish to simulate from a point that belongs to H_0 to check whether the chosen design maintains type I error of 5% or not. For this we consider, $\mu_c = 94.2$ and $\mu_t = 81.7$. Since in this case $\delta = 81.7 - 94.2 = -12.5$, this $(\mu_t, \mu_c) = (81.7, 94.2)$ point belongs to H_0 . Right-click on Sim 1 in the **Library** and select **Edit Simulation**. Go to the **Response Generation Info** tab in the upper pane and specify: **Mean control** = 94.2, **Mean Treatment** = 81.7, and **Std. Dev. of Paired Difference** (σ_D) = 8.18.

Simulation Parameters
Response Generation Info
Simulation Control Info

Specify Mean Responses

Mean Control (μ_c):

Mean Treatment (μ_t):

Std. Dev. of Paired Difference (σ_D):

Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 2. Select Sim 2 in the **Output Preview** and click  icon. Now double-click on Sim 2 in the **Library**.

9 Normal Equivalence Paired-Sample

The simulation output details will be displayed in the right pane under **Simulation** tab.

Simulation: Continuous Endpoint: One-Sample Test - Paired Design - Mean of Paired Differences

| Simulation Parameters | |
|--|-------------|
| Simulation ID | Sim2 |
| Design Type | Equivalence |
| Number of Looks | 1 |
| Test Type | 2-Sided |
| Sample Size (n) | 11 |
| Test Statistic | t |
| Lower Equivalence Limit (δ_L) | -12.5 |
| Upper Equivalence Limit (δ_U) | 12.5 |
| Avg. Power at Termination | 0.051 |
| Response Generation Parameters | |
| Mean Control (μ_c) | 94.2 |
| Mean Treatment (μ_t) | 81.7 |
| Std. Deviation (σ_c) | 8.18 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 11 |
| Average | 11 |

Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|--------|----------------|----------------|-------------------|----------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 11 | 1.812 | -1.812 | 508 | 0 | 10000 | 100.000% |
| Total | | | | 508 | 0 | 10000 | |
| % | | | | 5.080% | 0.000% | | |

Notice that the simulated power here is close to the pre-set type I error of 5%. The exact result of the simulations may differ slightly, depending on the seed.

9.2 Ratio of Paired Means

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of a outcome variable, X, with means μ_t and μ_c , respectively, and let σ_t^2 and σ_c^2 denote the respective variances. Here, the null hypothesis $H_0: \mu_t/\mu_c \leq \rho_L$ or $\mu_t/\mu_c \geq \rho_U$ is tested against the alternative hypothesis $H_1: \rho_L < \mu_t/\mu_c < \rho_U$. Let $\rho = \mu_t/\mu_c$ denotes the ratio of two means. Then the null hypothesis can be expressed as $H_0: \rho \leq \rho_L$ or $\rho \geq \rho_U$ and the alternative can be expressed as $H_1: \rho_L < \rho < \rho_U$. In practice, ρ_L and ρ_U are often chosen such that $\rho_L = 1/\rho_U$. The two one-sided tests (TOST) procedure of Schuirmann (1987) is commonly used for this analysis, and is employed in this section for a parallel-group study.

Let us have N such paired observation from T and C and (X_{it}, X_{ic}) denotes the i th pair of observations ($i = 1, \dots, N$). Then $\log X_{it} - \log X_{ic} = \log (X_{it}/X_{ic})$ denotes the logarithm of ratio of means for the i th subject. Here we assume that the each paired log-transformed observations on X from T and C, $(\log X_{it}, \log X_{ic})$ are bivariate normally distributed with common parameters. In other words, (X_{it}, X_{ic}) is distributed as a bivariate log-normal distribution.

Since we have translated the ratio hypothesis into a difference hypothesis using the log transformation, we can perform the test for difference as discussed in section 9.1. Note that we need the standard deviation of log of ratios. Sometimes, we are provided with information on coefficient of variation (CV) of ratios instead, and the standard

deviation of log ratios can be obtained using: $sd = \sqrt{\ln(1 + CV^2)}$.

This is a test for the comparison of geometric means of ratio, as we are taking the mean of the logarithms of ratios.

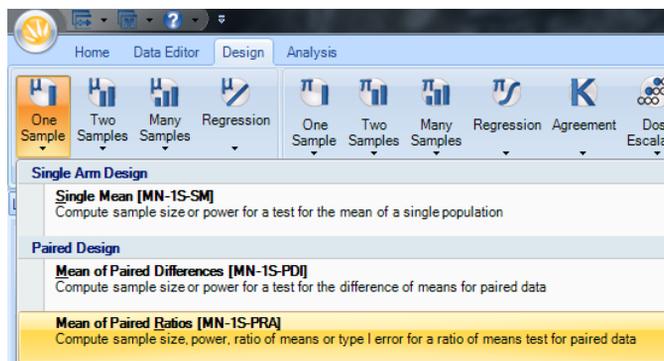
9.2.1 Trial Design

Here we will use the same example reported by Feng et al (2006). Denote the mean concentrations from Lab1 and Lab2 by μ_c and μ_t , and $\rho = \mu_t/\mu_c$ is the ratio between two means. Considering an equivalence limit of (0.85, 1.15) we can state our hypotheses as

$$H_0: \mu_t/\mu_c < 0.85 \text{ or } \mu_t/\mu_c > 1.15 \text{ against } H_1: 0.85 \leq \mu_t/\mu_c \leq 1.15$$

Based on the reported data, μ_c and μ_t are estimated as 94.2 pg mL⁻¹ and 89.9 pg mL⁻¹, respectively. Assume that the standard deviation of log ratio can be estimated is 0.086. As before, we want to design a study with 90% power at $\mu_c = 94.2$ and $\mu_t = 89.9$. We want to reject H_0 with type I error not exceeding 0.025.

Start East afresh. First, click **Continuous: One Sample** on the **Design** tab and then click **Paired Design: Mean of Paired Ratios** as shown below.



This will launch a new window.

Select **Equivalence** for **Trial Type**, and enter 0.025 for **Type I Error**, and 0.9 for **Power**. Then select **Individual Means** for **Input Method**, and enter the **Mean Control** (μ_c) as 94.2, **Mean Treatment** (μ_t) as 89.9, and **Std. Dev. of Log Ratio** as 0.086. Enter 0.85 for **Lower Equiv. Limit** (ρ_L) and 1.15 for **Upper Equiv. Limit**

9 Normal Equivalence Paired-Sample

(ρ_U). The upper pane should appear as below:

Click **Compute**. This will calculate the sample size for this design and the output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed sample size (8 samples) is highlighted in yellow.

| ID | Design Type | Test Type | Specified α | Power | Input Method | Sample Size | Test Statistic | μ_C | Mean Treatment (Alt.) | ρ_L | ρ_U | Std Dev Log Ratio |
|------|-------------|-----------|--------------------|-------|------------------|-------------|----------------|---------|-----------------------|----------|----------|-------------------|
| Des1 | Equivalence | 2-Sided | 0.025 | 0.901 | Individual Means | 8 | t | 94.2 | 89.9 | 0.85 | 1.15 | 0.086 |

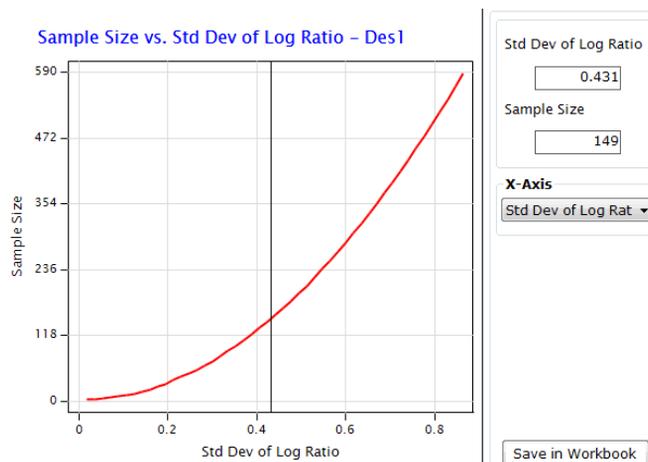
This design has default name Des 1. Select this design and click  in the **Output Preview** toolbar. Some of the design details will be displayed in the upper pane, labeled as **Output Summary**.

| Des1 | |
|--------------------------------------|------------------|
| Mnemonic | MN-1S-PRA |
| Test Parameters | |
| Design Type | Equivalence |
| Test Type | 2-Sided |
| Specified α | 0.025 |
| Power | 0.901 |
| Model Parameters | |
| Mean Control (μ_C) | 94.2 |
| Mean Treatment (μ_T) | 89.9 |
| Equivalence Lower Limit (ρ_L) | 0.85 |
| Equivalence Upper Limit (ρ_U) | 1.15 |
| Std.Dev. of Log Ratio | 0.086 |
| Input Method | Individual Means |
| Test Statistic | t |
| Sample Size | |
| Maximum | 8 |

In the **Output Preview** select Des 1 and click  in the toolbar to save this design to Wbk1 in the **Library**.

Plotting

Suppose you want to see how the standard deviation influences the sample size. In order to visualize this relationship, select Des 1 in the **Library**, click  on the **Library** toolbar, and then click **General (User Defined Plot)**. Select **Std Dev of Log Ratio** for **X-Axis** in right of the plot. This will display the sample size vs. standard deviation plot.



Close this plot before continuing.

9.2.2 Simulation

Now we want to check by simulation whether the sample size of 8 provides at least 90% power. Select Des 1 in the **Library**, and click  in the toolbar. Click on the **Response Generation Info** tab, and specify: **Mean control** = 94.2, **Mean Treatment** = 89.9, and **Std Dev. of Log Ratio**= 0.086.

| Simulation Parameters | Response Generation Info | Simulation Control Info |
|----------------------------|--------------------------|------------------------------|
| Mean Control(μ_c): | 94.2 | Std Dev. of Log Ratio: 0.086 |
| Mean Treatment(μ_t): | 89.9 | |

9 Normal Equivalence Paired-Sample

Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1. Notice that the simulated power is very close to the design power.

| ▲ | ID | Design Type | Test Type | Specified α | Power | Input Method | Sample Size | Test Statistic | μ c | Mean Treatment (Alt.) | ρ L | ρ U | Std Dev Log Ratio |
|---|-------|-------------|-----------|--------------------|-------|------------------|-------------|----------------|---------|-----------------------|----------|----------|-------------------|
| # | Des 1 | Equivalence | 2-Sided | 0.025 | 0.901 | Individual Means | 8 | t | 94.2 | 89.9 | 0.85 | 1.15 | 0.086 |
| # | Sim 1 | Equivalence | 2-Sided | | 0.902 | | | t | 94.2 | 89.9 | 0.85 | 1.15 | 0.086 |

10 Normal Superiority Two-Sample

To demonstrate the superiority of a new treatment over the control, it is often necessary to randomize subjects to the control and treatment arms, and contrast the group-dependent means of the outcome variables. In this chapter, we show how East supports the design and interim monitoring of such experiments.

10.1 Difference of Means

10.1.1 Trial Design (Weight

Control Trial of Orlistat)

10.1.2 IM of the Orlistat trial

10.1.3 t-Test Design

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of a normally distributed outcome variable, X , with means μ_t and μ_c , respectively, and with a common variance σ^2 . We intend to monitor the data up to K times after accruing $n_1, n_2, \dots, n_K \equiv n_{\max}$ patients. The information fraction at the j th look is given by $t_j = n_j/n_{\max}$. Let r denote the fraction randomized to treatment T.

Define the treatment difference to be

$$\delta = \mu_t - \mu_c .$$

The null hypothesis of interest is

$$H_0 : \delta = 0 .$$

We wish to construct a K -look group sequential level α test of H_0 having $1 - \beta$ power at the alternative hypothesis

$$H_1 : \delta = \delta_1 .$$

Let $\bar{X}_t(t_j)$ and $\bar{X}_c(t_j)$ be the mean responses of the experimental and control groups, respectively, at time t_j . Then

$$\hat{\delta}(t_j) = \bar{X}_t(t_j) - \bar{X}_c(t_j) \tag{10.1}$$

and

$$\text{var}[\hat{\delta}(t_j)] = \frac{\sigma^2}{n_j r(1-r)} . \tag{10.2}$$

Therefore, by the Scharfstein, Tsiatis and Robins (1997), Jennison and Turnbull (1997) theorem the stochastic process

$$W(t_j) = \sqrt{t_j} \frac{\bar{X}_t(t_j) - \bar{X}_c(t_j)}{\sqrt{\frac{\sigma^2}{n_j r(1-r)}}}, \quad j = 1, 2, \dots, K, \tag{10.3}$$

is $N(\eta t_j, t_j)$ with independent increments, where $\eta = 0$ under H_0 and $\eta = \delta_1 \sqrt{I_{\max}}$ under H_1 . We refer to η as the drift parameter.

10 Normal Superiority Two-Sample

10.1.1 Trial Design (Weight Control Trial of Orlistat)

Eighteen U.S. research centers participated in this trial, where obese adults were randomized to receive either Orlistat or placebo, combined with a dietary intervention for a period of two years (Davidson et al, 1999). Orlistat is an inhibitor of fat absorption, and the trial was intended to study its effectiveness in promoting weight loss and reduce cardiovascular risk factors. The study began in October 1992. More than one outcome measure was of interest, but we shall consider only body weight changes between baseline and the end of the first year intervention. We shall consider a group sequential design even though the original study was not intended as such. The published report does not give details concerning the treatment effect of interest or the desired significance level and power of the test. It does say, however, that 75% of subjects had been randomized to the Orlistat arm, probably to maximize the number of subjects receiving the active treatment.

Single-Look Design Suppose that the expected mean body weight change after one year of treatment was 9 kg in the Orlistat arm and 6 kg in the control arm. Assume also that the common standard deviation of the observations (weight change) was 8 kg. The standardized difference of interest would therefore be $(9 - 6)/8 = 0.375$. We shall consider a one sided test with 5% significance level and 90% power, and an allocation ratio (treatment:control) of 3:1; that is, 75% of the patients are randomized to the Treatment (Orlistat) arm.

First, click **Continuous: Two Samples** on the **Design** tab, and then click **Parallel Design: Difference of Means**.

In the upper pane of this window is the Input dialog box, which displays default input values. The effect size can be specified in one of three ways, selected from **Input Method**: (1) individual means and common standard deviation, (2) difference of means and common standard deviation, or (3) standardized difference of means. We will use the **Individual Means** method. Enter the appropriate design parameters so that the dialog box appears as shown. Remember to set the **Allocation Ratio** to 3.

Then click **Compute**.

The design is shown as a row in the **Output Preview**, located in the lower pane of this window. The computed sample size is 325 subjects.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | Input Method | δ | μ_c | Mean Treatment (Alt.) | σ | Test Statistic |
|------|-------------|--------------|-----------|--------------------|-------|-------|-------------|------------------|----------|---------|-----------------------|----------|----------------|
| Des1 | Superiority | 1 | 1-Sided | 0.05 | 0.9 | 3 | 325 | Individual Means | 3 | 6 | 9 | 8 | Z |

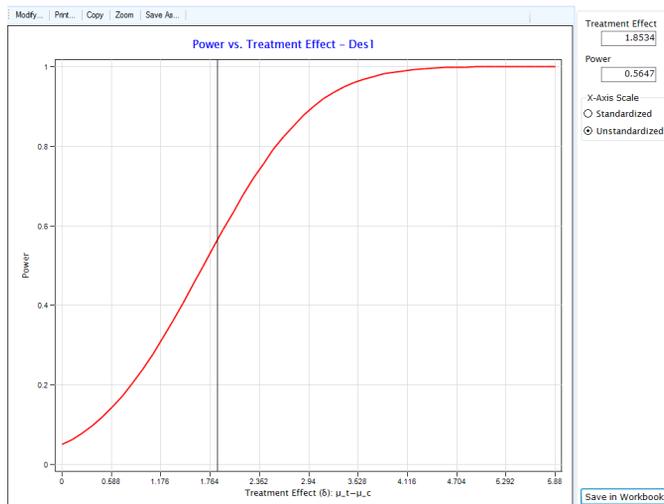
You can select this design by clicking anywhere along the row in the **Output Preview**.

On the **Output Preview** toolbar, click  to display a summary of the design details in the upper pane. Then, in the **Output Preview** toolbar, click  to save this design to Wbk1 in the **Library**. If you hover the cursor over Des1 in the **Library**, a tooltip will appear that summarizes the input parameters of the design.

With Des1 selected in the **Library**, click  on the **Library** toolbar, and then click **Power vs Treatment Effect (δ)**. The resulting power curve for this design is shown. You can save this chart to the **Library** by clicking **Save in Workbook**. You can also export the chart in one of several image formats (e.g., Bitmap or JPEG) by

10 Normal Superiority Two-Sample

clicking **Save As...** For now, you may close the chart before continuing.



Three-Look Design Create a new design by selecting Des1 in the **Library**, and clicking  on the **Library** toolbar, or by right-clicking and selecting **Edit Design**. In the Input, change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. A new tab for **Boundary Info** should appear. Click this tab to reveal the stopping boundary parameters. By default, the **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The left side contains details for the **Efficacy** boundary, and the right side contains details for the **Futility** boundary. By default, there is an efficacy boundary (to reject H_0) selected, but no futility boundary (to reject H_1). The **Boundary Family** specified is of the **Spending Functions** type. The default **Spending Function** is the **Lan-DeMets** (Lan & DeMets, 1983), with **Parameter OF** (O'Brien-Fleming), which generates boundaries that are very similar, though not identical, to the classical stopping boundaries of O'Brien and Fleming (1979).

The cumulative alpha spent, and the boundary values, are displayed in the table below.

Design Type: Superiority Number of Looks: 3

Design Parameters **Boundary Info**

Efficacy
 Boundary Family: Spending Functions
 Spending Function: Lan-DeMets
 Parameter: OF
 Type I Error (α): 0.05

Futility
 Boundary Family: None

Spacing of Looks Equal Unequal Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.333 | 0.001 | 3.200 |
| 2 | 0.667 | 0.016 | 2.141 |
| 3 | 1.000 | 0.050 | 1.695 |

Expected sample size and stopping probabilities

Click **Compute** to generate output for Des2. Select both Des1 and Des2 in the **Output Preview** and click  . The maximum and expected sample sizes are highlighted in yellow.

| | Des1 | Des2 |
|--|------------------|------------------|
| Mnemonic | MN-2S-DI | MN-2S-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Input Method | Individual Means | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 3 | 3 |
| Mean Control (μ_c) | 6 | 6 |
| Mean Treatment (μ_t) | 9 | 9 |
| Std. Deviation (σ) | 8 | 8 |
| Test Statistic | Z | Z |
| Allocation Ratio (n _t /n _c) | 3 | 3 |
| Boundary Parameters | | |
| Efficacy Boundary | | LD (OF) |
| Spacing of Looks | | Equal |
| Sample Size | | |
| Maximum | 325 | 331 |
| Expected Under H ₀ | | 329.115 |
| Expected Under H ₁ | | 256.591 |

The price to be paid for multiple looks is the commitment of a higher maximum sample size (331 patients) compared to that of a single-look design (325 patients). However, if the alternative hypothesis *H*₁ holds, the study has a chance of stopping at one of the two interim analyses and saving patient accrual: on average, Des2 will stop

10 Normal Superiority Two-Sample

with 257 patients if the alternative is true. The expected sample size under the null is 329, less than the maximum since there is a small probability of stopping before the last look and, wrongly, rejecting the null.

With Des2 selected in the **Output Preview**, click  to save Des2 to the **Library**. In order to see the stopping probabilities, as well as other characteristics, double-click Des2 in the **Library**. The clear advantage of this sequential design resides in the high probability of stopping by the second look, if the alternative is true, with a sample size of 221 patients, which is well below the requirements for a fixed sample study (325 patients). Even under the null, however, there is a small chance for the test statistic to cross the boundary for its early rejection (type-1 error probability) at the first or second look. Close the Details window before continuing.

Design: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means

| Test Parameters | |
|--------------------------------|------------------|
| Design ID | Des2 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Test Statistic | Z |
| Input Method | Individual Means |
| Mean Control (μ_c) | 6 |
| Mean Treatment (μ_t) | 9 |
| $\delta = \mu_t - \mu_c$ | |
| Under H0 | 0 |
| Under H1 | 3 |
| Std. Deviation (σ) | 8 |
| Allocation Ratio (n_t/n_c) | 3 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|-------------------------------|-------------|---------------|---------|
| Sample Size (n) | | | |
| Maximum | 83 | 248 | 331 |
| Expected H1 | 64.143 | 192.448 | 256.591 |
| Expected H0 | 82.521 | 246.594 | 329.115 |
| Maximum Information (I): 0.97 | | | |

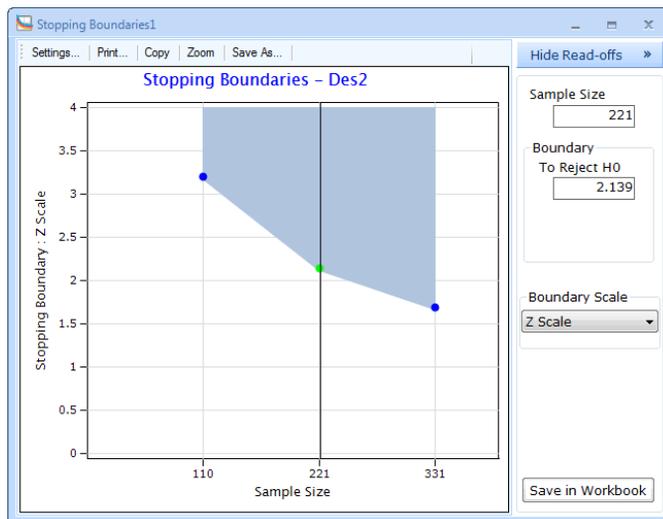
Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|------------|---|----------|
| | | | | | Under H0 | Under H1 |
| | | | | Efficacy Z | Efficacy | Efficacy |
| 1 | 0.332 | 110 | 6.741E-4 | 3.206 | 6.741E-4 | 0.066 |
| 2 | 0.668 | 221 | 0.016 | 2.139 | 0.016 | 0.543 |
| 3 | 1 | 331 | 0.05 | 1.695 | 0.034 | 0.291 |

Examining stopping boundaries and spending functions

Plot the boundary values of Des2 by clicking  on the **Library** toolbar, and then

selecting **Stopping Boundaries**. The following chart will appear:

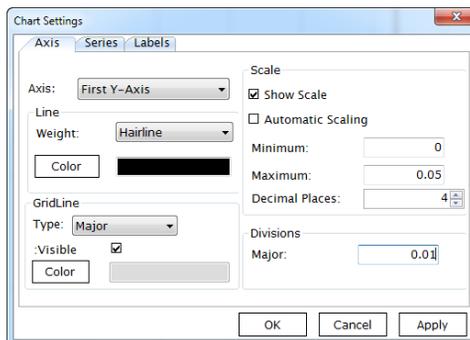


The three solid dots correspond to the actual boundary values to be used at the three planned analyses. Although the three looks are assumed to be equally spaced at design time, this assumption need not hold at analysis time. Values of the test-statistic (z-test) greater than the upper boundary values would warrant early stopping in favor of H_1 , that Orlistat is better than placebo. The horizontal axis expresses the total number of patients at each of the three analysis time-points. The study is designed so that the last analysis time point coincides with the maximum sample size required for the chosen design, namely 331 patients. By moving the vertical line cursor from left to right, one can observe the actual values of the stopping boundaries at each interim analysis time-point. The boundaries are rather conservative: for example, you would need the standardized test statistic to exceed 2.139 in order to stop the trial at the second look.

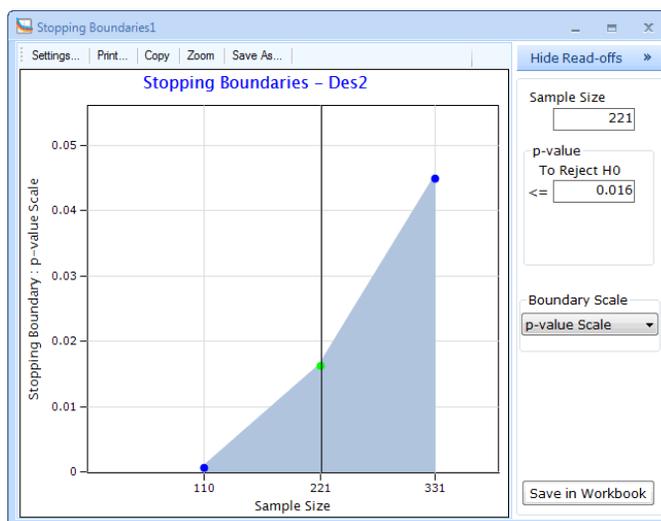
It is sometimes convenient to display the stopping boundaries on the p-value scale. Under **Boundary Scale**, select the **p-value Scale**. The chart now displays the cumulative number of patients on the X-axis and the nominal p-value (1-sided) that we would need in order to stop the trial at that interim look. To change the scale of this chart, click **Settings...** and in the **Chart Settings** dialog box, change the **Maximum** to

10 Normal Superiority Two-Sample

0.05, and the **Divisions: Major** to 0.01, and click **OK**.



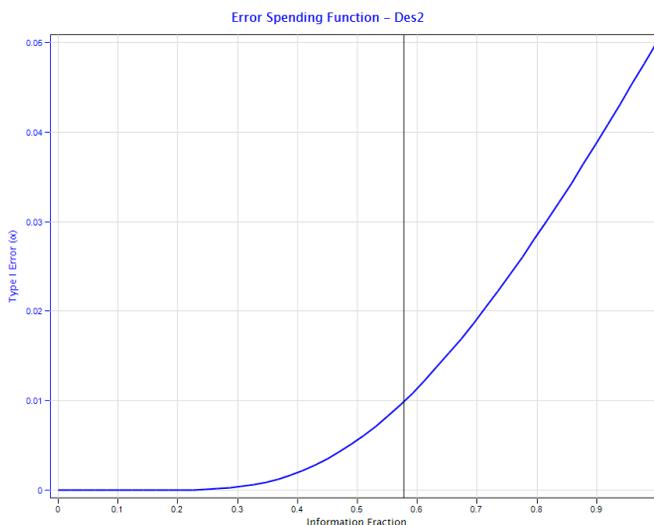
The following chart will be displayed.



For example, at the second look, after 221 subjects have been observed, we require a p-value smaller than 0.016 in order to stop the study. Notice that the p-value at the 3rd and final look needs to be smaller than 0.045, rather than the usual 0.05 that one would require for a single-look study. This is the penalty we pay for the privilege of taking three looks at the data instead of one. You may like to display the boundaries in the delta scale. In this scale, the boundaries are expressed in units of the effect size, or the difference in means. We need to observe a difference in average weight loss of 2.658

kg or more, in order to cross the boundary at the second look.

Close these charts, and click  and then **Error Spending**. The following chart will appear.



This spending function was proposed by Lan and DeMets (1983), and for one-sided tests has the following functional form:

$$\alpha(t) = 2 - 2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{t}}\right). \tag{10.4}$$

Observe that very little of the total type-1 error is spent early on, but more is spent rapidly as the information fraction increases, and reaches 0.05 at an information fraction of 1. A recursive method for generating stopping boundaries from spending functions is described in the Appendix G. Close this chart before continuing.

Lan and DeMets (1983) also provided a function for spending the type-1 error more aggressively. This spending function is denoted by **PK**, signifying that it is the Lan-DeMets spending function for generating stopping boundaries that closely resemble the classical Pocock (1977) stopping boundaries. It has the functional form:

$$\alpha(t) = \alpha \ln[1 + (e - 1)t] \tag{10.5}$$

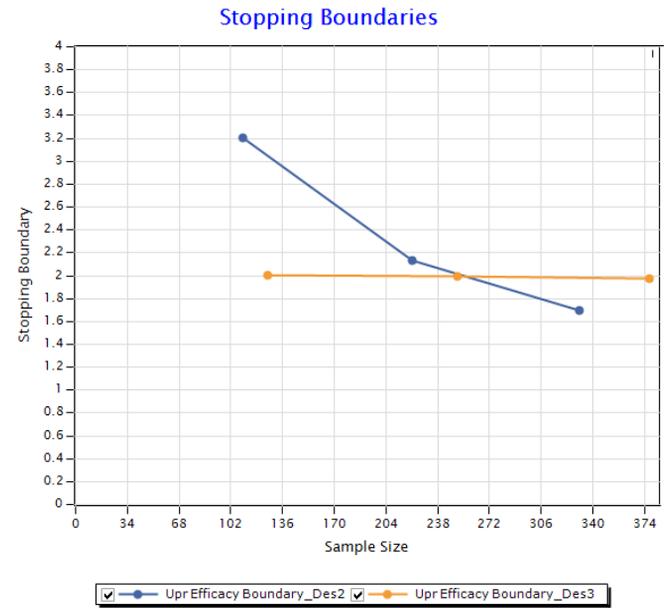
10 Normal Superiority Two-Sample

Select Des2 in the **Library**, and click  on the **Library** toolbar. On the **Boundary Info** tab, change the **Parameter** from **OF** to **PK**, and click **Compute**. With Des3 selected in the **Output Preview**, click . In the **Library**, select both Des2 and Des3, by holding the Ctrl key, and then click . The upper pane will display the details of the two designs side-by-side:

| | Des 2 | Des 3 |
|---|------------------|------------------|
| Mnemonic | MN-2S-DI | MN-2S-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 3 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Input Method | Individual Means | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 3 | 3 |
| Mean Control (μ_c) | 6 | 6 |
| Mean Treatment (μ_t) | 9 | 9 |
| Std. Deviation (σ) | 8 | 8 |
| Test Statistic | Z | Z |
| Allocation Ratio (nt/nc) | 3 | 3 |
| Boundary Parameters | | |
| Efficacy Boundary | LD (OF) | LD (PK) |
| Spacing of Looks | Equal | Equal |
| Sample Size | | |
| Maximum | 331 | 377 |
| Expected Under H0 | 329.115 | 369.359 |
| Expected Under H1 | 256.591 | 229.499 |

In the **Output Summary** toolbar, click  to compare the two designs according to **Stopping Boundaries**. Notice that the stopping boundaries for Des3 (PK) are relatively flat; almost the same critical point is used at all looks to declare significance.

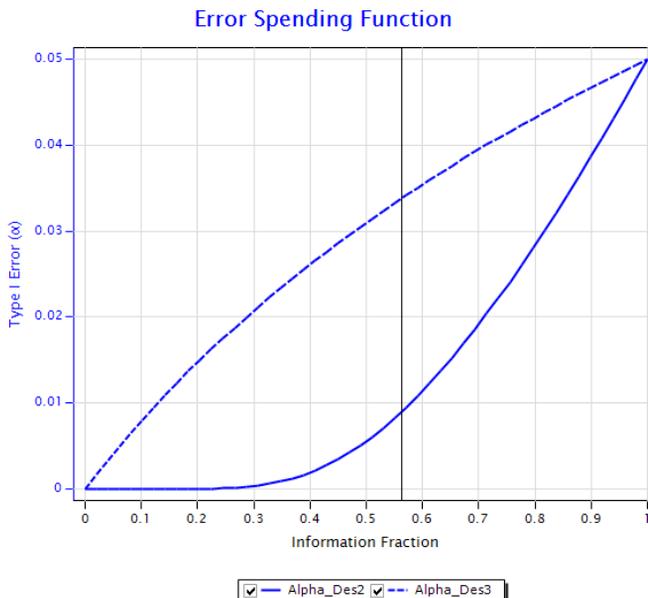
Close the chart before continuing.



Click  and select **Error Spending**. Des3 (PK) spends the type-1 error

10 Normal Superiority Two-Sample

probability at a much faster rate than Des2 (OF). Close the chart before continuing.



Wang and Tsiatis Power Boundaries

The stopping boundaries generated by the Lan-Demets **OF** and **PK** functions closely resemble closely the classical O’Brien-Fleming and Pocock stopping boundaries, respectively. These classical boundaries are a special case of a family of power boundaries proposed by Wang and Tsiatis (1987). For a two-sided level- ψ test, using K equally spaced looks, the power boundaries for the standardized test statistic Z_j at the j -th look are of the form

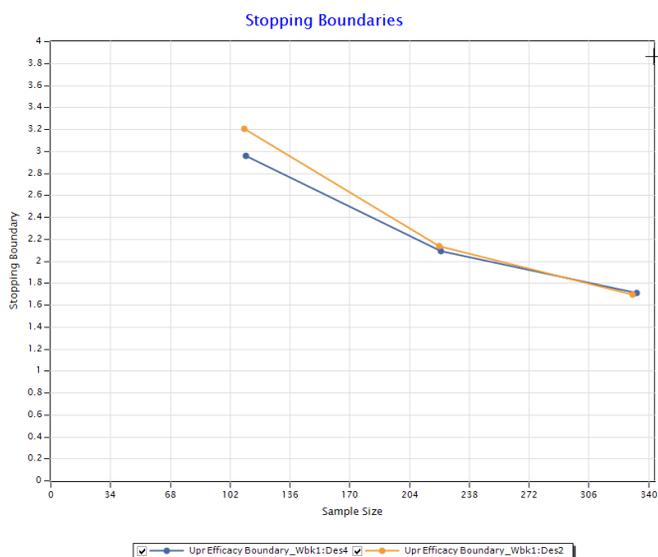
$$Z_j \geq \frac{C(\Delta, \alpha, K)}{(j/K)^{0.5-\Delta}} \tag{10.6}$$

The normalizing constant $C(\Delta, \alpha, K)$ is evaluated by recursive integration so as to ensure that the K -look group sequential test has type-1 error equal to α (see Appendix G for details), and Δ is a parameter characterizing the shape of the stopping boundary. For example, if $\Delta = 0.5$, the boundaries are constant at each of the K looks. These are the classical Pocock stopping boundaries (Pocock, 1977). If $\Delta = 0$, the width of the boundaries is inversely proportional to the square root of the information fraction j/K at the j -th look. These are the classical O’Brien-Fleming stopping boundaries (O’Brien and Fleming, 1979). Other choices produce boundaries of different shapes. Notice from equation (10.6) that power boundaries have a specific

functional form, and can be evaluated directly, or tabulated, once the normalizing constant $C(\Delta, \alpha, K)$ has been worked out for various combinations of α and K . In contrast, spending function boundaries are evaluated indirectly by inverting a pre-specified spending function as shown in Appendix F.

Right-click Des3 in the **Library** and select **Edit Design**. On the **Boundary Info** tab, change the **Boundary Family** from **Spending Functions** to **Wang-Tsiatis**. Leave the default value of Δ as 0, and click **Compute**. With Des4 selected in the **Output Preview**, click  .

In the **Library**, select both Des2 and Des4 by holding the Ctrl key. Click  and select **Stopping Boundaries**. As expected from our discussion above, the boundary values for Des2 (Lan-Demets, OF) and for Des4 (Wang-Tsiatis, $\Delta = 0$) are very similar. Close the chart before continuing.

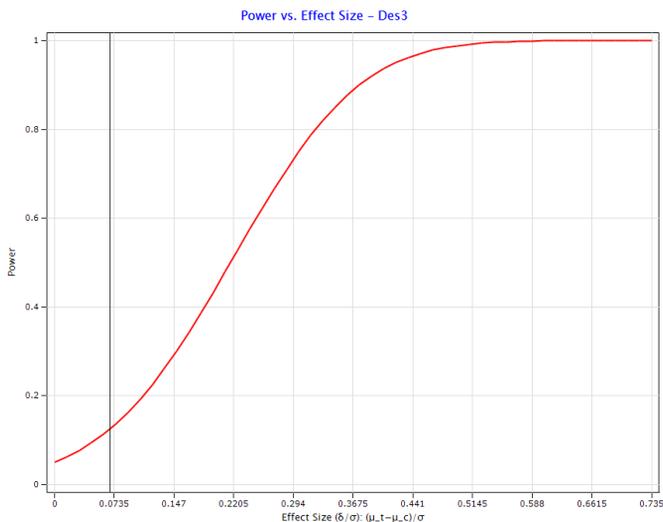


More charts

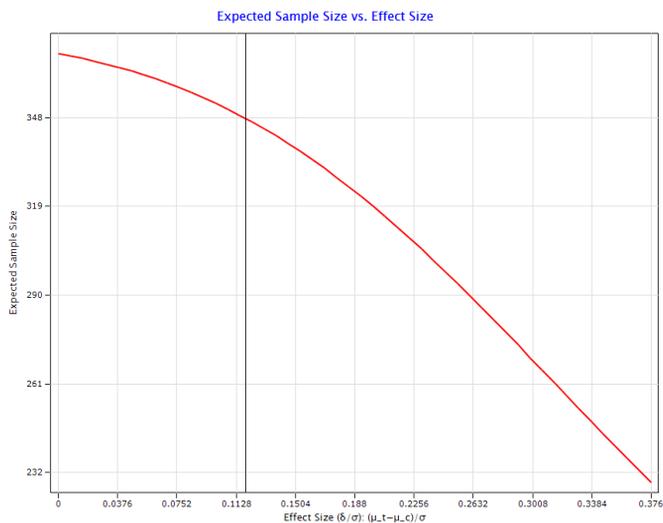
Select Des3 in the **Library**, click  , and then click **Power vs. Treatment effect** (δ). Click the radiobutton for **Standardized** under **X-Axis Scale**. By scrolling from left to right with the vertical line cursor, one can observe the power for various

10 Normal Superiority Two-Sample

values of the effect size.



Close this chart, and with Des3 selected, click  again. Then click **Expected Sample Size**. Click the radio button for **Standardized** under **X-Axis Scale**. The following chart appears:



By scrolling from left to right with the vertical line cursor we can observe how the expected number of events decreases as the effect size increases. Close this chart before continuing.

Unequally spaced analysis time points

In the above designs, we have assumed that analyses were equally spaced. This assumption can be relaxed if you know *when* interim analyses are likely to be performed (e.g., for administrative reasons). In either case, departures from this assumption are allowed during the actual interim monitoring of the study, but sample size requirements will be more accurate if allowance is made for this knowledge.

With Des3 selected in the **Library**, right-click **Edit Design**. Under **Spacing of Looks** in the **Boundary Info** tab, click the **Unequal** radio button.

The column titled **Info. Fraction** can be edited to modify the relative spacing of the analyses. The information fraction refers to the proportion of the maximum (yet unknown) sample size. By default, this table displays equal spacing, but suppose that the two interim analyses will be performed with 0.25 and 0.5 of the maximum sample size. Click **Recalc** to recompute the cumulative alpha spent and the efficacy boundary values.

Spacing of Looks Equal Unequal

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.25 | 0.018 | 2.100 |
| 2 | 0.5 | 0.031 | 2.077 |
| 3 | 1 | 0.050 | 1.913 |

After entering these new information fraction values, click **Compute**. Select Des5 in the **Output Preview** and click  to save it in the **Library** for now.

Arbitrary amounts of error probability to be spent at each analysis

Another feature of East is the possibility to specify arbitrary amounts of cumulative error probability to be used at each look. This option can be combined with the option of unequal spacing of the analyses. With Des5 selected in the **Library**, click  on the **Library** toolbar. Under the **Boundary Info** tab, select **Interpolated** for the **Spending Function**. In the column titled **Cum. α Spent**, enter 0.005 for the first look

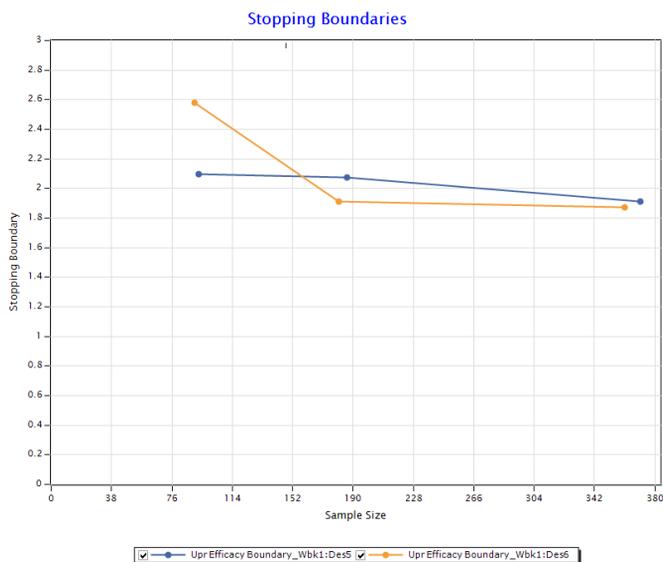
10 Normal Superiority Two-Sample

and 0.03 for the second look, click **Recalc**, and then **Compute**.

Spacing of Looks Equal Unequal

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.250 | 0.005 | 2.576 |
| 2 | 0.500 | 0.03 | 1.913 |
| 3 | 1.000 | 0.05 | 1.874 |

Select Des6 in the **Output Preview** and click . From the **Library**, select Des5 and Des6 by holding the Ctrl key. Click , and select **Stopping Boundaries**. The following chart will be displayed.



The advantage of Des6 over Des5 is the more conservative boundary (less type-1 error probability spent) at the first look. Close these charts before continuing.

Computing power for a given sample size

East can compute the achieved power, given the other design parameters such as sample size. Select Des6 in the **Library** and right-click **Edit Design**. On the **Design Parameters** tab, click the radio button for **Power**. You will notice that the field for power will contain the word “Computed”. You may now enter a value for the sample

size: Enter 250, and click **Compute**. As expected, the achieved power is less than 0.9, namely 0.781.

Test Type: 1-Sided
 Type I Error (α): 0.05
 Power: 0.781
 Sample Size (n): 250
 Allocation Ratio: 3
 (n_1/n_2)

To delete this design, click Des7 in the **Output Preview**, and click  in the toolbar. East will display a warning to make sure that you want to delete the selected row. Click **Yes** to continue.

Spending function boundaries for early stopping in favor of H_0 or H_1

So far we have considered only efficacy boundaries, which allow for early stopping in favor of the alternative. It may be of interest, in addition, to consider futility boundaries, which allow for early stopping when there is *lack* of evidence against the null hypothesis. Select Des2 in the **Library** and click . On the **Boundary Info** tab, you can select from one of several types of futility boundaries, such as from a spending function, or by conditional power. Note that some of these options are available for one-sided tests only.

Futility
 Boundary Family: Spending Functions
 Spending Function: Spending Functions
 Parameter: p-value
 Conditional Power
 δ/σ

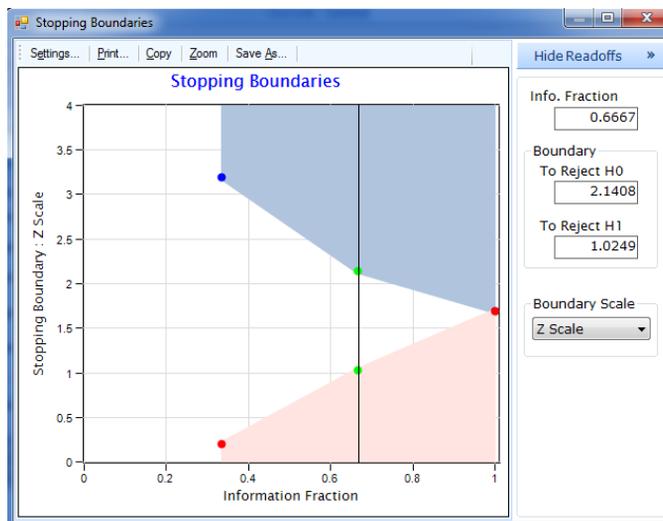
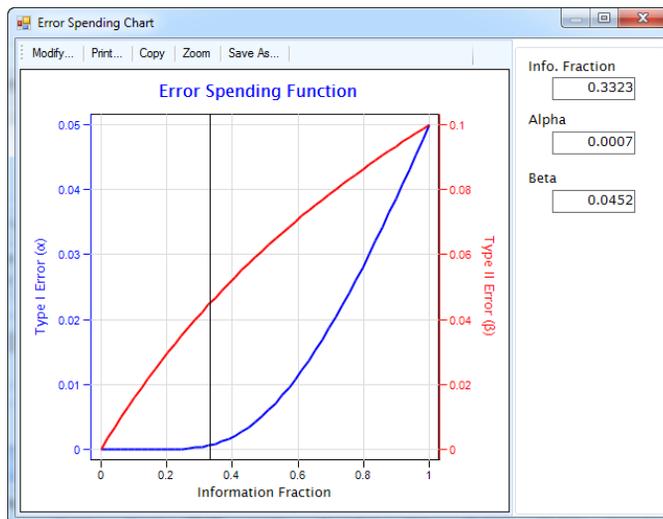
Select **Spending Functions** under **Boundary Family**. Select **PK** for the **Parameter**, and leave all other default settings. See the updated values of the stopping boundaries populated in the table below.

| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Cum. β Spent | Futility Boundary |
|--------|----------------|-------------------------------------|-------------------------------------|---------------------|-------------------|--------------------|-------------------|
| 1 | 0.333 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.0007 | 3.2001 | 0.0453 | 0.1974 |
| 2 | 0.667 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.0164 | 2.1408 | 0.0763 | 1.0249 |
| 3 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.0500 | 1.6948 | 0.1000 | 1.6948 |

On the **Boundary Info** tab, you may also like to click the  or  icons to

10 Normal Superiority Two-Sample

view plots of the error spending functions, or stopping boundaries, respectively.



Click **Compute**, and with Des7 selected in the **Output Preview**, click . To view the design details, double-click Des7 in the **Library**. Because not all the type-2 error is spent at the final look, this trial has a chance of ending early if the null

hypothesis is true. This is demonstrated by the low expected sample size under the null (209 patients), compared to those of the other designs considered so far. Close the Output window before continuing.

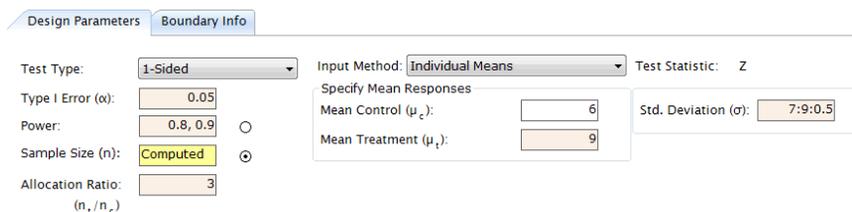
Before continuing to the next section, we will save the current workbook, and open a new workbook. Select Wbk1 in the **Library** and right-click, then click **Save**.

Next, click the  button, click **New**, and then **Workbook**. A new workbook, Wbk2, should appear in the **Library**. Delete all designs from the **Output Preview** before continuing.

Creating multiple designs

To create more than one design from the Input, one simply enters multiple values in any of the highlighted input fields. Multiple values can be entered in two ways. First, one can enter a comma-separated list (e.g., “0.8, 0.9”). Second, one can use colon notation (e.g., “7:9:0.5”) to specify a range of values, where “a:b:c” is read as from ‘a’ to ‘b’ in step size ‘c’.

Suppose that we wished to explore multiple variations of Des7. With Des7 selected in the **Library**, right-click and select **Edit Design**. In the **Design Parameters** tab of the Input, enter multiple values for the **Power(1-β)** (0.8, 0.9) and **Std.Deviation(σ)** (7 : 9 : 0.5) and click **Compute**:



The screenshot shows the 'Design Parameters' dialog box with the following settings:

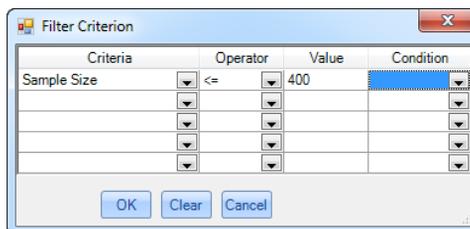
- Test Type: 1-Sided
- Type I Error (α): 0.05
- Power: 0.8, 0.9
- Sample Size (n): Computed
- Allocation Ratio: 3
- Input Method: Individual Means
- Mean Control (μ_c): 6
- Mean Treatment (μ_t): 9
- Std. Deviation (σ): 7:9:0.5
- Test Statistic: Z

We have specified 10 designs here, from the combination of 2 distinct values of the power and 5 distinct values of the standard deviation. To view all 10 designs on the screen, click  to maximize the **Output Preview**. The designs within the **Output Preview** can be sorted in ascending or descending order, according to one of the column variables. For example, if you click once on the column titled **Sample Size**, the designs will be sorted (from top to bottom) in ascending order of the total sample size.

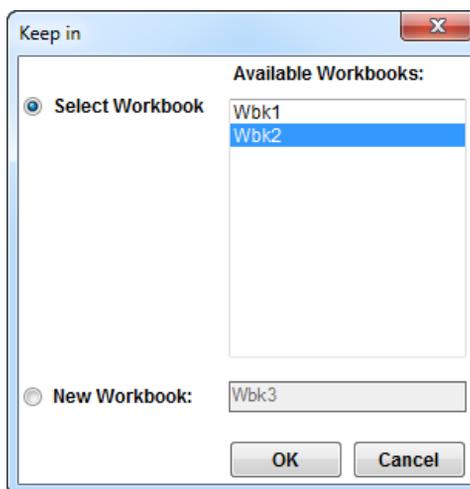
In addition, you may wish to filter and select designs that meet certain criteria. Click

10 Normal Superiority Two-Sample

 on the **Output Preview** toolbar, and in the filter criterion box, select only those designs for which the maximum sample size is less than or equal to 400, as follows:



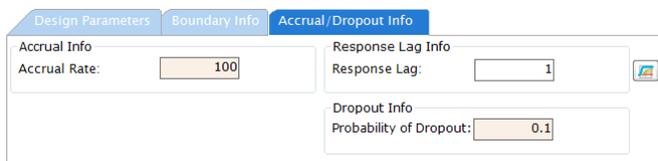
From the remaining designs, select Des8 in the **Output Preview**, and click . You will be asked to nominate the workbook in which this design should be saved. Select Wbk2 and click **OK**.



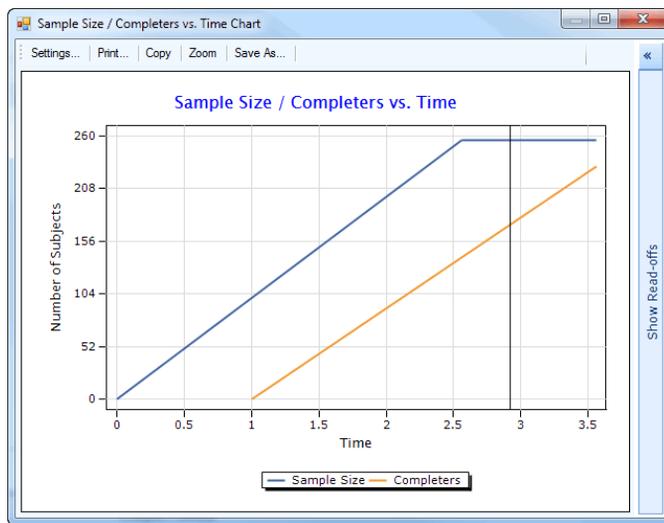
Accrual and dropout information

More realistic assumptions regarding the patient accrual process – namely, accrual rate, response lag, and probability of dropout – can be incorporated into the design stage. First, the accrual of patients may be estimated to occur at some known rate. Second, because the primary outcome measure is change in body weight from baseline to end of first year, the response lag is known to be 1 year. Finally, due to the long-term nature of the study, it is estimated that a small proportion of patients is likely to drop out over the course of the study.

With Des8 selected in the **Library**, click . Click **Include Options** in the top right hand corner of the Input, and then click **Accrual/Dropout Info**. A new tab should appear to the right of **Design Parameters** and **Boundary Info**. Click on this **Accrual/Dropout Info** tab, and enter the following information as shown below: The accrual rate is 100 patients per year, the response lag is 1 year, and the probability that a patients drops out before completing the study is 0.1.



A plot of the predicted accruals and completers over time can be generated by clicking .



Click **Compute** to generate the design. Select Des18 in the **Output Preview**, and click . Select Wbk2 and click **OK**. Double-click Des18 in the **Library**. The output details reveal that in order to ensure that data can be observed for 153 completers by the second look, one needs to have accrued 255 subjects. Close this Output window

10 Normal Superiority Two-Sample

before continuing.

Completers, Sample Size, Dropouts, Pipeline and Analysis Times: Look by Look

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Completers (s) | Dropouts (d) | Pipeline (n-s-d) | Analysis Time |
|--------|--------------------------|-----------------|----------------|--------------|------------------|---------------|
| 1 | 0.335 | 186 | 77 | 9 | 100 | 1.86 |
| 2 | 0.665 | 255 | 153 | 17 | 85 | 2.7 |
| 3 | 1 | 255 | 230 | 25 | 0 | 3.55 |

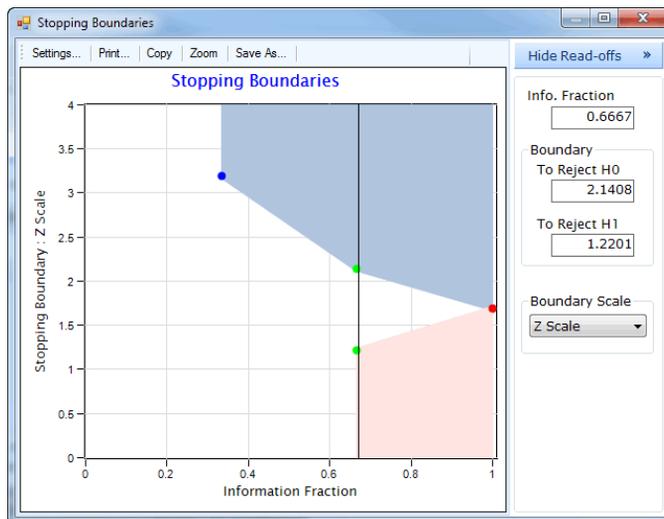
Select individual looks

With Des8 selected in Wbk2, click . In the look details table of the **Boundary Info** tab, notice that there are ticked checkboxes under the columns **Stop for Efficacy** and **Stop for Futility**. East gives you the flexibility to remove one of the stopping boundaries at certain looks. For example, untick the checkbox in the first look under the **Stop for Futility** column, and click **Recalc**.

| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Cum. β Spent | Futility Boundary |
|--------|----------------|-------------------------------------|-------------------------------------|---------------------|-------------------|--------------------|-------------------|
| 1 | 0.333 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | 0.0007 | 3.2001 | | NA |
| 2 | 0.667 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.0164 | 2.1408 | 0.1527 | 1.2201 |
| 3 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.0500 | 1.6948 | 0.2000 | 1.6948 |

Click  to view the new boundaries. Notice that the futility boundary does not

begin until the second look.



Simulation of the Orlistat trial

Suppose you now wish to simulate Des4 in Wbk1. Select Des4 in the **Library**, and click the **S** from the **Library** toolbar. Alternatively, right-click on Des4 and select **Simulate**. A new Simulation worksheet will appear. Click on the **Response Generation Info** tab, and input the following values: **Mean control** = 6; **Mean Treatment** = 6; **(Common) Std. Deviation** = 8. In other words, we are simulating from a population in which there is no true difference between the control and treatment means. This simulation will allow us to check the type-1 error rate when using Des4.

| Simulation Parameters | Response Generation Info | Simulation Control Info |
|---------------------------------------|---|-------------------------|
| Generate Data Using: Individual Means | <input checked="" type="checkbox"/> Common Standard Deviation | |
| Mean Control (μ_c): 6 | SD Control (σ_c): 8 | |
| Mean Treatment (μ_t): 6 | SD Treatment (σ_t): 8 | |

Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled Sim1.

10 Normal Superiority Two-Sample

With Sim1 selected in the **Output Preview**, click  , then double-click Sim1 in the **Library**. The simulation output details will be displayed in the upper pane. In the **Overall Simulation Result** table, notice that the percentage of times the upper efficacy stopping boundary was crossed is largely consistent with a type-1 error of 5%. The exact values of your simulations may differ, depending on your seed.

☯ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | | Efficacy | Count |
| | | Upper | | | |
| 1 | 111 | 2.961 | 18 | 18 | 0.180% |
| 2 | 222 | 2.094 | 180 | 180 | 1.800% |
| 3 | 333 | 1.71 | 306 | 9802 | 98.020% |
| Total | | | 504 | 10000 | |
| % | | | 5.040% | | |

Right-click Sim1 in the **Library** and click **Edit Simulation**. In the **Response Generation Info** tab, enter 9 for **Mean Treatment**. Leave all other values, and click **Simulate**. With Sim2 selected in the **Output Preview**, click  , then double-click Sim2 in the **Library**. Notice that the percentage of times the efficacy stopping boundary was crossed is largely consistent with 90% power for the original design. Feel free to experiment further with other simulation options before continuing.

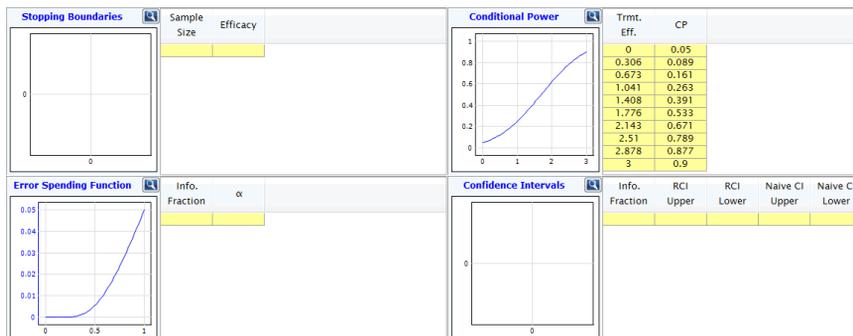
10.1.2 Interim monitoring of the Orlistat trial

Suppose we decided to adopt Des2. Select Des2 in the **Library**, and click **IM** on the **Library** toolbar. Alternatively, right-click on Des2 and select **Interim Monitoring**. The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | Repeated 95% CI for δ | | Repeated p-value |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|------------------------------|-------|------------------|
| | | | | | | | Upper | Lower | |
| | | | | | | | | | |
| 1 | | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | | | | | | | | | |

The bottom section contains four charts, each with a corresponding table to its right. These charts provide graphical and numerical descriptions of the progress of the

clinical trial and are useful tools for decision making by a data monitoring committee.



Making Entries in the Interim Monitoring Dashboard

Although the study has been designed assuming three equally spaced analyses, departures from this strategy are permissible using the spending function methodology of Lan and DeMets (1983) and its extension to boundaries for early stopping in favor of H_0 proposed by Pampallona, Tsiatis and Kim (2001). At each interim analysis time point, East will determine the amount of type-1 error probability and type-2 error probability that it is permitted to spend based on the chosen spending functions specified in the design. East will then re-compute the corresponding stopping boundaries. This strategy ensures that the overall type-1 error will not exceed the nominal significance level α . We shall also see how East proceeds so as to control the type-2 error probability.

Open the **Test Statistic Calculator** by clicking on the **Enter Interim Data** button. Assume that we take the first look after 110 patients (**Sample Size (Overall)**), with an **Estimate of δ** as 3, and **Standard Error of Estimate of δ** as 1.762. Click **OK** to

10 Normal Superiority Two-Sample

continue.

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Cumulative Sample Size:

Input for Normal end point

Estimate of δ :
 $\delta = (\mu_t - \mu_c)$

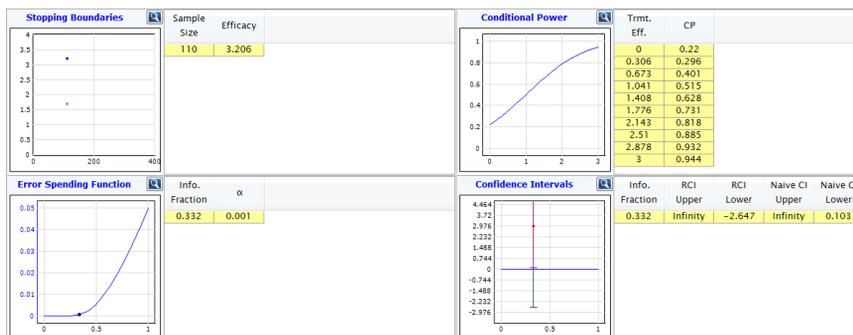
Standard Error of Estimate of δ :

Output

Test Statistic:

Recalc OK Cancel

East will update the charts and tables in the dashboard accordingly. For example the Stopping Boundaries Chart displays recomputed stopping boundaries and the path traced out by the test statistic. The Error Spending Function Chart displays the cumulative error spent at each interim look. The Conditional Power (CP) Chart shows the probability of crossing the upper stopping boundary, given the most recent information. Finally, the RCI (Repeated Confidence Interval) Chart displays repeated confidence intervals (Jennison & Turnbull, 2000).



Repeat the input procedure from above with the second look after 221 patients (Sample Size (Overall), Estimate of δ as 2, and Standard Error of Estimate of δ as 1. Click **Recalc** and **OK** to continue.

For the final look, make sure to tick the box **Set Current Look as Last**. Input the following estimates: 331 patients (**Sample Size (Overall)**), with an **Estimate of δ** as 3, and **Standard Error of Estimate of δ** as 1. Click **Recalc** and **OK** to continue.

The upper boundary has been crossed. The dashboard will be updated, and the **Final Inference** table shows the final outputs. For example, the adjusted p-value is 0.017, consistent with the rejection of the null.

| Final Inference | |
|--|-------|
| Final Outputs at Look # | 3 |
| Adj. p-value | 0.017 |
| Adj. Pt. Est. for δ | 2.505 |
| Adj. 90% CI for δ | |
| Upper Confidence Bound | 4.314 |
| Lower Confidence Bound | 0.582 |
| Post-Hoc Power | 0.9 |

10.1.3 Trial Design Using a t-Test (Single Look)

In Section 10.1.1 the sample size obtained to correctly power the trial relied on asymptotic approximation for the distribution of a Wald-type statistic. In the single look setting this statistic is

$$Z = \frac{\hat{\delta}}{\sqrt{\text{var}[\hat{\delta}]}} \quad (10.7)$$

with

$$\text{var}[\hat{\delta}] = \frac{\hat{\sigma}^2}{nr(1-r)}. \quad (10.8)$$

In a small single-look trial a more accurate representation of the distribution of Z is obtained by using Student's t-distribution with $(n-1)$ degrees of freedom.

Consider the Orlistat trial described in Section 10.1.1 where we would like to test the null hypothesis that treatment does not lead to weight loss, $H_0: \delta = 0$, against the alternative hypothesis that the treatment does result in a loss of weight, $H_1: \delta > 0$. We will now design this same trial in a different manner, using the t-distribution for the test statistic. Start East afresh. Click **Continuous: Two Samples** on the **Design** tab, and then click **Parallel Design: Difference of Means**. Enter the following design parameters so that the dialog box appears as shown. Remember to select a **1-Sided** for **Trial Type**, and enter an **Allocation Ratio** of 3. These values are the same as those

10 Normal Superiority Two-Sample

from Des1, except that under **Dist. of Test Stat.**, select **t**. Then click **Compute**.

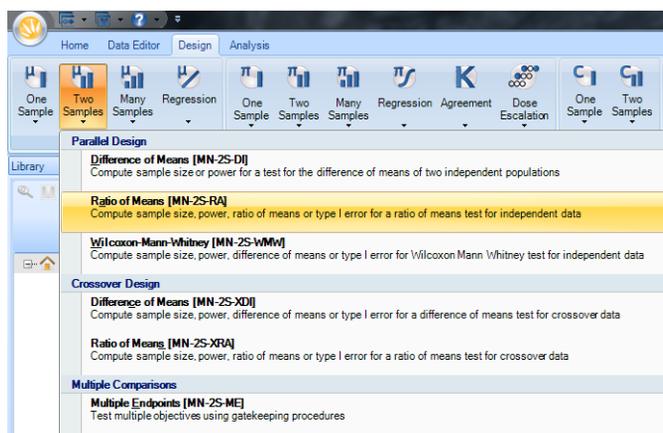
The screenshot shows the Minitab Design tool interface for a "Design: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means". The "Design Parameters" tab is active. The "Design Type" is set to "Superiority" and "Number of Looks" is 1. The "Test Type" is "1-Sided". The "Type I Error (α)" is 0.05, "Power" is 0.9, and "Sample Size (n)" is 327. The "Allocation Ratio" is 3. The "Input Method" is "Individual Means", with "Mean Control (μ_c)" at 6 and "Mean Treatment (μ_t)" at 9. The "Test Statistic" is "t", "Variance" is "Equal", and "Std. Deviation (σ)" is 8. A "Compute" button is visible at the bottom right.

We observe that the required sample size for this study is 327 patients. Contrast this to the 325 patients obtained using the normal distribution in Section 10.1.1.

10.2 Ratio of Means for Independent Data (Superiority)

Let σ_t and σ_c denote the standard deviations of the treatment and control group responses respectively. It is assumed that the coefficient of variation (CV), defined as the ratio of the standard deviation to the mean, is the same for both groups: $\frac{\sigma_t}{\mu_t} = \frac{\sigma_c}{\mu_c}$. Finally let $\rho = \frac{\mu_t}{\mu_c}$. For a Superiority trial, the null hypothesis $H_0 : \rho = \rho_0$ is tested against the two-sided alternative hypothesis $H_1 : \rho \neq \rho_0$ or a one-sided alternative hypothesis $H_1 : \rho < \rho_0$ or $H_1 : \rho > \rho_0$.

First, click **Continuous: Two Samples** on the **Design** tab, and then click **Parallel Design: Ratio of Means**.



Suppose that we wish to determine the sample size required for a one sided test to achieve a type-1 error of .05, and power of 90%, to detect a ratio of means of 1.25. We also need to specify the $CV = 0.25$. Enter the appropriate design parameters so that the input dialog box appears as below, and click **Compute**.

10 Normal Superiority Two-Sample

Design: Continuous Endpoint: Two-Sample Test - Parallel Design - Ratio of Means

Design Type: Superiority Include Options

Design Parameters

| | | | | | | | |
|------------------------------------|---------|--------------------------------|----------------|--|------|--------------|------|
| Test Type: | 1-Sided | Input Method: | Ratio of Means | Test Statistic: | Z | | |
| Type I Error (α): | 0.05 | Specify Alternative Hypothesis | | Ratio of Means ($\rho = \mu_1 / \mu_2$): | 1.25 | Coeff. Var.: | 0.25 |
| Power: | 0.9 | | | | | | |
| Sample Size (n): | 42 | | | | | | |
| Allocation Ratio: (n_1/n_2) | 1 | | | | | | |

Compute

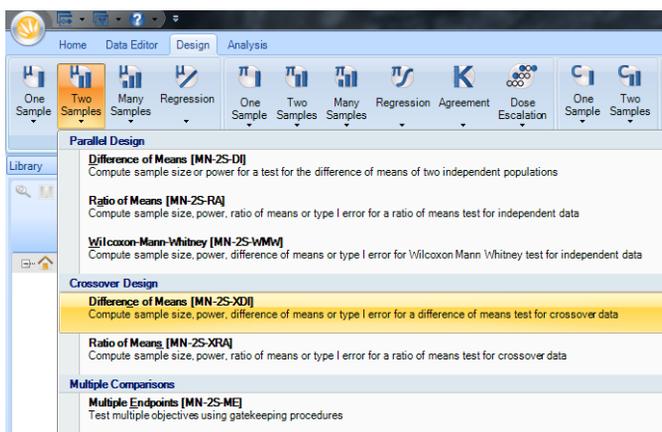
The computed sample size (42 subjects) is highlighted in yellow.

10.3 Difference of Means for Crossover Data (Superiority)

In a crossover trial, each experimental subject receives two or more different treatments. The order in which each subject receives the treatments depends on the particular design chosen for the trial. The simplest design is a 2×2 crossover trial, where each subject receives two treatments, say A and B. Half of the subjects receive A first and then, after a suitably chosen period of time, crossover to B. The other half receive B first and then crossover to A.

The null and alternative hypotheses are the same as for a two sample test for difference of means for independent data. However, a key advantage of the crossover design is that each subject serves as his/her own control. The test statistic also needs to account for not only treatment effects, but period and carryover effects.

We will demonstrate this design for a Superiority trial. First, click **Continuous: Two Samples** on the **Design** tab, and then click **Crossover Design: Difference of Means**.



Suppose that we wish to determine the sample size required to achieve a type-1 error of .05, and power of 90%, to detect a difference of means of 75 with standard deviation of the difference of 150. Enter the appropriate design parameters so that the input

10 Normal Superiority Two-Sample

dialog box appears as below, and click **Compute**.

Design: Continuous Endpoint: Two-Sample Test - Crossover Design - Difference of Means

Design Type: Superiority

Design Parameters

Test Type: 2-Sided

Type I Error (α): 0.05

Power: 0.9

Sample Size (n): 45

Allocation Ratio: 1
(n_{α} / n_{α})

Input Method: Difference of Means

Specify Alternative Hypothesis
Difference of Means (δ_1): 75

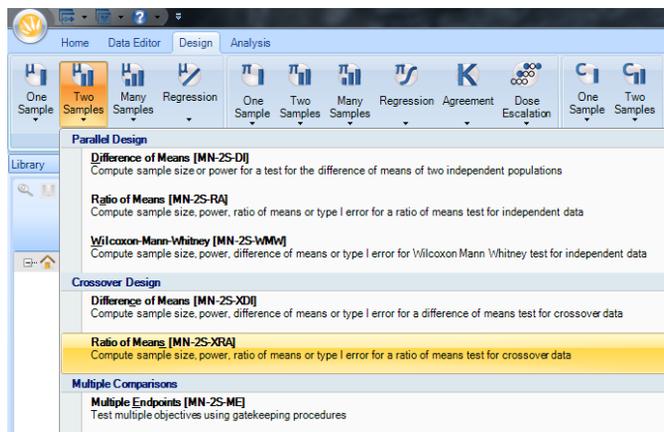
Std. Dev. of Diff.: 150

Compute

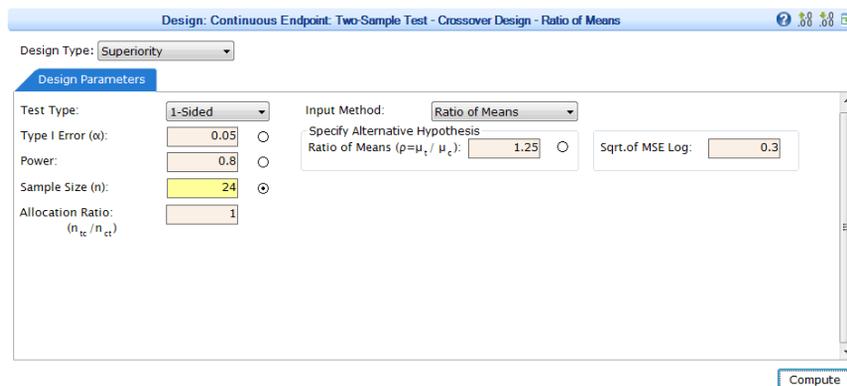
The computed sample size (45 subjects) is highlighted in yellow.

10.4 Ratio of Means for Crossover Data (Superiority)

We will demonstrate this design for a Superiority trial. The null hypothesis $H_0 : \rho = \rho_0$ is tested against the two-sided alternative hypothesis $H_1 : \rho \neq \rho_0$ or a one-sided alternative hypothesis $H_1 : \rho < \rho_0$ or $H_1 : \rho > \rho_0$. First, click **Continuous: Two Samples** on the **Design** tab, and then click **Crossover Design: Ratio of Means**.



Suppose that we wish to determine the sample size required for a one sided test to achieve a type-1 error of .05, and power of 80%, to detect a ratio of means of 1.25 with square root of MSE of 0.3. Enter the appropriate design parameters so that the input dialog box appears as below, and click **Compute**.



The computed sample size (24 subjects) is highlighted in yellow.

10 Normal Superiority Two-Sample

10.5 Assurance (Probability of Success)

Assurance, or probability of success, is a Bayesian version of power, which corresponds to the (unconditional) probability that the trial will yield a statistically significant result. Specifically, it is the prior expectation of the power, averaged over a prior distribution for the unknown treatment effect (see O’Hagan et al., 2005). For a given design, East allows you to specify a prior distribution, for which the assurance or probability of success will be computed.

Select Des2 in the **Library**, and click  on the **Library** toolbar. Alternatively, recompute this design with the following inputs: A 3-look design with Lan-Demets(OF) efficacy only boundary, Superiority Trial, 1-sided, 0.05 type-1 error, 90% power, allocation ratio = 3, mean control = 6, mean treatment = 9, and standard deviation = 8.

Select the **Assurance** checkbox in the Input window.

Suppose that we wish to specify a Normal prior distribution for the treatment effect δ , with a mean of 3, and standard deviation of 2. Thus, rather than assuming $\delta = 3$ with certainty, we use this prior distribution to reflect the uncertainty about the true treatment effect.

In the **Distribution** list, click **Normal**, and in the **Input Method** list, click **E(δ) and SD(δ)**.

Type 3 in the **E(δ)** box, and type 2 in the **SD(δ)** box, and then click **Compute**.

The computed probability of success (0.72) is shown below. Note that for this prior, assurance is less than the specified power (0.9); incorporating the uncertainty about δ has yielded a less optimistic estimate of power.

Assurance (Probability of Success): 0.72

Prior Distribution for: δ Distribution: Normal

Input Method: E(δ) and SD(δ)

User Specified

E(δ): 3

SD(δ): 2

In the Output Preview, right-click the row corresponding to this design, and rename the design ID as **Bayes1**, and save it to the **Library**.

Return to the input window. Type 0.001 in the **SD(δ)** box, and click **Compute**. Such a prior approximates the non-Bayesian power calculation, where one specifies a fixed treatment effect.

As shown below, such a prior yields a probability of success that is similar to the specified power.

Assurance (Probability of Success): 0.9

Prior Distribution for: δ Distribution: Normal

Input Method: E(δ) and SD(δ)

User Specified

E(δ): 3

SD(δ): 0.001

East also allows you to specify an arbitrary prior distribution through a CSV file. In the **Distribution** list, click **User Specified**, and then click **Browse...** to select the CSV file where you have constructed a prior.

10 Normal Superiority Two-Sample

Assurance (Probability of Success) Computed
 Prior Distribution for: δ Distribution: User Specified
 Browse...

The CSV file should contain two columns, where the first column lists the grid points for the parameter of interest (in this case, δ), and the second column lists the prior probability assigned to each grid point. For example, we consider a 5-point prior with probability = 0.2 at each point. The prior probabilities can be entered as weights that do not sum to one, in which case East will re-normalize for you.

| | A | B |
|---|---|-----|
| 1 | 1 | 0.2 |
| 2 | 2 | 0.2 |
| 3 | 3 | 0.2 |
| 4 | 4 | 0.2 |
| 5 | 5 | 0.2 |

Once the CSV filename and path has been specified, click **Compute** to calculate the assurance, which will be displayed in the box below:

Assurance (Probability of Success): 0.751
 Distribution: User Specified

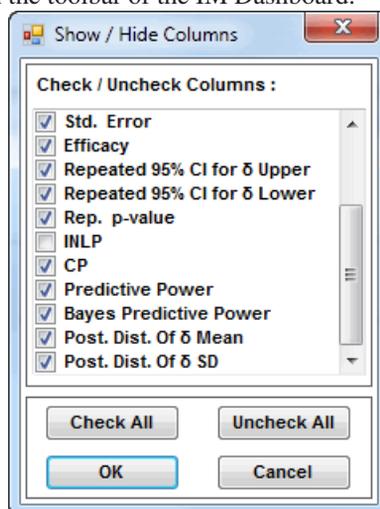
As stated in O’Hagan et al. (2005, p.190): “Assurance is a key input to decision-making during drug development and provides a reality check on other methods of trial design.” Indeed, it is not uncommon for assurance to be much lower than the specified power. The interested reader is encouraged to refer to O’Hagan et al. for further applications and discussions on this important concept.

10.6 Predictive Power and Bayesian Predictive Power

Similar Bayesian ideas can be applied to conditional power for interim monitoring. Rather than calculating conditional power for a single assumed value of the treatment effect, δ , such as at $\hat{\delta}$, we may account for the uncertainty about δ by taking a weighted average of conditional powers, weighted by the posterior distribution for δ . For normal

endpoints, East assumes a posterior distribution for δ that results from a *diffuse* prior distribution, which produces an average power called the **predictive power** (Lan, Hu, & Proschan, 2009). In addition, if the user specified a normal prior distribution at the design stage to calculate assurance, then East will also calculate the average power, called **Bayesian predictive power**, for the corresponding posterior. We will demonstrate these calculations for the design renamed as Bayes1 earlier.

In the **Library**, right-click Bayes1 and click **Interim Monitoring**, then click (Show/Hide Columns) in the toolbar of the IM Dashboard.



In the Show/Hide Columns window, make sure to show the columns for: CP (Conditional Power), Predictive Power, Bayes Predictive Power, Posterior Distribution of δ Mean, and Posterior Distribution of δ SD, and click **OK**. The following columns will be displayed in the main grid of the IM Dashboard.

| CP | Predictive Power | Bayes Predictive Power | Post. Dist. Of δ | |
|----|------------------|------------------------|-------------------------|----|
| | | | Mean | SD |
| | | | | |
| | | | | |

Assume that we observed interim data after 110 patients, with an estimate of $\delta = 1$, and a standard error of the estimate = 0.7. Enter these values in the **Test Statistic Calculator** by clicking **Enter Interim Data**, and click **OK**.

10 Normal Superiority Two-Sample

The IM Dashboard will be updated. In particular, notice the differing values for CP and the Bayesian measures of power.

| CP | Predictive Power | Bayes Predictive Power | Post. Dist. Of δ | |
|------|------------------|------------------------|-------------------------|-------|
| | | | Mean | SD |
| 0.84 | 0.719 | 0.487 | 1.218 | 0.661 |
| | | | | |
| | | | | |

11 Nonparametric Superiority Two Sample

The Wilcoxon-Mann-Whitney nonparametric test is a commonly used test for the comparison of two distributions when the observations cannot be assumed to come from normal distributions. It is used when the distributions differ only in a location parameter and is especially useful when the distributions are not symmetric. For Wilcoxon-Mann-Whitney test, East supports single look superiority designs only.

11.1 Wilcoxon-Mann-Whitney Test

Let X_1, \dots, X_{n_t} be the n_t observations from the treatment (T) with distribution function F_t and Y_1, \dots, Y_{n_c} be the n_c observations from the control (C) with distribution function F_c . F_t and F_c are assumed to be continuous with corresponding densities f_t and f_c , respectively. The primary objective in Wilcoxon-Mann-Whitney test is to investigate whether there is a shift of location, which indicates the presence of the treatment effect. Let θ represents the treatment effect. Then we test the null hypothesis $H_0: \theta = 0$ against the two-sided alternative $H_1: \theta \neq 0$ or a one-sided alternative hypothesis $H_1: \theta < 0$ or $H_1: \theta > 0$. Let U denote the number of pairs (X_i, Y_j) such that $X_i < Y_j$, so $U = \sum_{i=1}^{n_t} \sum_{j=1}^{n_c} I(X_i, Y_j)$ where $I(a, b) = 1$ if $a < b$ and $I(a, b) = 0$ if $a \geq b$. Then $U/n_c n_t$ is a consistent estimator of

$$p = P(X < Y) = \int_{-\infty}^{\infty} F_t(y) f_c(y) dy = \int_0^1 F_t[F_c^{-1}(u)] du. \quad (11.1)$$

The power is approximated using the asymptotic normality of U and depends on the value of p , and thus depends on F_c and F_t . In order to find the power for a given sample size or to find the sample size for a given power, we must specify p . However, this is often a difficult task. If we are willing to specify F_c and F_t , then p can be computed. East computes p assuming that F_c and F_t are normal distributions with means μ_c and μ_t and a common standard deviation σ , by specifying the values of the difference in the means and the standard deviation. With this assumption, equation (11.1) results in

$$p = \Phi\left(\frac{\mu_t - \mu_c}{\sqrt{2}\sigma}\right) \quad (11.2)$$

Using the results of Noether (1987), with $n_t = rN$, the total sample size for an α level two-sided test to have power $1 - \beta$ for a specified value of p is approximated by

$$N = \frac{(z_{\alpha/2} + z_{\beta})^2}{12r(1-r)(p - .5)^2}.$$

11 Nonparametric Superiority Two Sample

11.2 Example: Designing a single look superiority study

Based on a pilot study of an anti-seizure medication, we want to design a 12-month placebo-controlled study of a treatment for epilepsy in children. The primary efficacy variable is the percent change from baseline in the number of seizures in a 28-day period. The mean percent decrease was 2 for the control and 8 for the new treatment, with an estimated standard deviation of 25. We plan to design the study to test the null hypothesis $H_0:\theta = 0$ against $H_1:\theta \neq 0$. We want to design a study that would have 90% power at $\mu_c = 2$ and $\mu_t = 8$ under H_1 and maintains type I error at 5%.

11.2.1 Designing the study

Click **Continuous: Two Samples** on the **Design** tab and then click **Parallel Design: Wilcoxon-Mann-Whitney**.

Design: Continuous Response: Wilcoxon Mann Whitney Test for Independent Data

Include Options

Design Parameters

Side: Input Method:

Type I Error (α):

Power ($1 - \beta$):

Sample Size (n):

Allocation Ratio (n_1/n_2):

Specify Mean Responses

Mean Control (μ_c):

Mean Treatment (μ_t):

Std. Deviation:

Compute

This will launch a new window. The upper pane of this window displays several fields with default values. Select **2-Sided** for **Test Type** and enter **0.05** for **Type I Error**. Select **Individual Means** for **Input Method** and then specify **Mean Control** (μ_c) as 2 and **Mean Treatment** (μ_t) as 8. Specify **Std. Deviation** as 25. Click **Compute**. The upper pane now should appear as below:

Design: Continuous Response: Wilcoxon Mann Whitney Test for Independent Data

Design Parameters

Side: Input Method:

Type I Error (α):

Power (1-β):

Sample Size (n):

Allocation Ratio (n₁ / n₂):

Specify Mean Responses

Mean Control (μ_c):

Mean Treatment (μ_t):

The required sample size for this design is shown as a row in the **Output Preview**, located in the lower pane of this window. The computed total sample size (772 subjects) is highlighted in yellow.

| Output Preview | | | | | | | | | | | |
|---|------|-----|------------|-----------|------------------|-----------|-----------|-------------|---------|--------------|--|
| Design | α | 1-β | Maximum SS | Side Type | Allocation Ratio | Ctrl Mean | Trtm Mean | Probability | Std Dev | Actual Power | |
|  Design1 | 0.05 | 0.9 | 772 | 2-Sided | 1 | 2 | 8 | 0.567 | 25 | 0.9 | |

This design has default name Des 1 and results in a total sample size of 772 subjects in order to achieve 90% power. The probability displayed in the row is 0.567, which indicates the approximate probability $P[X < Y]$ assuming $X \sim N(8, 25^2)$ and $Y \sim N(2, 25^2)$. This is in accordance with the equation 11.2.

Select this design by clicking anywhere along the row in the **Output Preview** and click  in the **Output Preview** toolbar. Some of the design details will be

11 Nonparametric Superiority Two Sample

displayed in the upper pane, labeled as **Output Summary**.

| Parameters | Design 1 |
|-----------------------------------|----------|
| Test Parameters | |
| Type-1 Error (α) | 0.05 |
| Power ($1-\beta$) | 0.9 |
| Test Side | 2-Sided |
| Allocation Ratio (nt/nc) | 1 |
| Sample Size | |
| Total Sample Size | 772 |
| Model Parameters | |
| Actual Power | 0.9 |
| Control Mean (μ_c) | 2 |
| Treatment Mean (μ_t) | 8 |
| Probability (Control < Treatment) | 0.567 |
| Std Dev | 25 |

In the **Output Preview** toolbar, click  to save this design to Wbk1 in the **Library**. Double-click Des 1 in the **Library** to see the details of the design.

Design: Wilcoxon MannWhitney Test:Comparing two Independent Location Parameters

Input Parameters

| | |
|--|----------------|
| Design ID: | Design1 |
| Side Type: | 2-Sided |
| Type-1 Error (α): | 0.05 |
| Power ($1-\beta$): | 0.9 |
| Sample Size (n): | To be Computed |
| Sample Size Allocation Ratio (nt/nc): | 1 |
| Mean Response under Control (μ_c): | 2 |
| Mean Response under Treatment (μ_t): | 8 |
| Standard Deviation (σ): | 25 |
| Pr(Control < Treatment): | 0.567 |

Output

| | |
|---|-------|
| Sample Size (n): | 772 |
| Mean Difference ($\delta_1 = \mu_t - \mu_c$): | 6 |
| Sample Size under Control (nc): | 386 |
| Sample Size under Treatment (nt): | 386 |
| Actual Power for Computed Sample Size: | 0.9 |
| Lower Critical Point: | -1.96 |
| Upper Critical Point: | 1.96 |

Summary

The study will require a total of 772 subjects to detect a difference in response distributions between the treatment and control samples, with desired power 90.02%. This test is based on a 2-Sided non-parametric Wilcoxon Mann Whitney test for continuous response data with significance level 0.05. The treatment group mean is 8 and the control group mean is 2. The test assumes the common standard deviation of 25 for treatment and control groups. The probability that an observation in the control group is less than an observation in the treatment group under the alternative hypothesis is 0.567. The allocation ratio is 1 and therefore, of the 772 subjects, 386 are allocated to the control arm and 386 are allocated to the treatment arm.

According to this summary, the study needs a total of 772 subjects. Of these 772 subjects, 386 will be allocated to the treatment group and remaining 386 will be allocated to the control group.

Since the sample size is inversely proportional to $(p - .5)^2$, it is sensitive to mis-specification of p (see equation (11.1)). The results of the pilot study included several subjects who worsened over the baseline and thus the difference in the means might not be an appropriate approach to determining p . To obtain a more appropriate value of p , we have several alternative approaches. We can further examine the results

of the pilot study after exclusion of some of the extreme values, which will decrease the standard deviation and provide a difference in the means, which may be a more reasonable measure of the difference between the distributions. The difference in the medians may be a more reasonable measure of the difference between the distributions, especially when used with a decreased standard deviation.

The median percent decrease was 10 for the control and 18 for the new treatment, with an estimated standard deviation of 25. Create a new design by selecting Des 1 in the **Library**, and clicking  on the **Library** toolbar. In the Input, change the **Mean Control** (μ_c) and **Mean Treatment** (μ_t) to 10 and 18, respectively.

Design: Continuous Response: Wilcoxon Mann Whitney Test for Independent Data

Include Options

Design Parameters

Side: 2-Sided Input Method: Individual Means

Type I Error (α): 0.05

Power (1 - β): 0.9

Sample Size (n): Computed

Allocation Ratio (n_1/n_2): 1

Specify Mean Responses

Mean Control (μ_c): 10

Mean Treatment (μ_t): 18

Std.Deviation

25

Click **Compute** to generate output for Des 2. To compare Des 1 and Des 2, select both rows in **Output Preview** using the Ctrl key, and click  icon in the **Output Preview** toolbar. Both designs will be displayed in the **Output Summary** pane.

| Parameters | Design 1 | Design 2 |
|-----------------------------------|----------|----------|
| Test Parameters | | |
| Type-I Error (α) | 0.05 | 0.05 |
| Power (1 - β) | 0.9 | 0.9 |
| Test Side | 2-Sided | 2-Sided |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Sample Size | | |
| Total Sample Size | 772 | 438 |
| Model Parameters | | |
| Actual Power | 0.9 | 0.901 |
| Control Mean (μ_c) | 2 | 10 |
| Treatment Mean (μ_t) | 8 | 18 |
| Probability (Control < Treatment) | 0.567 | 0.59 |
| Std Dev | 25 | 25 |

The sample size required for Des 2 is only 438 subjects as compared to 772 subjects in Des 1. Now we consider decreasing the standard deviation to 20 to lessen the impact of the extreme values. Select Des 2 in the **Output Preview**, and click  icon in the

11 Nonparametric Superiority Two Sample

toolbar. In the Input, change the **Std. Deviation** to 20. Click **Compute** to generate output for this design. Select all the rows in **Output Preview** and click  in the **Output Preview** toolbar to see them in the **Output Summary** pane. This design results in a total sample size of 283 subjects in order to attain 90% power.

| Compare Designs | | | |
|----------------------------|---------|---------|---------|
| Parameters | Design1 | Design2 | Design3 |
| Test Parameters | | | |
| Type-1 Error (α) | 0.05 | 0.05 | 0.05 |
| Power ($1-\beta$) | 0.9 | 0.9 | 0.9 |
| Test Side | 2-Sided | 2-Sided | 2-Sided |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Sample Size | | | |
| Total Sample Size | 772 | 438 | 283 |
| Model Parameters | | | |
| Actual Power | 0.9 | 0.901 | 0.901 |
| Control Mean (μ_c) | 2 | 10 | 10 |
| Treatment Mean (μ_t) | 8 | 18 | 18 |
| Probability (Contr | 0.567 | 0.59 | 0.611 |
| Std Dev | 25 | 25 | 20 |

12 Normal Non-inferiority Two-Sample

In a noninferiority trial, the goal is to establish that an experimental treatment is *no worse than* the standard treatment, rather than attempting to establish that it is superior. A therapy that is demonstrated to be non-inferior to the current standard therapy for a particular indication might be an acceptable alternative if, for instance, it is easier to administer, cheaper, or less toxic. Non-inferiority trials are designed by specifying a non-inferiority margin. The amount by which the mean response on the experimental arm is worse than the mean response on the control arm must fall within this margin in order for the claim of non-inferiority to be sustained. In this chapter, we show how East supports the design and interim monitoring of such experiments, with a normal endpoint.

12.1 Difference of Means

12.1.1 Trial design

12.1.2 Three-Look Design

12.1.3 Simulation

12.1.4 Interim Monitoring

12.1.5 Trial Design Using a t-Test (Single Look)

12.1.1 Trial design

Consider the design of an antihypertension study comparing an ACE inhibitor to a new AII inhibitor. Let μ_c be the mean value of a decrease in systolic blood pressure level (in mmHg) for patients in the ACE inhibitor (control) group and μ_t be the mean value of a decrease in blood pressure level for patients in the AII inhibitor (treatment) group. Let $\delta = \mu_t - \mu_c$ be the treatment difference. We want to demonstrate that the AII inhibitor is non-inferior to the ACE inhibitor. For this example, we will consider a non-inferiority margin equal to one-third of the mean response in control group. From historical data, $\mu_c = 9$ mmHg and therefore the non-inferiority margin is 3 mmHg. Accordingly we will design the study to test the null hypothesis of inferiority $H_0 : \delta \geq -3$, against the one sided non-inferiority alternative $H_1 : \delta < -3$. The test is to be conducted at a significance level (α) of 0.025 and is required to have 90% power at $\delta = 0$. We assume that σ^2 , the variance of the patient response, is the same for both groups and is equal to 100.

Start East afresh. Click **Continuous: Two Samples** on the **Design** tab and then click **Parallel Design: Difference of Means**.

Single-look design

In the input window, select **Noninferiority** for **Design Type**. The effect size can be specified in one of three ways by selecting different options for **Input Method**: (1) individual means and common standard deviation, (2) difference of means and common standard deviation, or (3) standardized difference of means. We will use the **Individual Means** method. Select **Individual Means** for **Input Method**, specify the **Mean Control** (μ_c) as 9 and **Noninferiority margin** (δ_0) as -3 and specify the

12 Normal Non-inferiority Two-Sample

Std. Deviation (σ) as 10. Specify 0 for **Difference in Means** (δ_1). The upper pane should appear as below:

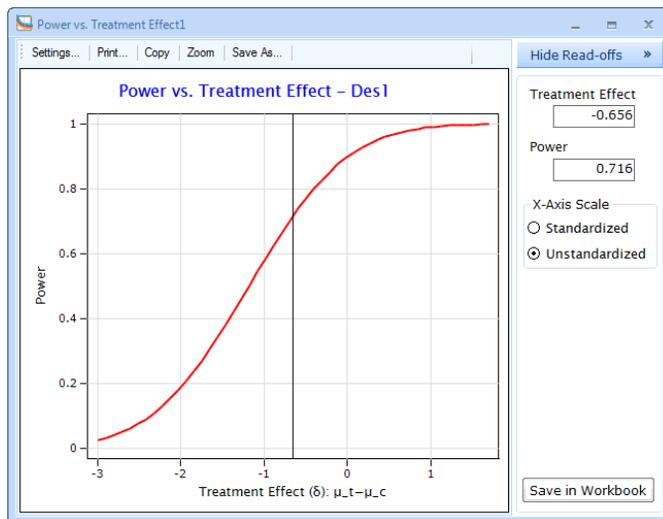
Click **Compute**. This will calculate the sample size for this design, and the output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed sample size (467 subjects) is highlighted in yellow.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | Input Method | δ_1 | δ_0 | μ_c | Mean Treatment (Alt.) | Mean Treatment (Null) | σ |
|------|----------------|--------------|-----------|--------------------|-------|-------|-------------|------------------|------------|------------|---------|-----------------------|-----------------------|----------|
| Des1 | Noninferiority | 1 | 1-Sided | 0.025 | 0.9 | 1 | 467 | Individual Means | 0 | -3 | 9 | 9 | 6 | 10 |

This design has default name Des 1. Select this design by clicking anywhere along the row in the **Output Preview** and click . In the **Output Preview** toolbar, click to save this design to Wbk1 in the **Library**. If you hover the cursor over Des 1 in the **Library**, a tooltip will appear that summarizes the input parameters of the design.

With Des 1 selected in the **Library**, click on the **Library** toolbar, and then click **Power vs Treatment Effect** (δ). The resulting power curve for this design will

appear.



You can save this chart to the Library by clicking **Save in Workbook**. In addition, you can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...** For now, you may close the chart before continuing.

12.1.2 Three-Look Design

Create a new design by selecting Des 1 in the **Library**, and clicking  on the **Library** toolbar. In the Input, change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. A new tab for **Boundary Info** should appear. Click this tab to reveal the stopping boundary parameters. By default, the **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The left side contains details for the **Efficacy** boundary, and the right side contains details for the **Futility** boundary. By default, there is an efficacy boundary (to reject H_0) selected, but no futility boundary (to reject H_1). The **Boundary Family** specified is of the **Spending Functions** type. The default **Spending function** is the **Lan-DeMets** (Lan & DeMets, 1983), with **Parameter OF** (O'Brien-Fleming), which generates boundaries that are very similar, though not identical, to the classical stopping

12 Normal Non-inferiority Two-Sample

boundaries of O'Brien and Fleming (1979).

Design Type: Noninferiority Number of Looks: 3

Design Parameters Boundary Info

Efficacy
 Boundary Family: Spending Functions
 Spending Function: Lan-DeMets
 Parameter: OF
 Type I Error (α): 0.025

Futility
 Boundary Family: None

Spacing of Looks Equal Unequal Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.333 | 0.000 | 3.710 |
| 2 | 0.667 | 0.006 | 2.511 |
| 3 | 1.000 | 0.025 | 1.993 |

Click **Compute** to generate output for Des 2. Save this design in the current workbook by selecting the corresponding row in **Output Preview** and clicking . To compare Des 1 and Des 2, select both rows in the **Output Preview** using the Ctrl key and click . Both designs will be displayed in the **Output Summary**.

| | Des 1 | Des 2 |
|--|------------------|------------------|
| Mnemonic | MN-2S-DI | MN-2S-DI |
| Test Parameters | | |
| Design Type | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Input Method | Individual Means | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 0 | 0 |
| Noninf. Margin ($\delta_0 = \mu_{t0} - \mu_c$) | -3 | -3 |
| Mean Control (μ_c) | 9 | 9 |
| Mean Treatment (μ_t) | 9 | 9 |
| Mean Treatment (μ_{t0}) | 6 | 6 |
| Std. Deviation (σ) | 10 | 10 |
| Test Statistic | Z | Z |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Boundary Parameters | | |
| Efficacy Boundary | | LD (OF) |
| Spacing of Looks | | Equal |
| Sample Size | | |
| Maximum | 467 | 473 |
| Expected Under H0 | | 472.032 |
| Expected Under H1 | | 379.185 |

The maximum sample size with Des 2 is 473, which is only a slight increase over the fixed sample size in Des 1. However, the expected sample size with Des 2 is 379 patients under H_1 , a saving of almost 100 patients. In order to see the stopping probabilities, double-click Des 2 in the **Library**.

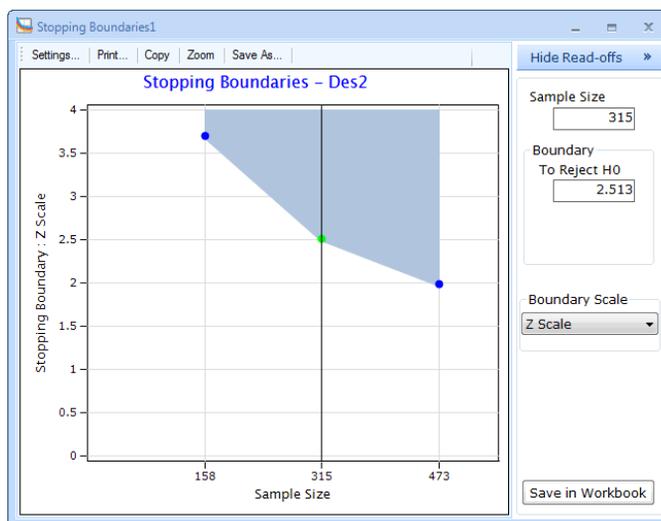
☺ **Stopping Boundaries: Look by Look**

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries Efficacy Z | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|--------------------------|---|----------|
| | | | | | Under H0 | Under H1 |
| | | | | | Efficacy | Efficacy |
| 1 | 0.334 | 158 | 1.053E-4 | 3.706 | 1.053E-4 | 0.034 |
| 2 | 0.666 | 315 | 0.006 | 2.513 | 0.006 | 0.525 |
| 3 | 1 | 473 | 0.025 | 1.993 | 0.019 | 0.341 |

The clear advantage of this sequential design resides in the high probability of stopping by the second look, if the alternative is true, with a sample size of 315 patients, which is well below the requirements for a fixed sample study (467 patients). Close the Output window before continuing.

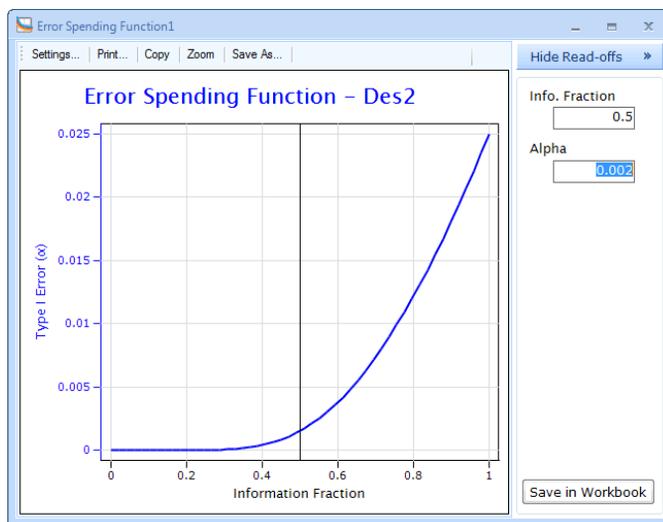
Examining stopping boundaries and spending functions

You can plot the boundary values of Des 2 by clicking  on the **Library** toolbar, and then clicking **Stopping Boundaries**. The following chart will appear:



12 Normal Non-inferiority Two-Sample

You can choose a different **Boundary Scale** from the corresponding drop down box. The available boundary scales include: Z scale, Score Scale, δ Scale, δ/σ Scale and p -value scale. To plot the error spending function for Des 2, select Des 2 in the **Library**, click the  in the toolbar, and then click **Error Spending**. The following chart will appear:



The above spending function is according to Lan and DeMets (1983) with O'Brien-Fleming flavor, and for one-sided tests has the following functional form:

$$\alpha(t) = 2 - 2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{t}}\right)$$

Observe that very little of the total type-1 error is spent early on, but more is spent rapidly as the information fraction increases, and reaches 0.025 at an information fraction of 1.

Feel free to explore other plots by clicking the  icon in the **Library** toolbar. Close all charts before continuing. To obtain the tables used to generate these plots, click the  icon.

Select Des 2 in the **Library**, and click  on the **Library** toolbar. In the **Boundary Info** tab, change the **Boundary Family** from **Spending Functions** to

Wang-Tsiatis. The Wang-Tsiatis (1989) power boundaries are of the form

$$c(t_j) = C(\Delta, \alpha, K)t_j^\Delta$$

for $j = 1, 2, \dots, K$, where Δ is a shape parameter that characterizes the boundary shape and $C(\Delta, \alpha, K)$ is a positive constant. The choice $\Delta = 0$ will yield the classic O'Brien-Fleming stopping boundary, whereas the $\Delta = 0.5$ will yield the classic Pocock stopping boundary. Other choices of parameters in the range -0.5 to 0.5 are also permitted. Accept the default parameter 0 and click **Compute** to obtain the sample size.

A new row will be added to the **Output Preview** with design name as Des 3.

Select all three rows in **Output Preview** using the Ctrl key and click  . All three designs will be displayed in the **Output Summary**.

| | Des 1 | Des 2 | Des 3 |
|---|------------------|------------------|------------------|
| Mnemonic | MN-2S-DI | MN-2S-DI | MN-2S-DI |
| Test Parameters | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 | 3 |
| Test Type | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 | 0.025 |
| Power | 0.9 | 0.9 | 0.9 |
| Model Parameters | | | |
| Input Method | Individual Means | Individual Means | Individual Means |
| Diff. in Means ($\delta = \mu_T - \mu_C$) | 0 | 0 | 0 |
| Noninf. Margin ($\delta_0 = \mu_T - \mu_C$) | -3 | -3 | -3 |
| Mean Control (μ_C) | 9 | 9 | 9 |
| Mean Treatment (μ_T) | 9 | 9 | 9 |
| Mean Treatment (μ_0) | 6 | 6 | 6 |
| Std. Deviation (σ) | 10 | 10 | 10 |
| Test Statistic | Z | Z | Z |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Boundary Parameters | | | |
| Efficacy Boundary | | LD (OF) | WT (0) |
| Spacing of Looks | | Equal | Equal |
| Sample Size | | | |
| Maximum | 467 | 473 | 475 |
| Expected Under H0 | | 472.032 | 473.825 |
| Expected Under H1 | | 379.185 | 373.38 |

12 Normal Non-inferiority Two-Sample

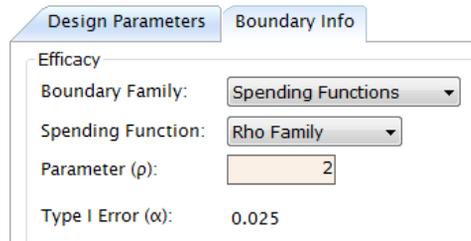
Note that the total sample size and the expected sample size under H_1 for Des 3 are close to those for Des 2. This is expected because the Wang-Tsiatis power family with shape parameter 0 yields the classic O'Brien-Fleming stopping boundaries. Save this design in the current workbook by selecting the corresponding row in **Output Preview** and clicking  on the **Output Preview** toolbar.

Select Des 2 in the **Library**, and click the  on the **Library** toolbar. In the **Boundary Info** tab, change the **Spending Function** from **Lan-DeMets** to **Rho Family**. The Rho spending function was first published by Kim and DeMets (1987) and was generalized by Jennison and Turnbull (2000). It has following functional form:

$$\alpha(t) = \alpha t^\rho \quad \rho > 0$$

When $\rho = 1$, the corresponding stopping boundaries resemble the Pocock stopping boundaries. When $\rho = 3$, the boundaries resemble the O'Brien-Fleming boundaries. Larger value of ρ yield increasingly conservative boundaries.

Specify parameter (ρ) as 2, and click **Compute**



| Design Parameters | | Boundary Info | |
|----------------------------|--------------------|---------------|--|
| Efficacy | | | |
| Boundary Family: | Spending Functions | | |
| Spending Function: | Rho Family | | |
| Parameter (ρ): | 2 | | |
| Type I Error (α): | 0.025 | | |

A new row will be added to the **Output Preview** with design name as Des 4. Select all four rows in **Output Preview** using the Ctrl key and click . All the designs will

be displayed in the **Output Summary**.

| | Des1 | Des2 | Des3 | Des4 |
|--|------------------|------------------|------------------|------------------|
| Mnemonic | MN-2S-DI | MN-2S-DI | MN-2S-DI | MN-2S-DI |
| Test Parameters | | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 | 3 | 3 |
| Test Type | 1-Sided | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 | 0.025 | 0.025 |
| Power | 0.9 | 0.9 | 0.9 | 0.9 |
| Model Parameters | | | | |
| Input Method | Individual Means | Individual Means | Individual Means | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 0 | 0 | 0 | 0 |
| Noninf. Margin ($\delta_0 = \mu_{t0} - \mu_c$) | -3 | -3 | -3 | -3 |
| Mean Control (μ_c) | 9 | 9 | 9 | 9 |
| Mean Treatment (μ_t) | 9 | 9 | 9 | 9 |
| Mean Treatment (μ_{t0}) | 6 | 6 | 6 | 6 |
| Std. Deviation (σ) | 10 | 10 | 10 | 10 |
| Test Statistic | Z | Z | Z | Z |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 | 1 |
| Boundary Parameters | | | | |
| Efficacy Boundary | | LD (OF) | WT (0) | Rho (2) |
| Spacing of Looks | | Equal | Equal | Equal |
| Sample Size | | | | |
| Maximum | 467 | 473 | 475 | 487 |
| Expected Under H0 | | 472.032 | 473.825 | 484.745 |
| Expected Under H1 | | 379.185 | 373.38 | 350.617 |

Observe that Des 4 requires a total sample size of 14 more subjects than Des 2. The expected sample size under H_1 for Des 4 is only 351 patients, compared to 379 patients for Des 2 and 467 patients for Des 1. Save Des 4 to the Library by selecting the corresponding row in the **Output Preview** and clicking  .

12.1.3 Simulation

Select Des 4 in the **Library**, and click  in the toolbar. Alternatively, right-click on Des 4 and select **Simulate**. A new window for simulation will appear. Click on the **Response Generation Info** tab, and specify: **Mean control** = 9; **Mean Treatment** = 9; **SD Control** = 10.

Simulation Parameters
Response Generation Info
Simulation Control Info

Generate Data Using: Individual Means

Mean Control (μ_c): 9

Mean Treatment (μ_t): 9

Common Standard Deviation

SD Control (σ_c): 10

SD Treatment (σ_t): 10

Click **Simulate**. Once the simulation run has completed, East will add an additional

12 Normal Non-inferiority Two-Sample

row to the **Output Preview** labeled as Sim 1.

Select Sim 1 in the **Output Preview** and click . Double-click Sim 1 in the **Library**. The simulation output details will be displayed.

| Simulation Parameters | |
|--------------------------------|------------------|
| Simulation ID | Sim1 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 487 |
| Noninf. Margin (δ_0) | -3 |
| Test Statistic | t |
| Avg. Power at Termination | 0.899 |
| Response Generation Parameters | |
| Generate Data Using | Individual Means |
| Mean Control (μ_c) | 9 |
| Mean Treatment (μ_t) | 9 |
| SD Control (σ_c) | 10 |
| SD Treatment (σ_t) | 10 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

⊖ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 162 |
| 2 | 325 |
| 3 | 487 |
| Average | 350.215 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 162 | 2.774 | 2026 | 2026 | 20.260% |
| 2 | 325 | 2.346 | 4379 | 4379 | 43.790% |
| 3 | 487 | 2.062 | 2580 | 3595 | 35.950% |
| Total | | | 8985 | 10000 | |
| % | | | 89.850% | | |

The upper efficacy stopping boundary was crossed around 90% of times, out of 10,000 simulated trials, which is consistent with the power of 90%. The exact result of the simulations may differ slightly, depending on the seed.

12.1.4 Interim Monitoring

Select Des 4 in the **Library**, and click  from the **Library** toolbar. Alternatively,

right-click on Des 4 and select **Interim Monitoring**.

| Enter Interim Data | | | | | | | | | | | Interim Monitoring: Des4 | | | | |
|--------------------|----------------------|------------------------|----------------|----------|----------------|----------|------------------------|-------|------------------|----|--------------------------|--|--|--|--|
| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 97.5% RCI for δ | | Repeated p-value | CP | Predicti... Power | | | | |
| | | | | | | | Upper | Lower | | | | | | | |
| 1 | | | | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | | | |

Effect Size under H1: $\delta_1 = (\mu_1 - \mu_2) = 0$

Select the Look # 1 row for which data entry is desired and click the "Enter Interim Data" button on the toolbar.

Stopping Boundaries

| Sample Size | Efficacy |
|-------------|----------|
| | |

Conditional Power

| Trmt. Eff. | CP |
|------------|-------|
| -3 | 0.025 |
| -2.694 | 0.051 |
| -2.327 | 0.107 |
| -1.959 | 0.198 |
| -1.592 | 0.326 |
| -1.224 | 0.479 |
| -0.857 | 0.636 |
| -0.49 | 0.773 |
| -0.122 | 0.875 |

Error Spending

| Info. Fraction | α |
|----------------|----------|
| | |

Confidence Intervals

| Info. Fraction | RCI Upper | RCI Lower |
|----------------|-----------|-----------|
| | | |

The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs. The bottom section contains four charts, each with a corresponding table to its right. These charts provide graphical and numerical descriptions of the progress of the clinical trial and are useful tools for decision making by a data monitoring committee.

Although the study has been designed assuming three equally spaced analyses, departures from this strategy are permissible using the spending function methodology of Lan and DeMets (1983) and its extension to boundaries for early stopping in favor of H_0 proposed by Pampallona, Tsiatis and Kim (2001). At each interim analysis time point, East will determine the amount of type-1 error probability and type-2 error probability that it is permitted to spend based on the chosen spending functions specified in the design. East will then re-compute the corresponding stopping boundaries. This strategy ensures that the overall type-1 error does not exceed the nominal significance level α .

12 Normal Non-inferiority Two-Sample

Let us take the first look after accruing 200 subjects. The test statistic at look j for testing non-inferiority is given by

$$Z_j = \frac{\hat{\delta}_j - \delta_0}{SE(\hat{\delta}_j)}$$

where $\hat{\delta}_j$ and δ_0 indicate estimated treatment difference and the non-inferiority margin, respectively. SE denotes the standard error. Suppose we have observed $\hat{\delta}_j = 2.3033$ and $SE(\hat{\delta}_j) = 2.12132$. With $\delta_0 = -3$, the value of test statistic at first look would be $Z_1 = (2.3033 + 3)/2.12132$ or 2.5.

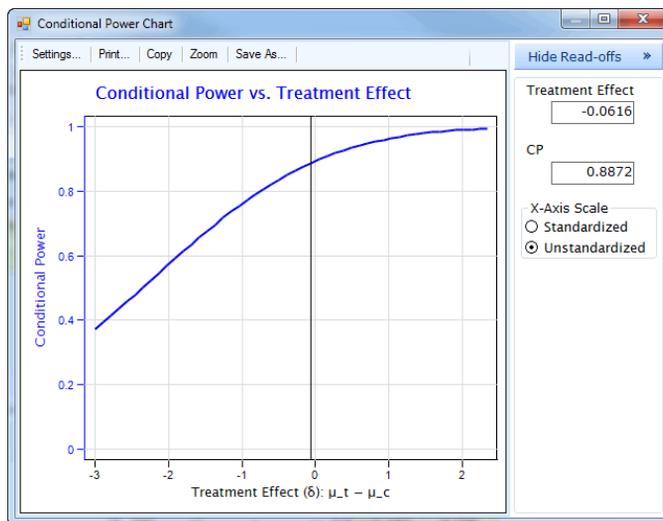
To pass these values to East, click **Enter Interim Data** to open the **Test Statistic Calculator**. Enter the following values: 200 for **Cumulative Sample Size**, 2.3033 as **Estimate of δ** and 2.12132 as **Standard Error of Estimate of δ** . Click **Recalc**, and then **OK**.

The screenshot shows a window titled "Test Statistic Calculator". It has a "Recalc" button on the left and "OK" and "Cancel" buttons on the right. The main area is titled "Editing Look #1" and contains a checkbox "Set Current Look as Last" which is unchecked. Below this are three input fields: "Cumulative Sample Size" with the value 200, "Estimate of δ " with the value 2.3033, and "Standard Error of Estimate of δ " with the value 2.1232. Below these are two output fields: "Estimate of $\delta - \delta_0$ " with the value 5.303 and "Test Statistic" with the value 2.498. The output fields have a yellow background.

The value of test statistic is 2.498, which is very close to the stopping boundary 2.634. The lower bound of 97.5% repeated confidence interval (RCI) for δ is -3.29.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 97.5% RCI for δ | |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|------------------------|--------|
| | | | | | | | Upper | Lower |
| 1 | 0.411 | 200 | 2.5 | 2.303 | 2.121 | 2.634 | Infinity | -3.285 |
| 2 | | | | | | | | |
| 3 | | | | | | | | |

Click the  icon in the **Conditional Power** chart located in lower part of the dashboard.

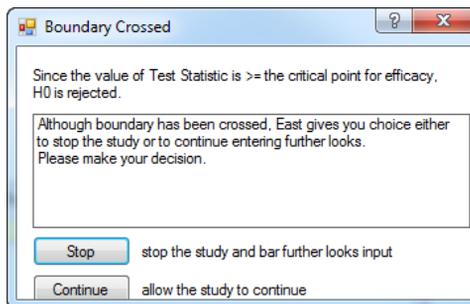


The conditional power at the current effect size 2.303 is over 99.3%.

Suppose we take the next interim look after accruing 350 subjects. Enter **350** for **Cumulative Sample Size**, **2.3033** for **Estimate of δ** and **1.71047** for **Standard Error of Estimate of δ** . Click **Recalc** and **OK** to update the charts and tables in the dashboard.

12 Normal Non-inferiority Two-Sample

Now the stopping boundary is crossed and the following dialog box appears.



Click **Stop**. The dashboard will now include the following table.

| Final Inference | |
|--|--------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.005 |
| Adj. Pt. Est. for δ | 2.013 |
| Adj. 95% CI for δ | |
| Upper Confidence Bound | 5.486 |
| Lower Confidence Bound | -1.692 |
| Post-Hoc Power | |

The adjusted confidence interval and p-value are calculated according to the approach proposed by Tsiatis, Rosner and Mehta (1984) and later extension by Kim and DeMets (1987). The basic idea here is to search for the confidence bounds such that the p-value under the alternative hypothesis just becomes statistically significant.

12.1.5 Trial Design Using a t-Test (Single Look)

In Section 12.1 the sample size is obtained based on asymptotic approximation of the distribution of the test statistics

$$\frac{\hat{\delta} - \delta_0}{\sqrt{\text{var}[\hat{\delta}]}}$$

If the study under consideration is small, the above asymptotic approximation of the distribution may be poor. Using the student's t-distribution with $(n - 1)$ degrees of freedom, we may better size the trial to have appropriate power to reject the H_0 . In East, this can be done through specifying distribution of test statistic as t . We shall illustrate this by designing the study described in Section 12.1 that aims to demonstrate that the AII inhibitor is non-inferior to the ACE inhibitor.

Select Des 1 from the **Library**. Click  from the toolbar. Change the **Test Statistic** from **Z** to **t**. The entries for the other fields need not be changed. Click **Compute**. East will add an additional row to the **Output Preview** labeled as Des 5. The required sample size is 469. Select the rows corresponding to Des 1 and Des 5 and . This will display both designs in the **Output Summary**.

| | Des 1 | Des 5 |
|---|------------------|------------------|
| Mnemonic | MN-2S-DI | MN-2S-DI |
| Test Parameters | | |
| Design Type | Noninferiority | Noninferiority |
| No. of Looks | 1 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Input Method | Individual Means | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 0 | 0 |
| Noninf. Margin ($\delta_0 = \mu_t - \mu_c$) | -3 | -3 |
| Mean Control (μ_c) | 9 | 9 |
| Mean Treatment (μ_t) | 9 | 9 |
| Mean Treatment (μ_0) | 6 | 6 |
| Std. Deviation (σ) | 10 | 10 |
| Test Statistic | Z | t |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Sample Size | | |
| Maximum | 467 | 469 |

Des 5, which used the t distribution, requires us to commit a combined total of 469 patients to the study, up from 467 in Des 1, which used the normal distribution. The extra patients are needed to compensate for the extra variability due to estimation of the $\text{var}[\hat{\delta}]$.

12.2 Ratio of Means

12.2.1 Trial design

12.2.2 Designing the study

12.2.3 Simulation

Let μ_t and μ_c denote the means of the observations from the experimental treatment (T) and the control treatment (C), respectively, and let σ_t^2 and σ_c^2 denote the corresponding variances of the observations. It is assumed that $\sigma_t/\mu_t = \sigma_c/\mu_c$, i.e. the coefficient of variation $CV = \sigma/\mu$ is the same for t and c . Finally, let $\rho = \mu_t/\mu_c$.

For a non-inferiority trial with ratio of means we define the null hypothesis as

- $H_0 : \rho \leq \rho_0$ if $\rho_0 < 1$
- $H_0 : \rho \geq \rho_0$ if $\rho_0 > 1$

where ρ_0 denotes the noninferiority margin. Consider the case when $\rho_0 < 1$. Now define $\delta = \ln(\rho) = \ln(\mu_t) - \ln(\mu_c)$, so the null hypothesis becomes $H_0: \delta \leq \delta_0$ where $\delta_0 = \ln(\rho_0)$.

Since we can translate the ratio hypothesis into a difference hypothesis, we can perform the test for difference as discussed in section 12.1 on log-transformed data.

12 Normal Non-inferiority Two-Sample

Here, we need the standard deviation of log transformed data. If we are provided with the coefficient of variation (CV) instead, the standard deviation of log transformed data can be obtained using the relation $sd = \sqrt{\ln(1 + CV^2)}$.

12.2.1 Trial design

For illustration, we consider the example cited by Laster and Johnson (2003): A randomized clinical study of a new anti-hypertensive therapy known to produce fewer side-effects than a standard therapy but expected to be almost 95% effective ($\rho_1 = 0.95$). To accept the new therapy, clinicians want a high degree of assurance that it is at least 80% as effective in lowering blood pressure as the standard agent. Accordingly we plan to design the study to test:

$$H_0 : \mu_t / \mu_c \leq 0.8$$

against

$$H_1 : \mu_t / \mu_c > 0.8$$

Reductions in seated diastolic blood pressure are expected to average 10 mmHg ($= \mu_c$) with standard therapy with standard deviation as 7.5 mmHg ($= \sigma_c$). Therefore, CV in the standard therapy is $7.5/10 = 0.75$. We also assume that CV in both therapies are equal. We need to design a study that would have 90% power at $\rho_1 = 0.95$ under H_1 and maintains one-sided type I error at 5%.

12.2.2 Designing the study

Start East afresh. Click **Continuous: Two Samples**, under the **Design** tab, and then click **Parallel Design: Ratio of Means**.

In the input window, select **Noninferiority** for **Design Type**. Select **Individual Means** for **Input Method** and then specify the **Mean Control** (μ_c) as 10, **Noninferiority Margin** (ρ_0) as 0.8 and **Ratio of Means** (ρ_1) as 0.95. Specify 0.75

value for **Coeff. Var.**. The upper pane should appear as below:

Click **Compute**. This will calculate the sample size for this design, and the output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed total sample size (636 subjects) is highlighted in yellow.

| ID | Design Type | Test Type | Specified α | Power | Input Method | Sample Size | nt/nc | Test Statistic | ρ | μ_c | Mean Treatment (Alt.) | ρ_0 | C V of Data |
|-------|----------------|-----------|--------------------|-------|------------------|-------------|-------|----------------|--------|---------|-----------------------|----------|-------------|
| Des 1 | Noninferiority | 1-Sided | 0.025 | 0.9 | Individual Means | 636 | 1 | Z | 0.95 | 10 | 9.5 | 0.8 | 0.75 |

This design has default name Des 1. Select this design by clicking anywhere along the row in the **Output Preview** and click . Some of the design details will be

12 Normal Non-inferiority Two-Sample

displayed in the **Output Summary**.

| | Des 1 |
|------------------------------------|------------------|
| Mnemonic | MN-2S-RA |
| Test Parameters | |
| Design Type | Noninferiority |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Ratio of Means (ρ) | 0.95 |
| Mean Control (μ_c) | 10 |
| Mean Treatment (μ_t) | 9.5 |
| Noninferiority Margin (ρ_0) | 0.8 |
| Coeff. of Variation Data | 0.75 |
| Input Method | Individual Means |
| Allocation Ratio (nt/nc) | 1 |
| Test Statistic | Z |
| Sample Size | |
| Maximum | 636 |

In the **Output Preview** toolbar, click  to save this design to Wbk1 in the **Library**. Double-click on Des 1 in the **Library** to see the details of the design.

Design: Continuous Endpoint: Two-Sample Test - Parallel Design - Ratio of Means

| Test Parameters | |
|---------------------------------------|------------------|
| Design ID | Des1 |
| Design Type | Noninferiority |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| Sample Size (n) | To be Computed |
| Model Parameters | |
| Test Statistic | Z |
| Input Method | Individual Means |
| Mean Treatment (μ_t) | 9.5 |
| Mean Control (μ_c) | 10 |
| Mean Ratio ($\rho = \mu_t / \mu_c$) | 0.95 |
| $\rho = \mu_t / \mu_c$ | |
| Under H0 | 0.8 |
| Under H1 | 0.95 |
| Noninferiority Margin (ρ_0) | 0.8 |
| Allocation Ratio (n_t/n_c) | 1 |
| Coefficient of Variation of Data | 0.75 |

Sample Size Information

| | |
|-----------------------------|-----|
| Sample Size (n) | 636 |
| Treatment (n _t) | 318 |
| Control (n _c) | 318 |

Critical Points

| | |
|----------------|------|
| Critical Point | 1.96 |
|----------------|------|

Unequal allocation ratio

Since the profile of standard therapy is well established and comparatively little is known about the new therapy, you want to put more subjects on the new therapy. You can do this by specifying allocation ratio greater than 1. Suppose you want 50% more subjects on new therapy compared to standard one. Then we need to specify allocation ratio (n_t/n_c) as 1.5.

Create a new design by selecting Des 1 in the **Output Preview**, and clicking  on the Output toolbar. In the Input, change the **Allocation Ratio** from 1 to 1.5. Click **Compute** to obtain sample size for this design. A new row will be added labeled as Des 2.

Save this design in the current workbook by selecting the corresponding row in **Output Preview** and clicking  on the **Output Preview** toolbar. Select both rows in **Output Preview** using the Ctrl key and click .

| | Des 1 | Des 2 |
|------------------------------------|------------------|------------------|
| Mnemonic | MN-25-RA | MN-25-RA |
| Test Parameters | | |
| Design Type | Noninferiority | Noninferiority |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Ratio of Means (ρ) | 0.95 | 0.95 |
| Mean Control (μ_c) | 10 | 10 |
| Mean Treatment (μ_t) | 9.5 | 9.5 |
| Noninferiority Margin (ρ_0) | 0.8 | 0.8 |
| Coeff. of Variation Data | 0.75 | 0.75 |
| Input Method | Individual Means | Individual Means |
| Allocation Ratio (n_t/n_c) | 1 | 1.5 |
| Test Statistic | Z | Z |
| Sample Size | | |
| Maximum | 636 | 662 |

t distribution test statistic

Create a new design by selecting Des 2 in the **Output**, and clicking  on the Output toolbar. In the Input, change the **Test Statistic** from Z to t. Click **Compute** to obtain sample size for this design. A new row will be added labeled as Des 3.

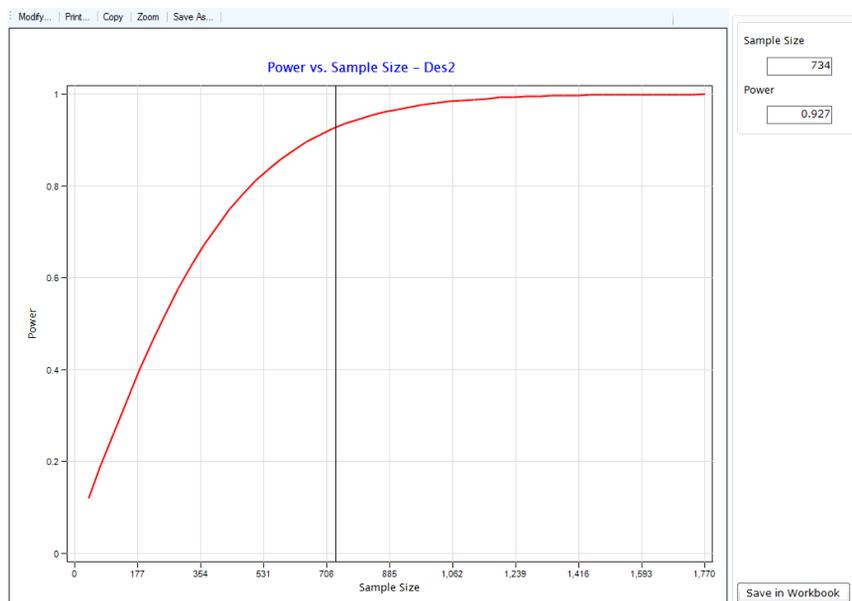
| ID | Trial Type | Test Type | Specified α | Specified Power | Attained Power | Input Method | Sample Size |
|------|----------------|-----------|--------------------|-----------------|----------------|------------------|-------------|
| Des1 | Noninferiority | 1-Sided | 0.025 | 0.9 | 0.9 | Individual Means | 636 |
| Des2 | Noninferiority | 1-Sided | 0.025 | 0.9 | 0.9 | Individual Means | 662 |
| Des3 | Noninferiority | 1-Sided | 0.025 | 0.9 | 0.9 | Individual Means | 664 |

12 Normal Non-inferiority Two-Sample

A sample size of 664 will be needed, which is very close to the sample size 662 obtained in Des 2 under the normal distribution.

Plotting

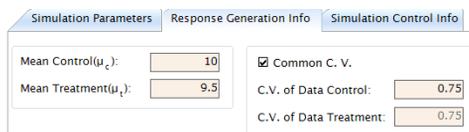
With Des 2 selected in the **Library**, click  on the **Library** toolbar, and then click **Power vs Sample Size**. The resulting power curve for this design will appear.



You can export this chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...** Feel free to explore other plots as well. Once you have finished, close all charts before continuing.

12.2.3 Simulation

Select Des 2 in the **Library**, and click  in the toolbar. Alternatively, right-click on Des 2 and select **Simulate**. A new Simulation window will appear. Click on the **Response Generation Info** tab, and specify: **Mean control** = 10; **Mean Treatment** = 9.5; **CV of Data Control** = 0.75.



| Simulation Parameters | Response Generation Info | Simulation Control Info |
|-----------------------------|--------------------------|--|
| Mean Control (μ_c): | 10 | <input checked="" type="checkbox"/> Common C. V. |
| Mean Treatment (μ_t): | 9.5 | C.V. of Data Control: 0.75 |
| | | C.V. of Data Treatment: 0.75 |

Click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1.

Select Sim 1 in the **Output Preview** and click . Double-click on Sim 1 in the **Library**. The simulation output details will be displayed.

Simulation: Continuous Endpoint: Two-Sample Test - Parallel Design - Ratio of Means

| Simulation Parameters | |
|---|----------------|
| Simulation ID | Sim2 |
| Trial Type | Noninferiority |
| Test Type | 1-Sided |
| Sample Size (n) | 662 |
| Test Statistic | t |
| Noninferiority Margin (ρ_0) | 0.8 |
| Response Generation Parameters | |
| Mean Response under Control (μ_c) | 10 |
| Mean Response under Treatment (μ_t) | 9.5 |
| C.V. of Data Control | 0.75 |
| C.V. of Data Treatment | 0.75 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Simulation Boundaries

Critical Point: 1.96

Overall Simulation Results

| | Upper H0 | Lower H0 |
|-------------------|----------|----------|
| No. of Rejections | 9041 | NA |
| % | 90.41 | NA |

Starting Seed: 1500235
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:03

Out of 10,000 simulations, close to 90% are rejected for non-inferiority. Therefore, the simulation result verifies that the design attains 90% power. The simulation result might vary depending on the starting seed value chosen.

12 Normal Non-inferiority Two-Sample

12.3 Difference of Means in Crossover Designs

12.3.1 Trial Design

In a 2×2 crossover design each subject is randomized to one of two sequence groups. Subjects in sequence group 1 receive the test drug (T) formulation in a first period, have their outcome variable, X recorded, wait out a washout period to ensure that the drug is cleared from their system, then receive the control drug formulation (C) in period 2 and finally have the measurement on X again. In sequence group 2, the order in which the T and C are assigned is reversed. The table below summarizes this type of trial design.

| Group | Period 1 | Washout | Period 2 |
|-------|----------|---------|----------|
| 1(TC) | Test | — | Control |
| 2(CT) | Control | — | Test |

The resulting data are commonly analyzed using a linear model. The response y_{ijk} in period j on subject k in sequence group i , where $i = 1, 2$, $j = 1, 2$, and $k = 1, \dots, n_i$ is modeled as a linear function of an overall mean response μ , formulation effect τ_t and τ_c , period effects π_1 and π_2 , and sequence effects γ_1 and γ_2 . The fixed effects model can be displayed as:

| Group | Period 1 | Washout | Period 2 |
|-------|-----------------------------------|---------|-----------------------------------|
| 1(TC) | $\mu + \tau_t + \pi_1 + \gamma_1$ | — | $\mu + \tau_c + \pi_2 + \gamma_1$ |
| 2(CT) | $\mu + \tau_c + \pi_1 + \gamma_2$ | — | $\mu + \tau_t + \pi_2 + \gamma_2$ |

Let $\mu_t = \mu + \tau_t$ and $\mu_c = \mu + \tau_c$ denote the means of the observations from the test and control formulations, respectively, and let MSE denote the mean-squared error.

In a noninferiority trial, we test $H_0 : \delta \leq \delta_0$ against $H_0 : \delta > \delta_0$ if $\delta_0 < 0$ or $H_0 : \delta \geq \delta_0$ against $H_0 : \delta < \delta_0$ if $\delta_0 > 0$, where δ_0 indicates the noninferiority margin.

East uses following test statistic to test the above null hypothesis

$$T_L = \frac{(\bar{y}_{11} - \bar{y}_{12} - \bar{y}_{21} + \bar{y}_{22})/2 - \delta_0}{\sqrt{\frac{\hat{\sigma}^2}{2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where, \bar{y}_{ij} is the mean of the observations from group i and period j and $\hat{\sigma}^2$ is the estimate of error variance. T_τ is distributed with Student's t distribution with $(n_1 + n_2 - 2)$ degrees of freedom.

12.3.1 Trial Design

Consider a 2×2 crossover trial between a Test drug (T) and a Reference Drug (C) where the noninferiority need to be established in terms of some selected treatment response. Let μ_T and μ_C denote the mean of Test and Reference drugs, respectively. Let $\delta = \mu_t - \mu_c$ be the difference in averages. The noninferiority margin were set at -3. Accordingly we plan to design the study to test:

$$H_0 : \mu_t - \mu_c \leq -3$$

against

$$H_1 : \mu_t - \mu_c > -3$$

For this study, we consider $\mu_c = 21.62$ ng.h/mL and $\mu_t = 23.19$ ng.h/mL under H_1 . Further we assume mean squared error (MSE) would be 2.5. We want to design a study that would have 90% power at $\delta_1 = 23.19 - 21.62 = 1.57$ under H_1 . We want to perform this test at a one sided 0.025 level of significance.

Start East afresh. First, **Continuous: Two Samples** on the **Design** tab, and then click **Crossover Designs: Difference of Means**.

In the input window, select **Noninferiority** for **Design Type**. Select **Individual Means** for **Input Method** and then specify the **Mean Control (μ_c)** as 21.62 and **Mean Treatment (μ_t)** as 23.19. Enter the **Type I Error (α)** as 0.025. Select **Sqrt (MSE)** from the drop-down list and enter as 2.5. Finally, enter **Noninferiority Margin (δ_0)** as -3. The upper pane should appear as below:

Design: Continuous Endpoint: Two-Sample Test - Crossover Design - Difference of Means

Design Type: Noninferiority

Design Parameters

Test Type: 1-Sided

Type I Error (α): 0.025

Power: 0.9

Sample Size (n): Computed

Allocation Ratio: 1
(n_{TC} / n_{CT})

Input Method: Individual Means

Specify Null Hypothesis
Noninferiority Margin (δ_0): -3

Sqrt(MSE): 2.5

Specify Mean Responses
Mean Control (μ_c): 21.62
Mean Treatment (μ_t): 23.19

Compute

Click **Compute**. The sample size required for this design is highlighted in yellow. Save this design in the current workbook by selecting the corresponding row in

12 Normal Non-inferiority Two-Sample

Output Preview and clicking  on the **Output Preview** toolbar. Double-click Des 1 in **Library**. This will display the design details. The sample size required for Des 1 is only 9 to establish non-inferiority with 90% power.

Design: Continuous Endpoint: Two-Sample Test - Crossover Design - Difference of Means

| Test Parameters | |
|--------------------------------------|------------------|
| Design ID | Des1 |
| Design Type | Noninferiority |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.912 |
| Sample Size (n) | To be Computed |
| Model Parameters | |
| Test Statistic | t |
| Input Method | Individual Means |
| Mean Treatment (μ_t) | 23.19 |
| Mean Control (μ_c) | 21.62 |
| $\delta = \mu_t - \mu_c$ | |
| Under H0 | -3 |
| Under H1 | 1.57 |
| Noninferiority Margin (δ_0) | -3 |
| Sqrt(MSE) | 2.5 |
| Allocation Ratio (n_c/n_t) | 1 |

| Sample Size Information | |
|-------------------------|---|
| Sample Size (n) | 9 |
| Treatment (n_t) | 4 |
| Control (n_c) | 5 |

| Critical Points | |
|-----------------|-------|
| Critical Point | 2.365 |

12.4 Ratio of Means in Crossover Designs

12.4.1 Trial Design

We consider the same anti-hypertensive therapy example discussed in section 12.2, but this time we will assume that the data has come from a crossover design. We wish to test the following hypotheses:

$$H_0 : \mu_t / \mu_c \leq 0.8$$

against

$$H_1 : \mu_t / \mu_c > 0.8$$

We want the study to have at least 90% power at $\rho_1 = 0.95$ and maintains one-sided test I error at 5%. As before, we will consider $CV = 0.75$ for both treatment arms.

Start East afresh. First, click **Continuous: Two Samples** under the **Design** tab, and then click **Crossover Design: Ratio of Means**.

In the input window, select **Noninferiority** for **Design Type**. Select **Individual Means** for **Input Method** and then specify the **Noninferiority**

Margin (ρ_0) as 0.8, **Mean Control** (μ_c) as 10, and **Mean Treatment** (μ_t) as 9.5. Using the relationship between CV (=0.75) and standard deviation of log-transformed data mentioned in section 12.2, we have standard deviation for log-transformed data as 0.45. Specify 0.45 for **Sqrt. of MSE Log**. The upper pane should appear as below:

Click **Compute**. The sample size required for this design is highlighted in yellow in the Output Preview pane. Save this design in the current workbook by selecting the corresponding row in **Output Preview** and clicking on the **Output Preview** toolbar. Select Des 1 in **Library** and click . This will display the design details.

Design: Continuous Endpoint: Two-Sample Test - Crossover Design - Ratio of Means

| Test Parameters | |
|--|------------------|
| Design ID | Des1 |
| Design Type | Noninferiority |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.902 |
| Sample Size (n) | To be Computed |
| Model Parameters | |
| Test Statistic | t |
| Input Method | Individual Means |
| Mean Treatment (μ_t) | 9.5 |
| Mean Control (μ_c) | 10 |
| $\rho = \mu_t / \mu_c$ | |
| Under H0 | 0.8 |
| Under H1 | 0.95 |
| Noninferiority Margin (ρ_0) | 0.8 |
| Sqrt(MSE) Log | 0.45 |
| Allocation Ratio (n_{tc} / n_{ct}) | 1 |

Sample Size Information

| | |
|-----------------------------|-----|
| Sample Size (n) | 147 |
| Treatment (n _t) | 73 |
| Control (n _c) | 74 |

Critical Points

| | |
|----------------|-------|
| Critical Point | 1.977 |
|----------------|-------|

12 *Normal Non-inferiority Two-Sample*

In general, a crossover design requires fewer subjects compared to its parallel design counterpart, and may be preferred whenever it is feasible.

13

Normal Equivalence Two-Sample

In many cases, the goal of a clinical trial is neither superiority nor non-inferiority, but equivalence. In Section 13.1, the problem of establishing the equivalence with respect to the difference of the means of two normal distributions using a parallel-group design is presented. The corresponding problem of establishing equivalence with respect to the log ratio of means is presented in Section 13.2. For the crossover design, the problem of establishing equivalence with respect to the difference of the means is presented in Section 13.3, and with respect to the log ratio of means in Section 13.4.

13.1 Difference in Means

13.1.1 Trial design

13.1.2 Simulation

In some experimental situations, we want to show that the means of two normal distributions are “close”. For example, a test formulation of a drug (T) and the control (or reference) formulation of the same drug (C) are considered to be bioequivalent if the rate and extent of absorption are similar. Let μ_t and μ_c denote the means of the observations from the test and reference formulations, respectively, and let σ^2 denote the common variance of the observations. The goal is to establish that $\delta_L < \mu_t - \mu_c < \delta_U$, where δ_L and δ_U are a-priori specified values used to define equivalence. The two one-sided tests (TOST) procedure of Schuirmann (1987) is commonly used for this analysis, and is employed in this section for a parallel-group study.

Let $\delta = \mu_t - \mu_c$ denote the true difference in the means. The null hypothesis $H_0: \delta \leq \delta_L$ or $\delta \geq \delta_U$ is tested against the two-sided alternative hypothesis $H_1: \delta_L < \delta < \delta_U$ at level α , using TOST. Here we perform the following two tests together:

- Test1: $H_{0L}: \delta \leq \delta_L$ against $H_{1L}: \delta > \delta_L$ at level α
- Test2: $H_{0U}: \delta \geq \delta_U$ against $H_{1U}: \delta < \delta_U$ at level α

H_0 is rejected in favor of H_1 at level α if and only if both H_{0L} and H_{0U} are rejected. Note that this is the same as rejecting H_0 in favor of H_1 at level α if the $(1 - 2\alpha)$ 100% confidence interval for δ is completely contained within the interval (δ_L, δ_U) .

Let N be the total sample size and $\hat{\mu}_t$ and $\hat{\mu}_c$ denote the estimates of the means T and C, respectively. Let $\hat{\delta} = \hat{\mu}_t - \hat{\mu}_c$ denote the estimated difference with standard error $se(\hat{\delta})$

We use the following two test statistics to apply Test1 and Test2, respectively:

$$T_L = \frac{(\hat{\delta} - \delta_L)}{se(\hat{\delta})}, T_U = \frac{(\hat{\delta} - \delta_U)}{se(\hat{\delta})}$$

13 Normal Equivalence Two-Sample

T_L and T_U are assumed to follow Student's t-distribution with $(N - 2)$ degrees of freedom under H_{0L} and H_{0U} , respectively. H_{0L} is rejected if $T_L \geq t_{1-\alpha, (N-2)}$, and H_{0U} is rejected if $T_U \leq t_{\alpha, (N-2)}$.

The null hypothesis H_0 is rejected in favor of H_1 if $T_L \geq t_{1-\alpha, (N-2)}$ and $T_U \leq t_{\alpha, (N-2)}$, or in terms of confidence intervals: Reject H_0 in favor of H_1 at level α if

$$\delta_L + t_{1-\alpha, (N-2)} 2\hat{\sigma} / \sqrt{N} < \hat{\delta} < \delta_U + t_{\alpha, (N-2)} 2\hat{\sigma} / \sqrt{N}. \quad (13.1)$$

We see that decision rule (13.1) is the same as rejecting H_0 in favor of H_1 if the $(1 - 2\alpha)$ 100% confidence interval for δ is entirely contained within interval (δ_L, δ_U) .

The above inequality (13.1) cannot hold if $4t_{1-\alpha, (N-2)} \hat{\sigma} / \sqrt{N} \geq (\delta_U - \delta_L)$, in which case H_0 is not rejected in favor of H_1 . Thus, we assume that $4t_{1-\alpha, (N-2)} \hat{\sigma} / \sqrt{N} < (\delta_U - \delta_L)$. The impact of this assumption was examined by Bristol (1993a).

The power or sample size of such a trial design is determined for a specified value of δ , denoted δ_1 , for a single-look study only. The choice $\delta_1 = 0$, i.e. $\mu_t = \mu_c$, is common. For a specified value of δ_1 , the power is given by

$$\Pr(\text{Reject } H_0) = 1 - \tau_\nu(t_{\alpha, \nu} | \Delta_1) + \tau_\nu(-t_{\alpha, \nu} | \Delta_2) \quad (13.2)$$

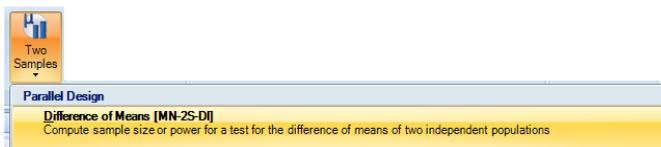
where $\nu = N - 2$ and Δ_1 and Δ_2 are non-centrality parameters given by $\Delta_1 = (\delta_1 - \delta_L) / se(\hat{\delta})$ and $\Delta_2 = (\delta_1 - \delta_U) / se(\hat{\delta})$, respectively. $t_{\alpha, \nu}$ denotes the upper $\alpha \times 100\%$ percentile from a Student's t distribution with ν degrees of freedom. $\tau_\nu(x | \Delta)$ denotes the distribution function of a non-central t distribution with ν degrees of freedom and non-centrality parameter Δ , evaluated at x .

Since we don't know the sample size N ahead of time, we cannot characterize the bivariate t-distribution. Thus solving for sample size must be performed iteratively by equating the formula (13.2) to the power $1 - \beta$.

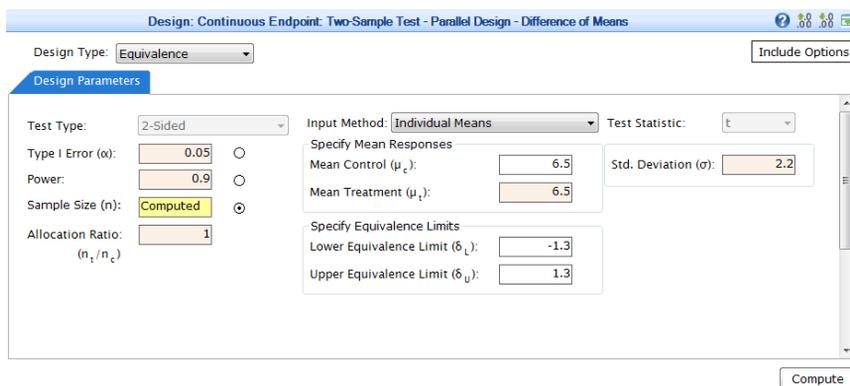
13.1.1 Trial design

Consider the situation where we need to establish equivalence between a test formulation of capsules (T) with the marketed capsules (C). The response variable is the change from baseline in total symptom score. Based on the studies conducted during the development program, it is assumed that $\mu_c = 6.5$. Based on this value, equivalence limits were set as $-\delta_L = \delta_U = 1.3 (= 20\% \mu_c)$. We assume that the common standard deviation is $\sigma = 2.2$. We want to have 90% power at $\mu_t = \mu_c$.

Start East afresh. Click **Continuous: Two Samples** on the **Design** tab and then click **Parallel Design: Difference of Means**.



This will launch a new window. The upper pane of this window displays several fields with default values. Select **Equivalence** for **Design Type**, and **Individual Means** for **Input Method**. Enter 0.05 for **Type I Error**. Specify both **Mean Control** (μ_c) and **Mean Treatment** (μ_t) as 6.5. We have assumed $\sigma = 2.2$. Enter this value for **Std. Deviation** (σ). Also enter -1.3 for **Lower Equivalence Limit** (δ_L) and 1.3 for **Upper Equivalence Limit** (δ_U). The upper pane should appear as below:



Click **Compute**. The output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed sample size (126 subjects) is highlighted in yellow.

| ID | Design Type | Test Type | Specified α | Power | nt/nc | Sample Size | Input Method | δ_L | μ_c | Mean Treatment (Alt.) | δ_L | δ_U | σ | Test Statistic |
|------|-------------|-----------|--------------------|-------|-------|-------------|------------------|------------|---------|-----------------------|------------|------------|----------|----------------|
| Des1 | Equivalence | 2-Sided | 0.05 | 0.902 | 1 | 126 | Individual Means | 0 | 6.5 | 6.5 | -1.3 | 1.3 | 2.2 | t |

This design has default name Des 1. Select this design and click  in the **Output Preview** toolbar. Some of the design details will be displayed in the upper pane,

13 Normal Equivalence Two-Sample

labeled as **Output Summary**.

| Des 1 | |
|---|------------------|
| Mnemonic | MN-2S-DI |
| Test Parameters | |
| Design Type | Equivalence |
| No. of Looks | 1 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.902 |
| Model Parameters | |
| Input Method | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 0 |
| Mean Control (μ_c) | 6.5 |
| Mean Treatment (μ_t) | 6.5 |
| Equivalence Lower Limit (δ_L) | -1.3 |
| Equivalence Upper Limit (δ_U) | 1.3 |
| Std. Deviation (σ) | 2.2 |
| Test Statistic | t |
| Allocation Ratio (nt / nc) | 1 |
| Sample Size | |
| Maximum | 126 |

A total of 126 subjects must be enrolled in order to achieve the desired 90% power under the alternative hypothesis. Of these 126 subjects 63 will be randomized to the test formulation, and the remaining 63 to the marketed formulation. In the **Output Preview** toolbar, select Des 1 and click  to save this design to Wbk1 in the **Library**.

Suppose that this sample size is not economically feasible and we want to examine power for a total sample size of 100. Select Des 1 in the **Library**, and click  on the **Library** toolbar. In the Input, click the radiobutton for **Power**, and enter **Sample Size** (n) as 100.

Design Parameters

Test Type: 2-Sided ▾

Type I Error (α): 0.05

Power: Computed

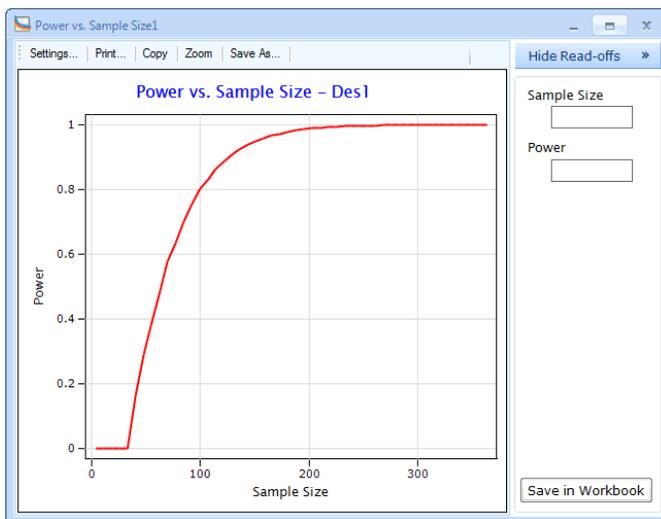
Sample Size (n): 100

Allocation Ratio: 1
(n_t / n_c)

Click **Compute**. This will add a new row to the **Output Preview** and the calculated power is highlighted in yellow. We see that a power of 80.3% can be achieved with 100 subjects.

| ID | Design Type | Test Type | Specified α | Power | nt/nc | Sample Size |
|------|-------------|-----------|--------------------|-------|-------|-------------|
| Des1 | Equivalence | 2-Sided | 0.05 | 0.902 | 1 | 126 |
| Des2 | Equivalence | 2-Sided | 0.05 | 0.803 | 1 | 100 |

Suppose we want to see how the design parameters such as power, sample size and treatment effect are interrelated. To visualize any particular relationship for Des 1, first select Des 1 from **Library** and then click  in the toolbar. You will see a list of options available. To plot power against sample size, click **Power vs Sample Size**.



Feel free to explore other plots and options available with them. Close the charts before continuing.

13.1.2 Simulation

We wish to make sure that Design 1 has the desired power of 90%, and maintains the type I error of 5%. This examination can be conducted using simulation. Select Des 1 in the **Library**, and click  in the toolbar. Alternatively, right-click Des 1 and

13 Normal Equivalence Two-Sample

select **Simulate**. A new Simulation window will appear. Click on the **Response Generation Info** tab. We will first simulate under H_1 . Leave the default values as below, and click **Simulate**.

| Simulation Parameters | | Response Generation Info | | Simulation Control Info | |
|-----------------------------|------------------|---|-----|-------------------------|--|
| Generate Data Using: | Individual Means | <input checked="" type="checkbox"/> Common Standard Deviation | | | |
| Mean Control (μ_c): | 6.5 | SD Control (σ_c): | 2.2 | | |
| Mean Treatment (μ_t): | 6.5 | SD Treatment (σ_t): | 2.2 | | |

Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1.

Select Sim 1 in the **Output Preview** and click . Now double-click on Sim 1 in the **Library**. The simulation output details, including the table below, will be displayed.

Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|--------|----------------|----------------|-------------------|----------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 126 | 1.657 | -1.657 | 9031 | 0 | 10000 | 100.000% |
| Total | | | | 9031 | 0 | 10000 | |
| % | | | | 90.310% | 0.000% | | |

Observe that out of the 10,000 simulated trials, the null hypothesis was around 90% of the time. (Note: The numbers on your screen might differ slightly because you might be using a different starting seed for your simulations.)

Next we will simulate from a point that belongs to the null hypothesis. Consider $\mu_c = 6.5$ and $\mu_t = 7.8$. Select Sim 1 in **Library** and click  icon. Go to the **Response Generation Info** tab in the upper pane and specify: **Mean Control** (μ_c) = 6.5 and **Mean Treatment** (μ_t) = 7.8.

| Simulation Parameters | | Response Generation Info | | Simulation Control Info | |
|-----------------------------|------------------|---|-----|-------------------------|--|
| Generate Data Using: | Individual Means | <input checked="" type="checkbox"/> Common Standard Deviation | | | |
| Mean Control (μ_c): | 6.5 | SD Control (σ_c): | 2.2 | | |
| Mean Treatment (μ_t): | 7.8 | SD Treatment (σ_t): | 2.2 | | |

Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 2. Select Sim 2 in the **Output Preview** and click . Now double-click on Sim 2 in the **Library**. You can see that when H_0 is true, the simulated power is close to the specified type I error of 5%.

13.2 Ratio of Means

13.2.1 Trial design

13.2.2 Simulation

For some pharmacokinetic parameters, the ratio of the means is a more appropriate measure of the distance between the treatments. Let μ_t and μ_c denote the means of the observations from the test formulation (T) and the reference (C), respectively, and let σ_t^2 and σ_c^2 denote the corresponding variances of the observations. It is assumed that $\sigma_t/\mu_t = \sigma_c/\mu_c$, i.e. the coefficient of variation $CV = \sigma/\mu$ is the same for T and C. Finally, let $\rho = \mu_t/\mu_c$.

The goal is to establish that $\rho_L < \rho < \rho_U$, where ρ_L and ρ_U are specified values used to define equivalence. In practice, ρ_L and ρ_U are often chosen such that $\rho_L = 1/\rho_U$. The two one-sided tests procedure of Schuirmann (1987) is commonly used for this analysis, and is employed here for a parallel-group study.

The null hypothesis $H_0: \rho \leq \rho_L$ or $\rho \geq \rho_U$ is tested against the two-sided alternative hypothesis $H_1: \rho_L < \rho < \rho_U$ at level α , using two one-sided tests. Schuirmann (1987) proposed working this problem on the natural logarithm scale. Thus, we are interested in the parameter $\delta = \ln(\rho) = \ln(\mu_t) - \ln(\mu_c)$ and the null hypothesis $H_0: \delta \leq \delta_L$ or $\delta \geq \delta_U$ is tested against the two-sided alternative hypothesis $H_1: \delta_L < \delta < \delta_U$ at level α , using two one-sided t-tests. Here $\delta_L = \ln(\rho_L)$ and $\delta_U = \ln(\rho_U)$.

Since we have translated the ratio hypothesis into a difference hypothesis, we can perform the test for difference as discussed in section 13.1. Note that we need the standard deviation for log transformed data. However, if we are provided with information on CV instead, the standard deviation of log transformed data can be obtained using the relation $sd = \sqrt{\ln(1 + CV^2)}$.

13.2.1 Trial design

Suppose that the logarithm of area under the curve (AUC), a pharmacokinetic parameter related to the efficacy of a drug, is to be analyzed to compare the two formulations of a drug. We want to show that the two formulations are bioequivalent by showing that the ratio of the means satisfies $0.8 < \mu_t/\mu_c < 1.25$. Thus $\rho_L = 0.8$ and $\rho_U = 1.25$. Also, based on previous studies, it is assumed that the coefficient of variation is $CV = 0.25$.

13 Normal Equivalence Two-Sample

Start East afresh. Click **Continuous: Two Samples** on the **Design** tab and then click **Parallel Design: Ratio of Means**.

This will launch a new window. The upper pane of this window displays several fields with default values. Select **Equivalence** for **Trial Type**, and enter 0.05 for the **Type I Error**. For the **Input Method**, specify **Ratio of Means**. Enter 1 for **Ratio of Means** (ρ_1), 0.8 for **Lower Equivalence Limit** (ρ_L) and 1.25 for **Upper Equivalence Limit** (ρ_U). Specify 0.25 for **Coeff. Var.**. The upper pane should appear as below:

Click **Compute**. The output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed total sample size (55 subjects) is highlighted in yellow.

| ▲ | ID | Design Type | Test Type | Specified α | Power | Input Method | Sample Size | nt/nc | Test Statistic | ρ | ρ_L | ρ_U | C V of Data |
|---|------|-------------|-----------|--------------------|-------|----------------|-------------|-------|----------------|--------|----------|----------|-------------|
| 📄 | Des1 | Equivalence | 2-Sided | 0.05 | 0.906 | Ratio of Means | 55 | 1 | t | 1 | 0.8 | 1.25 | 0.25 |

In the **Output Preview** toolbar, click  to save this design to Wbk1 in the **Library**. Double-click Des 1 in the **Library** to see the details of the designs. Close

this output window before continuing.

Design: Continuous Endpoint: Two-Sample Test - Parallel Design - Ratio of Means

| Test Parameters | |
|---------------------------------------|----------------|
| Design ID | Des1 |
| Design Type | Equivalence |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.906 |
| Sample Size (n) | To be Computed |
| Model Parameters | |
| Test Statistic | t |
| Input Method | Ratio of Means |
| Mean Ratio ($\rho = \mu_t / \mu_c$) | 1 |
| $\rho = \mu_t / \mu_c$ | |
| Under H01 | 0.8 |
| Under H02 | 1.25 |
| Under H1 | 1 |
| Allocation Ratio (n_t/n_c) | 1 |
| Coefficient of Variation of Data | 0.25 |

Sample Size Information

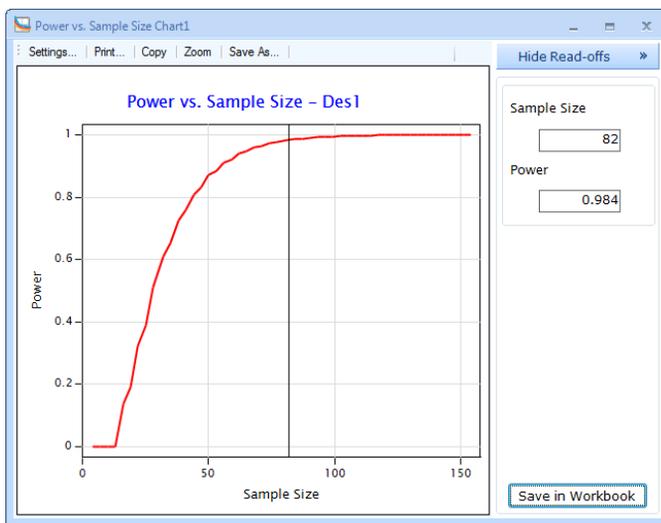
| | |
|-----------------------------|----|
| Sample Size (n) | 55 |
| Treatment (n _t) | 27 |
| Control (n _c) | 28 |

Critical Points

| | |
|----------------------|-------|
| Lower Critical Point | 1.675 |
| Upper Critical Point | 1.675 |

Plotting

With Des 1 selected in the **Library**, click  on the **Library** toolbar, and then click **Power vs Sample Size**. The resulting power curve for this design will appear.



13 Normal Equivalence Two-Sample

You can export this chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...** Feel free to explore charts. Close all chart before continuing.

13.2.2 Simulation

Suppose you suspect that CV will be smaller than 0.25; e.g., 0.2. Select Des 1 in the **Library**, and click  in the toolbar. Click on the **Response Generation Info** tab and change **C.V. of Data Control** to 0.20.

Simulation Parameters
Response Generation Info
Simulation Control Info

Mean Control(μ_c):

Mean Treatment(μ_t):

Common C. V.

C.V. of Data Control:

C.V. of Data Treatment:

Click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1. Select Sim 1 in the **Output Preview** and click . Now double-click on Sim 1 in the **Library**. The simulation output details will be displayed in the upper pane.

Simulation: Continuous Endpoint: Two-Sample Test - Parallel Design - Ratio of Means

| Simulation Parameters | |
|---|-------------|
| Simulation ID | Sim1 |
| Trial Type | Equivalence |
| Test Type | 2-Sided |
| Sample Size (n) | 55 |
| Test Statistic | t |
| Lower Equivalence Limit (ρ_L) | 0.8 |
| Upper Equivalence Limit (ρ_U) | 1.25 |
| Response Generation Parameters | |
| Mean Response under Control (μ_c) | 1 |
| Mean Response under Treatment (μ_t) | 1 |
| C.V. of Data Control | 0.2 |
| C.V. of Data Treatment | 0.2 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Simulation Boundaries

Upper Critical Point: 1.675
 Lower Critical Point: -1.675

Overall Simulation Results

| | Upper H0 | Lower H0 |
|--------------------------|----------|----------|
| No. of Rejections | 9868 | 0 |
| % | 98.68 | 0 |

Starting Seed: 5759915
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:02

Observe that out of 10,000 simulated trials, the null hypothesis was rejected over 98% of the time. (Note: The numbers on your screen might differ slightly depending on the starting seed.)

13.3 Difference of Means in Crossover Designs

13.3.1 Trial design

13.3.2 Simulation

Crossover trials are widely used in clinical and medical research. The crossover design is often preferred over a parallel design, because in the former, each subject receives all the treatments and thus each subject acts as their own control. This leads to the requirement of fewer subjects in a crossover design. In this chapter, we show how East supports the design and simulation of such experiments with endpoint as difference of means.

In a 2×2 crossover design each subject is randomized to one of two sequence groups (or, treatment sequences). Subjects in sequence group 1 receive the test drug (T) formulation in a first period, have their outcome variable, X recorded, wait out a washout period to ensure that the drug is cleared from their system, then receive the control drug formulation (C) in period 2 and finally have the measurement on X again. In sequence group 2, the order in which the T and C are assigned is reversed. The table below summarizes this type of trial design.

| Group | Period 1 | Washout | Period 2 |
|-------|----------|---------|----------|
| 1(TC) | Test | — | Control |
| 2(CT) | Control | — | Test |

The resulting data are commonly analyzed using a linear model. The response y_{ijk} on the k th subject in period j of sequence group i , where $i = 1, 2, j = 1, 2$, and $k = 1, \dots, n_i$ is modeled as a linear function of an overall mean response μ , formulation effect τ_t and τ_c , period effects π_1 and π_2 , and sequence effects γ_1 and γ_2 . The fixed effects model can be displayed as:

| Group | Period 1 | Washout | Period 2 |
|-------|-----------------------------------|---------|-----------------------------------|
| 1(TC) | $\mu + \tau_t + \pi_1 + \gamma_1$ | — | $\mu + \tau_c + \pi_2 + \gamma_1$ |
| 2(CT) | $\mu + \tau_c + \pi_1 + \gamma_2$ | — | $\mu + \tau_t + \pi_2 + \gamma_2$ |

Let $\mu_t = \mu + \tau_t$ and $\mu_c = \mu + \tau_c$ denote the means of the observations from the test and control formulations, respectively, and let MSE denote the mean-squared error of the log data obtained from fitting the model. This is nothing other than the MSE from a crossover ANOVA model for the 2×2 design (2 periods and 2 sequences).

13 Normal Equivalence Two-Sample

In an equivalence trial, the goal is to establish $\delta_L < \mu_t - \mu_c < \delta_U$, where δ_L and δ_U are specified values used to define equivalence. In practice, δ_L and δ_U are often chosen such that $\delta_L = -\delta_U$. The two one-sided tests (TOST) procedure of Schuirmann (1987) is commonly used for this analysis, and is employed here for a crossover study.

Let $\delta = \mu_t - \mu_c$ denotes the true difference in the means. The null hypothesis $H_0: \delta \leq \delta_L$ or $\delta \geq \delta_U$ is tested against the two-sided alternative hypothesis $H_1: \delta_L < \delta < \delta_U$ at level α , using TOST. Here we perform the following two tests together:

- Test1: $H_{0L}: \delta \leq \delta_L$ against $H_{1L}: \delta > \delta_L$ at level α
- Test2: $H_{0U}: \delta \geq \delta_U$ against $H_{1U}: \delta < \delta_U$ at level α

H_0 is rejected in favor of H_1 at level α if and only if both H_{0L} and H_{0U} are rejected. Note that this is the same as rejecting H_0 in favor of H_1 at level α if the $(1 - 2\alpha)$ 100% confidence interval for δ is completely contained within the interval (δ_L, δ_U) .

East uses following test statistic to test the above two null hypotheses

$$T_L = \frac{(\bar{y}_{11} - \bar{y}_{12} - \bar{y}_{21} + \bar{y}_{22})/2 - \delta_L}{\sqrt{\frac{MSE}{2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

and

$$T_U = \frac{(\bar{y}_{11} - \bar{y}_{12} - \bar{y}_{21} + \bar{y}_{22})/2 - \delta_U}{\sqrt{\frac{MSE}{2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where, \bar{y}_{ij} is the mean of the observations from group i and period j . Both T_L and T_U are distributed as Student's t distribution with $(n_1 + n_2 - 2)$ degrees of freedom.

The power of the test (i.e. probability of declaring equivalence) depends on the true value of $\mu_t - \mu_c$. The sample size (or power) is determined at a specified value of this difference, denoted δ_1 . The choice $\delta_1 = 0$, i.e. $\mu_t = \mu_c$, is common. Note that the power and the sample size depend only on δ_L , δ_U , δ_1 , and \sqrt{MSE} .

13.3.1 Trial design

Standard 2×2 crossover designs are often recommended in regulatory guidelines to establish bioequivalence of a generic drug with off patent brand-name drug. Consider a 2×2 bioequivalence trial between a Test drug (T) and a Reference Drug (C) where equivalence needs to be established in terms of the pharmacokinetic parameter Area Under the Curve (AUC). Let μ_T and μ_c denote the average AUC for Test and

Reference drugs, respectively. Let $\delta = \mu_t - \mu_c$ be the difference. To establish average bioequivalence, the calculated 90% confidence interval of δ should fall within a pre-specified bioequivalence limit. The bioequivalence limits are set at -3 and 3. Accordingly we plan to design the study to test:

$$H_0 : \mu_t - \mu_c \leq -3 \text{ or } \mu_t - \mu_c \geq 3$$

against

$$H_1 : -3 < \mu_t - \mu_c < 3$$

From this study, we consider $\mu_c = 21.62$ ng.h/mL and $\mu_t = 23.19$ ng.h/mL under H_1 . Further, we assume that the mean squared error (MSE) from ANOVA would be 2.5. We wish to design a study that would have 90% power at $\delta_1 = 23.19 - 21.62 = 1.57$ under H_1 .

Start East afresh. Click **Continuous: Two Samples** on the **Design** tab and then click **Crossover Design: Difference of Means**.

This will launch a new window. The upper pane displays several fields with default values. Select **Equivalence** for **Design Type**, and **Individual Means** for **Input Method**. Enter 0.05 for **Type I Error**. Specify the **Mean Control** (μ_c) as 21.62 and **Mean Treatment** (μ_t) as 23.19. Select **Sqrt (MSE)** from the drop-down list and specify as 2.5. Also specify the **Lower Equiv. Limit** (δ_L) and **Upper Equiv. Limit** (δ_U) as -3 and 3, respectively. The upper pane should appear as below:

Click **Compute**. The output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed sample size (54 subjects) is highlighted in

13 Normal Equivalence Two-Sample

yellow.

| ID | Design Type | Test Type | Specified α | Power | Input Method | Sample Size | ntc/nct | Test Statistic | μ | Mean Treatment (Alt.) | δ L | δ U | Sqrt(MSE) |
|-------|-------------|-----------|--------------------|-------|------------------|-------------|---------|----------------|-------|-----------------------|------------|------------|-----------|
| Des 1 | Equivalence | 2-Sided | 0.05 | 0.901 | Individual Means | 54 | 1 | t | 21.62 | 23.19 | -3 | 3 | 2.5 |

This design has default name Des 1. Select this design and click  in the **Output Preview** toolbar. Some of the design details will be displayed in the upper pane, labeled as **Output Summary**.

| Des 1 | |
|---------------------------------------|------------------|
| Mnemonic | MN-2S-XDI |
| Test Parameters | |
| Design Type | Equivalence |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.901 |
| Model Parameters | |
| Test Statistic | t |
| Mean Control (μ c) | 21.62 |
| Mean Treatment (μ t) | 23.19 |
| Equivalence Lower Limit (δ L) | -3 |
| Equivalence Upper Limit (δ U) | 3 |
| Sq.Root (Mean Square Error) | 2.5 |
| Input Method | Individual Means |
| Allocation Ratio (ntc/nct) | 1 |
| Sample Size | |
| Maximum | 54 |

In the **Output Preview** toolbar, click  to save this design to Wbk1 in the **Library**. Double-click Des 1 in the **Library** to see the details of the designs. Close the

output window before continuing.

Design: Continuous Endpoint: Two-Sample Test - Crossover Design - Difference of Means

| Test Parameters | |
|--|------------------|
| Design ID | Des1 |
| Design Type | Equivalence |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.901 |
| Sample Size (n) | To be Computed |
| Model Parameters | |
| Test Statistic | t |
| Input Method | Individual Means |
| Mean Treatment (μ_t) | 23.19 |
| Mean Control (μ_c) | 21.62 |
| $\delta = \mu_t - \mu_c$ | |
| Under H01 | -3 |
| Under H02 | 3 |
| Under H1 | 1.57 |
| Lower Equivalence Limit (δ_l) | -3 |
| Upper Equivalence Limit (δ_u) | 3 |
| Sqrt(MSE) | 2.5 |
| Allocation Ratio (n_c/n_t) | 1 |

Sample Size Information

| | |
|-----------------------------|----|
| Sample Size (n) | 54 |
| Treatment (n _t) | 27 |
| Control (n _c) | 27 |

Critical Points

| | |
|----------------------|--------|
| Lower Critical Point | -1.675 |
| Upper Critical Point | 1.675 |

13.3.2 Simulation

Select Des 1 in the **Library**, and click  in the toolbar. Alternatively, right-click Des 1 and select **Simulate**. A new Simulation window will appear. Click on the **Response Generation Info** tab, and specify: **Mean control** = 21.62; **Mean Treatment** = 23.19; **Sqrt(MSE)** = 2.5.

Simulation Parameters
Response Generation Info
Simulation Control Info

Mean Control(μ_c):

Mean Treatment(μ_t):

Sqrt(MSE)

Leave the other default values and click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1. Select Sim 1 in the **Output Preview** and click . Now double-click on Sim 1 in

13 Normal Equivalence Two-Sample

the **Library**. The simulation output details will be displayed.

Simulation: Continuous Endpoint: Two-Sample Test - Crossover Design - Difference of Means

| Simulation Parameters | |
|---|-------------|
| Simulation ID | Sim1 |
| Trial Type | Equivalence |
| Test Type | 2-Sided |
| Sample Size (n) | 54 |
| Lower Equivalence Limit (δ_L) | -3 |
| Upper Equivalence Limit (δ_U) | 3 |
| Response Generation Parameters | |
| Mean Response under Control (μ_c) | 21.62 |
| Mean Response under Treatment (μ_t) | 23.19 |
| Sqrt(MSE) | 2.5 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Simulation Boundaries

Upper Critical Point: 1.675
Lower Critical Point: -1.675

Overall Simulation Results

| | Upper H0 | Lower H0 |
|-------------------|----------|----------|
| No. of Rejections | 8982 | 0 |
| % | 89.82 | 0 |

Starting Seed: 6709419
Total Number of Simulations: 10000
Elapsed Time: 00:00:03

Notice that the number of rejections was close to 90% of the 10000 simulated trials. The exact result of the simulations may differ slightly, depending on the seed.

The simulation we have just done was under H_1 . We wish to simulate from a point that belongs to H_0 . Right-click Sim 1 in **Library** and select **Edit Simulation**. Go to the **Response Generation Info** tab in the upper pane and specify: **Mean control** = 21.62; **Mean Treatment** = 24.62; **Sqrt. MSE** = 2.5.

Simulation Parameters
Response Generation Info
Simulation Control Info

Mean Control(μ_c):

Mean Treatment(μ_t):

Sqrt.MSE:

Click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 2. Select Sim 2 in the **Output Preview** and click . Now double-click on Sim 2 in the **Library**. The simulation output

details will be displayed.

Simulation: Continuous Endpoint: Two-Sample Test - Crossover Design - Difference of Means

| Simulation Parameters | |
|---|-------------|
| Simulation ID | Sim2 |
| Trial Type | Equivalence |
| Test Type | 2-Sided |
| Sample Size (n) | 54 |
| Lower Equivalence Limit (δ_L) | -3 |
| Upper Equivalence Limit (δ_U) | 3 |
| Response Generation Parameters | |
| Mean Response under Control (μ_c) | 21.62 |
| Mean Response under Treatment (μ_t) | 24.62 |
| Sqrt(MSE) | 2.5 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Simulation Boundaries

Upper Critical Point: 1.675
Lower Critical Point: -1.675

Overall Simulation Results

| | Upper H0 | Lower H0 |
|-------------------|----------|----------|
| No. of Rejections | 488 | 0 |
| % | 4.88 | 0 |

Starting Seed: 6829218
Total Number of Simulations: 10000
Elapsed Time: 00:00:04

Notice that the upper efficacy stopping boundary was crossed very close to 5% of the 10000 simulated trials. The exact result of the simulations may differ slightly, depending on the seed.

13.4 Ratio of Means in Crossover Designs

Often in crossover designs, an equivalence hypothesis is tested in terms of ratio of means. These types of trials are very popular in establishing bioavailability or bioequivalence between two formulations in terms of pharmacokinetic parameters (FDA guideline on BA/BE studies for orally administered drug products, 2003). In particular, the FDA considers two products bioequivalent if the 90% confidence interval of the ratio of two means lie within (0.8, 1.25). In this chapter, we show how East supports the design and simulation of such experiments with endpoint as ratio of means.

In a 2×2 crossover design each subject is randomized to one of two sequence groups. We have already discussed 2×2 crossover design in section 13.3. However, unlike section 13.3, we are interested in the ratio of means. Let μ_t and μ_c denote the means of the observations from the experimental treatment (T) and the control treatment (C), respectively. In an equivalence trial with endpoint as ratio of means, the goal is to establish $\rho_L < \rho < \rho_U$, where ρ_L and ρ_U are specified values used to define equivalence. In practice, ρ_L and ρ_U are often chosen such that $\rho_L = 1/\rho_U$

The null hypothesis $H_0: \rho \leq \rho_L$ or $\rho \geq \rho_U$ is tested against the two-sided alternative hypothesis $H_1: \rho_L < \rho < \rho_U$ at level α , using two one-sided tests. Schuirmann (1987) proposed working this problem on the natural logarithm scale. Thus, we are interested in the parameter $\delta = \ln(\rho) = \ln(\mu_t) - \ln(\mu_c)$ and the null hypothesis $H_0: \delta \leq \delta_L$ or

13 Normal Equivalence Two-Sample

$\delta \geq \delta_U$ is tested against the two-sided alternative hypothesis $H_1: \delta_L < \delta < \delta_U$ at level α , using two one-sided t-tests. Here $\delta_L = \ln(\rho_L)$ and $\delta_U = \ln(\rho_U)$.

Since we have translated the ratio hypothesis into a difference hypothesis, we can perform the test for difference as discussed in section 13.1. Note that we need the standard deviation for log transformed data. However, if we are provided with information on CV instead, the standard deviation of log transformed data can be obtained using the relation $sd = \sqrt{\ln(1 + CV^2)}$.

13.4.1 Trial design

Standard 2×2 crossover designs are often recommended in regulatory guidelines to establish bioequivalence of a generic drug with off patent brand-name drug. Consider a 2×2 bioequivalence trial between a Test drug (T) and a Reference Drug (C) where the equivalence need to be established in terms of pharmacokinetic parameter Area Under the Curve (AUC). Let μ_T and μ_C denote the average AUC for Test and Reference drugs, respectively. Let $\rho = \mu_t/\mu_c$ be the ratio of averages. To establish average bioequivalence, the calculated 90% confidence interval of ρ should fall within a pre-specified bioequivalence limit. The bioequivalence limits are set at 0.8 and 1.25. Accordingly we plan to design the study to test:

$$H_0 : \mu_t/\mu_c \leq 0.8 \text{ or } \mu_t/\mu_c \geq 1.25$$

against

$$H_1 : 0.8 < \mu_t/\mu_c < 1.25$$

From this study, we consider $\mu_c = 21.62$ ng.h/mL and $\mu_t = 23.19$ ng.h/mL under H_1 . Further, we assume that the coefficient of variation (CV), or intrasubject variability, is 17%. For a lognormal population, the mean squared error (MSE) from ANOVA of log-transformed data, and CV, are related by: $MSE = \log(1 + CV^2)$. Thus in this case MSE is 0.0285 and its square-root is 0.169. We wish to design a study that would have 90% power at $\rho_1 = 23.19/21.62 = 1.073$ under H_1 .

Start East afresh. Click **Continuous: Two Samples** on the **Design** tab and then click **Crossover Design: Ratio of Means**.

This will launch a new window. The upper pane displays several fields with default values. Select **Equivalence** for **Design Type**, and **Individual Means** for **Input Method**. Enter 0.05 for **Type I Error**. Then specify the **Mean Control** (μ_c) as 21.62 and **Mean Treatment** (μ_t) as 23.19. Specify 0.169 for **Sqrt. of MSE Log**. Also

specify the **Lower Equiv. Limit** (ρ_L) and **Upper Equiv. Limit** (ρ_U) as 0.8 and 1.25, respectively. The upper pane should appear as below:

Click **Compute**. The output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed sample size (23 subjects) is highlighted in yellow.

| ID | Design Type | Test Type | Specified α | Power | Input Method | Sample Size | ntc/nct | Test Statistic | μ_c | Mean Treatment (Alt.) | ρ_L | ρ_U | Sqrt MSE Log |
|------|-------------|-----------|--------------------|-------|------------------|-------------|---------|----------------|---------|-----------------------|----------|----------|--------------|
| Des1 | Equivalence | 2-Sided | 0.05 | 0.907 | Individual Means | 23 | 1 | t | 21.62 | 23.19 | 0.8 | 1.25 | 0.169 |

This design has default name Des 1. Select this design and click  in the **Output Preview** toolbar. Some of the design details will be displayed in the upper pane, labeled as **Output Summary**.

| Des 1 | |
|--------------------------------------|------------------|
| Mnemonic | MN-25-XRA |
| Test Parameters | |
| Design Type | Equivalence |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.907 |
| Model Parameters | |
| Mean Control (μ_c) | 21.62 |
| Mean Treatment (μ_t) | 23.19 |
| Equivalence Lower Limit (ρ_L) | 0.8 |
| Equivalence Upper Limit (ρ_U) | 1.25 |
| Sq.Root (MSE) of Log Ratios | 0.169 |
| Input Method | Individual Means |
| Allocation Ratio (ntc/nct) | 1 |
| Test Statistic | t |
| Sample Size | |
| Maximum | 23 |

13 Normal Equivalence Two-Sample

In the **Output Preview** toolbar, click  to save this design to Wbk1 in the **Library**. Double-click Des 1 in the **Library** to see the details of the designs.

Design: Continuous Endpoint: Two-Sample Test - Crossover Design - Ratio of Means

| Test Parameters | |
|--------------------|----------------|
| Design ID | Des1 |
| Design Type | Equivalence |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.907 |
| Sample Size (n) | To be Computed |

| Model Parameters | |
|--------------------------------|------------------|
| Test Statistic | t |
| Input Method | Individual Means |
| Mean Treatment (μ_t) | 23.19 |
| Mean Control (μ_c) | 21.62 |
| $\rho = \mu_t / \mu_c$ | |
| Under H01 | 0.8 |
| Under H02 | 1.25 |
| Under H1 | 1.073 |
| Sqrt(MSE) Log | 0.169 |
| Allocation Ratio (n_t/n_c) | 1 |

| Sample Size Information | |
|-------------------------|----|
| Sample Size (n) | 23 |
| Treatment (n_t) | 11 |
| Control (n_c) | 12 |

| Critical Points | |
|----------------------|--------|
| Lower Critical Point | -1.725 |
| Upper Critical Point | 1.725 |

13.4.2 Simulation

Select Des 1 in the **Library**, and click  in the toolbar. Alternatively, right-click Des 1 and select **Simulate**. A new Simulation window will appear. Click on the **Response Generation Info** tab, and specify: **Mean control** = 21.62; **Mean Treatment** = 23.19; **Sqrt. of MSE Log** = 0.169.

| Simulation Parameters | Response Generation Info | Simulation Control Info |
|----------------------------|------------------------------------|---|
| Mean Control(μ_c): | <input type="text" value="21.62"/> | Sqrt.of MSE Log: <input type="text" value="0.169"/> |
| Mean Treatment(μ_t): | <input type="text" value="23.19"/> | |

Click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1.

Select Sim 1 in the **Output Preview** and click . Now double-click on Sim 1 in

the **Library**. The simulation output details will be displayed.

Simulation: Continuous Endpoint: Two-Sample Test - Crossover Design - Ratio of Means

| Simulation Parameters | |
|---|-------------|
| Simulation ID | Sim1 |
| Trial Type | Equivalence |
| Test Type | 2-Sided |
| Sample Size (n) | 23 |
| Lower Equivalence Limit (p_L) | 0.8 |
| Upper Equivalence Limit (p_U) | 1.25 |
| Response Generation Parameters | |
| Mean Response under Control (μ_C) | 21.62 |
| Mean Response under Treatment (μ_T) | 23.19 |
| Sqrt of MSE Log | 0.169 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Simulation Boundaries

Upper Critical Point: 1.725
 Lower Critical Point: -1.725

Overall Simulation Results

| | Upper H0 | Lower H0 |
|-------------------|----------|----------|
| No. of Rejections | 9070 | 0 |
| % | 90.7 | 0 |

Starting Seed: 8793362
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:03

Notice that the number of rejections was close to 90% of the 10,000 simulated trials. The exact result of the simulations may differ slightly, depending on the seed.

14

Normal: Many Means

In this section, we will illustrate various tests available for comparing more than two continuous means in East.

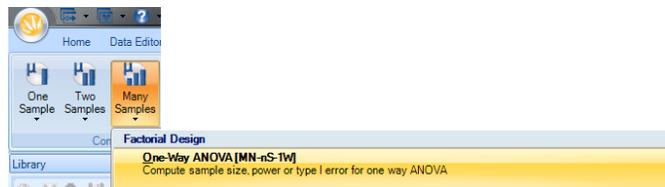
14.1 One Way ANOVA

14.1.1 One Way Contrast

In a one-way ANOVA test, we wish to test the equality of means across R independent groups. The two sample difference of means test for independent data is a one-way ANOVA test for 2 groups. The null hypothesis $H_0 : \mu_1 = \mu_2 = \dots = \mu_R$ is tested against the alternative hypothesis H_1 : for at least one pair (i, j) , $\mu_i \neq \mu_j$, where $i, j = 1, 2, \dots, R$.

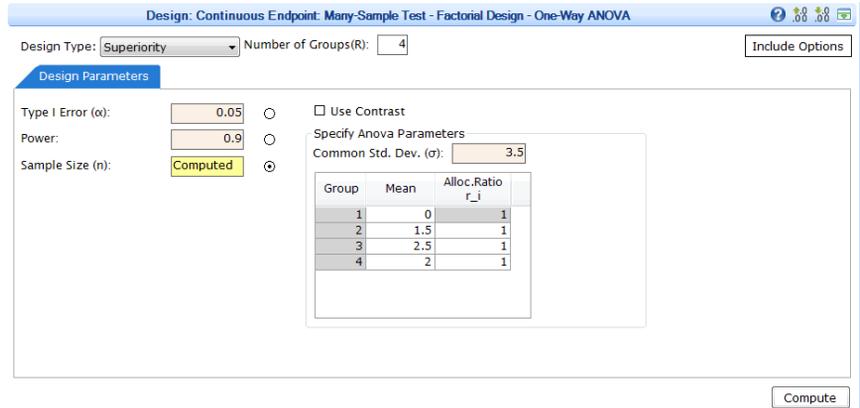
Suppose n patients have been allocated randomly to R treatments. We assume that the data of the R treatment groups comes from R normally distributed populations with the same variance σ^2 , and with population means $\mu_1, \mu_2, \dots, \mu_R$.

To design a one-way ANOVA study in East, first click **Continuous: Many Samples** on the **Design** tab, and then click **Factorial Design: One Way ANOVA**.



In the upper pane of this window is the input dialog box. Consider a clinical trial with four groups. Enter 4 in **Number of Groups(R)**. The trial is comparing three different doses of a drug against placebo in patients with Alzheimer's disease. The primary objective of the study is to evaluate the efficacy of these three doses, where efficacy is assessed by *difference from placebo* in cognitive performance measured on a 13-item cognitive subscale. On the basis of pilot data, the expected mean responses are 0, 1.5, 2.5, and 2, for Groups 1 to 4, respectively. The common standard deviation within each group is $\sigma = 3.5$. We wish to compute the required sample size to achieve 90% power with a type-1 error of 0.05. Enter these values into the dialog box as shown below.

Then, click **Compute**.



The design is shown as a row in the **Output Preview**, located in the lower pane of the window. The computed sample size (203) is highlighted in yellow.

| ▲ | ID | Design Type | Specified α | Power | Sample Size | Δ | Groups | σ | Variance of Means |
|---|------|-------------|--------------------|-------|-------------|----------|--------|----------|-------------------|
| | Des1 | Superiority | 0.05 | 0.901 | 203 | 0.071 | 4 | 3.5 | 0.875 |

Select this row, then click in the **Output Preview** toolbar to save this design to Workbook1 in the **Library**. With Des1 selected in the **Library**, click to

14 Normal: Many Means

display the following output.

Design: Continuous Endpoint: Many-Sample Test - Factorial Design - One-Way ANOVA

| Test Parameters | |
|-------------------------------|----------------|
| Design ID | Des1 |
| Design Type | Superiority |
| Specified α | 0.05 |
| Power | 0.901 |
| Sample Size (n) | To be Computed |
| Model Parameters | |
| Number of Groups (R) | 4 |
| Common Std. Dev. (σ) | 3.5 |
| Variance in Means | 0.875 |
| Effect Size (Δ) | 0.071 |

Sample Size Information

Sample Size (n) 203

Critical Points

Critical Point 2.65

Sample Size: Group by Group

| Group # | Proportion of Completers w.r.t. 1st Population | Sample Size Per Group |
|---------|--|-----------------------|
| | $r_i = n_i / n_1$ | |
| 1 | 1 | 51 |
| 2 | 1 | 51 |
| 3 | 1 | 50 |
| 4 | 1 | 51 |

ANOVA Parameters Table Values

| Group # | Mean | Proportion of Completers w.r.t. 1st Population. |
|---------|------|---|
| | | $r_i = n_i / n_1$ |
| 1 | 0 | 1 |
| 2 | 1.5 | 1 |
| 3 | 2.5 | 1 |
| 4 | 2 | 1 |

The output indicates that 51 patients per group is necessary to achieve the desired power. Close this output window before continuing.

14.1.1 One Way Contrast

A contrast of the population means is a linear combination of the μ_i 's. Let c_i denote the coefficient for population mean μ_i in the linear contrast. For a single contrast test of many means in a one-way ANOVA, the null hypothesis is $H_0 : \sum c_i \mu_i = 0$ versus a two-sided alternative $H_1 : \sum c_i \mu_i \neq 0$, or a one-sided alternative $H_1 : \sum c_i \mu_i < 0$ or $H_1 : \sum c_i \mu_i > 0$.

With Des1 selected in the **Library**, click . In the input dialog box, click the checkbox titled **Use Contrast**, and select a two-sided test. Ensure that the means for each group are the same as those from Des1 (0, 1.5, 2.5, and 2). In addition, we wish the test the following contrast: $-3, 1, 1, 1$, which compares the placebo group with the average of the three treatment groups. Finally, we may enter unequal allocation ratios such as: $1, 2, 2, 2$, which implies that twice as many patients will be assigned to each

treatment group as in the placebo group. Click **Compute**.

Use Contrast

Specify Anova Parameters

Common Std. Dev. (σ):

| Group | Mean | Contrast Coefficient | Alloc. Ratio r_i |
|-------|------|----------------------|--------------------|
| 1 | 0 | -3 | 1 |
| 2 | 1.5 | 1 | 2 |
| 3 | 2.5 | 1 | 2 |
| 4 | 2 | 1 | 2 |

The following row will be added to the **Output Preview**.

| ▲ | ID | Design Type | Specified α | Power | Sample Size | Δ | Groups | σ | Variance of Means | Test Type | Contrast | Scale |
|---|------|-------------|--------------------|-------|-------------|----------|--------|----------|-------------------|-----------|----------|-------|
| | Des1 | Superiority | 0.05 | 0.901 | 203 | 0.071 | 4 | 3.5 | 0.875 | | | |
| | Des2 | Superiority | 0.05 | 0.9 | 265 | 0.529 | 4 | 3.5 | | 2-Sided | 6 | 3.24 |

Given the above contrast and allocation ratios, this study would require a total of 265 patients to achieve 90% power.

14.2 One Way Repeated Measures (Const. Correlation) ANOVA

As with the one-way ANOVA discussed in subsection 14.1, the repeated measures ANOVA also tests for equality of means. However, in a repeated measures setting, all patients are measured under all levels of the treatment. As the sample is exposed to each condition in turn, the measurement of the dependent variable is repeated. Thus, there is some correlation between observations from the same patient, which needs to be accounted for. The constant correlation assumption means we assume that the correlation between observations from the same patient is constant for all patients. The correlation parameter (ρ) is an additional parameter that needs to be specified in the one way repeated measures study design.

Start East afresh. To design a repeated measure ANOVA study, click **Continuous: Many Samples**, and click **Factorial Design: One Way Repeated Measures (Constant Correlation) ANOVA**.

A specific type of repeated measures design is a longitudinal study in which patients are followed over a series of time points. As an illustration, we will consider a

14 Normal: Many Means

hypothetical study that investigated the effect of a dietary intervention on weight loss. The endpoint is decrease in weight (in kilograms) from baseline, measured at four time points: baseline, 4 weeks, 8 weeks, and 12 weeks. For **Number of Levels**, enter 4. We wish to compute the required sample size to achieve 90% power with a type-1 error of 0.05. The means at each of the four levels are: 0, 1.5, 2.5, 2 for Levels 1, 2, 3, and 4, respectively. Finally, enter $\sigma = 5$ and $\rho = 0.2$, and click **Compute**.

Design: Continuous Endpoint: Many-Sample Test - Factorial Design - One-Way Repeated Measures (Constant Correlation) ANOVA

Design Type: Superiority Number of Levels(M): 4

Design Parameters

Type I Error (α): 0.05

Power: 0.9

Sample Size (n): Computed

Use Contrast

Specify Anova Parameters

Between Level Correlation (ρ): 0.2

Std. Dev.at each Level (σ): 5

| Level | Mean |
|-------|------|
| 1 | 0 |
| 2 | 1.5 |
| 3 | 2.5 |
| 4 | 2 |

Compute

The design is shown as a row in the **Output Preview**, located in the lower pane of the window. The computed sample size (330) is highlighted in yellow.

| ID | Design Type | Specified α | Power | Sample Size | Levels | σ | Variance | Δ | Correlation |
|------|-------------|--------------------|-------|-------------|--------|----------|----------|----------|-------------|
| Des1 | Superiority | 0.05 | 0.901 | 330 | 4 | 5 | 0.875 | 0.044 | 0.2 |

Select this row, then click  in the **Output Preview** toolbar to save this design to Workbook1 in the **Library**. With Des1 selected in the **Library**, click  to

display the following output.

Design: Continuous Endpoint: Many-Sample Test - Factorial Design - One-Way Repeated Measures (Constant Correlation) ANOVA

| Test Parameters | |
|--------------------------------------|----------------|
| Design ID | Des 1 |
| Design Type | Superiority |
| Specified α | 0.05 |
| Power | 0.901 |
| Sample Size (n) | To be Computed |
| Model Parameters | |
| Number of Levels (M) | 4 |
| Common Std. Dev. (σ) | 5 |
| Variance in Means | 0.875 |
| Between Level Correlation (ρ) | 0.2 |
| Effect Size (Δ) | 0.044 |

Sample Size Information

Sample Size (n) 330

Critical Points

Critical Point 2.642

Sample Size: Group by Group

| Group # | Sample Size Per Group |
|---------|-----------------------|
| 1 | 83 |
| 2 | 82 |
| 3 | 83 |
| 4 | 82 |

ANOVA Parameters Table Values:

| Level | Mean |
|-------|------|
| 1 | 0 |
| 2 | 1.5 |
| 3 | 2.5 |
| 4 | 2 |

The output indicates that 83 patients per group is necessary to achieve the desired power.

14.3 Two Way ANOVA

In a two-way ANOVA, there are two factors to consider, say A and B. We can design a study to test equality of means across factor A, factor B, or the interaction between of A and B. In addition to the common standard deviation σ , you also need to specify the cell means.

For example, consider a study to determine the combined effects of sodium restriction and alcohol restriction on lowering of systolic blood pressure in hypertensive men (Parker et al., 1999). Let **Factor A** be sodium restriction and **Factor B** be alcohol restriction. There are two levels of each factor (restricted vs usual sodium intake, and restricted vs usual alcohol intake), producing four groups. Each patient is randomly assigned to one of these four groups.

Start East afresh. Click **Continuous: Many Samples**, and click **Factorial Design: Two-Way ANOVA**.

14 Normal: Many Means

Enter a type-1 error of 0.05. Then enter the following values in the input dialog box as shown below: **Number of Factor A Levels** as 2, **Number of Factor B Levels** as 2, **Common Std. Dev.** as 2, **A1/B1** as 0.5, **A1/B2** as 4.7, **A2/B1** as 0.4, and **A2/B2** as 6.9. We will first select **Power for A**, then click **Compute**.

Leaving the same input values, click **Compute** after selecting **Power for B** in the input window. Similarly, click **Compute** after selecting **Power for AB**. The **Output Preview** should now have three rows, as shown below.

| ID | Specified α | Specified Power A | Specified Power B | Specified Power AB | Attained Power A | Attained Power B | Attained Power AB | Sample Size | #Factor A Levels | #Factor B Levels | σ | Var VA | Var VB | Var VAB |
|------|-------------|-------------------|-------------------|--------------------|------------------|------------------|-------------------|-------------|------------------|------------------|---|--------|--------|---------|
| Des1 | 0.05 | 0.9 | NA | NA | 0.903 | 1 | 0.946 | 156 | 2 | 2 | 2 | 0.276 | 7.156 | 0.331 |
| Des2 | 0.05 | NA | 0.9 | NA | 0.127 | 0.981 | 0.142 | 12 | 2 | 2 | 2 | 0.276 | 7.156 | 0.331 |
| Des3 | 0.05 | NA | NA | 0.9 | 0.849 | 1 | 0.906 | 132 | 2 | 2 | 2 | 0.276 | 7.156 | 0.331 |

In order to achieve at least 90% power to detect a different across means in factor A, factor B, as well as the interaction, a sample size of 156 patients is necessary (i.e., Des1). Select Des1 in the **Output Preview**, then click  in the toolbar to save to Workbook1 in the **Library**. With Des1 selected in the **Library**, click  to

display the following output.

Design: Continuous Endpoint: Many-Sample Test - Factorial Design - Two-Way ANOVA

| Test Parameters | |
|--------------------------------|----------------|
| Design ID | Des1 |
| Specified α | 0.05 |
| Attained α : | 0.05 |
| Power for A | 0.903 |
| Power for B | 1 |
| Power for AB | 0.946 |
| Sample Size (n) | To be Computed |
| Model Parameters | |
| Number of Factor A Levels | 2 |
| Number of Factor B Levels | 2 |
| Common Std. Dev. (σ) | 2 |
| Variance in Means (V_A) | 0.276 |
| Variance in Means (V_B) | 7.156 |
| Variance in Means (V_{AB}) | 0.331 |

Sample Size Information

Sample Size (n) 156

Critical Points

Critical Point 3.904

⊖

| Sample Size Per Group | Factor B Levels | |
|-----------------------|-----------------|----|
| | B1 | B2 |
| A1 | 39 | 39 |
| A2 | 39 | 39 |

ANOVA Parameters Table Values

| Factor A Levels | Factor B Levels | | Row Mean |
|-----------------|-----------------|-----|----------|
| | B1 | B2 | |
| A1 | 0.5 | 4.7 | 2.6 |
| A2 | 0.4 | 6.9 | 3.65 |
| Column Mean | 0.45 | 5.8 | 3.125 |

The output indicates that 39 patients per group is necessary to achieve 90% power to test the main effect of A.

15 *Multiple Comparison Procedures for Continuous Data*

It is often the case that multiple objectives are to be addressed in one single trial. These objectives are formulated into a family of hypotheses. Formal statistical hypothesis tests can be performed to see if there is strong evidence to support clinical claims. Type I error is inflated when one considers the inferences together as a family. Failure to compensate for multiplicities can have adverse consequences. For example, a drug could be approved when actually it is not better than placebo. Multiple comparison (MC) procedures provides a guard against inflation of type I error due to multiple testing. Probability of making at least one type I error is known as family wise error rate (FWER). East supports several parametric and p-value based MC procedures. In this chapter we explain how to design a study using a chosen MC procedure that strongly maintains FWER.

In East, one can calculate the power from the simulated data under different MC procedures. With the information on power, one can choose the right MC procedure that provides maximum power yet strongly maintains the FWER. MC procedures included in East strongly control FWER. Strong control of FWER refers to preserving the probability of incorrectly claiming at least one null hypothesis. To contrast strong control with weak control of FWER, the latter controls the FWER under the assumption that all hypotheses are true. East supports following MC procedures based on continuous endpoint.

| Category | Procedure | Reference |
|---------------|-------------------------------|-----------------------------------|
| Parametric | Dunnett's Single Step | Dunnett CW (1955) |
| | Dunnett's Step Down | Dunnett CW and Tamhane AC (1991) |
| | Dunnett's Step Up | Dunnett CW and Tamhane AC (1992) |
| P-value Based | Bonferroni | Bonferroni CE (1935, 1936) |
| | Sidak | Sidak Z (1967) |
| | Weighted Bonferroni | Benjamini Y and Hochberg Y (1997) |
| | Holm's Step Down | Holm S (1979) |
| | Hochberg's Step Up | Hochberg Y (1988) |
| | Hommel's Step Up | Hommel G (1988) |
| | Fixed Sequence | Westfall PH, Krishen A (2001) |
| Fallback | Wiens B, Dimitrienko A (2005) | |

15.1 Parametric Procedures

15.1.1 Dunnett's single step

15.1.2 Dunnett's step-down and step-up procedures

Assume that there are k arms including the placebo arm. Let n_i be the number of subjects for i -th treatment arm ($i = 0, 2, \dots, k - 1$). Let $N = \sum_{i=0}^{k-1} n_i$ be the total sample size and the arm 0 refers to placebo. Let Y_{ij} be the response from subject j in treatment arm i and y_{ij} be the observed value of Y_{ij} ($i = 0, 2, \dots, k - 1, j = 1, 2, \dots, n_i$). Suppose that

$$Y_{ij} = \mu_i + e_{ij} \tag{15.1}$$

where $e_{ij} \sim N(0, \sigma^2)$. We are interested in the following hypotheses:

- For the right tailed test: $H_i : \mu_i - \mu_0 \leq 0$ vs $K_i : \mu_i - \mu_0 > 0$
- For the left tailed test: $H_i : \mu_i - \mu_0 \geq 0$ vs $K_i : \mu_i - \mu_0 < 0$

For the global null hypothesis at least one of the H_i is rejected in favor of K_i after controlling for FWER. Here H_i and K_i refer to null and alternative hypotheses, respectively, for comparison of i -th arm with the placebo arm.

East supports three parametric MC procedures - single step Dunnett test (Dunnett, 1955), step-down Dunnett test and step-up Dunnett test. These procedures make two parametric assumptions - normality and homoscedasticity. Let \bar{y}_i be the sample mean for treatment arm i and s^2 be the pooled sample variance for all arms. The test statistic for comparing treatment effect of arm i with placebo can be defined as

$$T_i = \frac{\bar{y}_i - \bar{y}_0}{s \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}} \tag{15.2}$$

Let t_i be the observed value of T_i and these observed values for $K - 1$ treatment arms can be ordered as $t_{(1)} \geq t_{(2)} \geq \dots \geq t_{(k-1)}$.

Detailed formula to obtain critical boundaries for single step Dunnett and step-down Dunnett tests are discussed in Appendix H.

In single step Dunnett test, the critical boundary remains same for all the $k - 1$ individual tests. Let c_α be the critical boundary that maintains FWER of α and \tilde{p}_i be the adjusted p -value associated with comparison of i -th arm and placebo arm. Then for a right tailed test, H_i is rejected if $t_i > c_\alpha$ and for a left tailed test H_i is rejected if $t_i < c_\alpha$.

Unlike in single step Dunnett test, the critical boundary does not remain same for all the $k - 1$ individual tests in step-down Dunnett test. Let c_i be the critical boundary and \tilde{p}_i be the adjusted p -value associated with comparison of i -th arm and placebo arm. For a right tailed test $H_{(i)}$ is rejected if $t_{(i)} > c_i$ and $H_{(1)}, \dots, H_{(c-i)}$ have been

15 Multiple Comparison Procedures for Continuous Data

already rejected. For a left tailed test $H_{(i)}$ is rejected if $t_{(i)} < c_{k-i}$ and $H_{(i-1)}, \dots, H_{(k-1)}$ have been already rejected.

Unlike step-down test, step-up Dunnett procedure starts with the least significant test statistic i.e., $t_{(k-1)}$. Let c_i be the critical boundary and \tilde{p}_i be the adjusted p-value associated with comparison of i -th arm and placebo arm. The i -th test statistic in order i.e., $t_{(i)}$ will be tested if and only if none of $H_{(i+1)}, \dots, H_{(k-1)}$ are rejected. If $H_{(i)}$ is rejected then stop and reject all of $H_{(i)}, \dots, H_{(1)}$. For a right tailed test, $H_{(i)}$ is rejected if $t_{(i)} > c_{(i)}$ and for a left tailed test $H_{(i)}$ is rejected if $t_{(i)} < c_{(i)}$.

For both single step Dunnett and step-down Dunnett tests, the global null hypothesis is rejected in favor of at least one right tailed alternative if $H_{(1)}$ is rejected and in favor of at least one left tailed alternative if $H_{(k-1)}$ is rejected .

Single step Dunnett test and step-down Dunnett test can be seen as the parametric version of Bonferroni procedure and Holm procedure, respectively. Parametric tests are uniformly more powerful than the corresponding p-value based tests when the parametric assumption holds or at least approximately holds, especially when there are a large number of hypotheses. Parametric procedures may not control FWER if the standard deviations are different.

15.1.1 Dunnett's single step

Dunnett's Single Step procedure is described below with an example.

Example: Alzheimer's Disease Clinical Trial

In this section, we will use an example to illustrate how to design a study using the MCP module in East. This is a randomized, double-blind, placebo-controlled, parallel study to assess three different doses (0.3 mg, 1 mg and 2 mg) of a drug against placebo in patients with mild to moderate probable Alzheimer's disease. The primary objective of this study is to evaluate the safety and efficacy of the three doses. The drugs are administered daily for 24 weeks to subjects with Alzheimer's disease who are either receiving concomitant treatment or not receiving any co-medication. The efficacy is assessed by cognitive performance based on the Alzheimer's disease assessment scale-13-item cognitive sub-scale. From previous studies, it is estimated that the common standard deviation of the efficacy measure is 5. It is expected that the dose-response relationship follows straight line within the dose range we are interested.

We would like to calculate the power for a total sample size of 200. This will be a balanced study with a one-sided 0.025 significance level to detect at least one dose

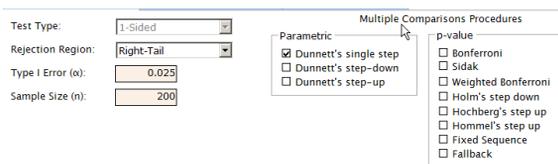
with significant difference from placebo. We will show how to simulate the power of such a study using the multiple comparison procedures listed above.

Designing the study

First, click  (**Continuous: Many Samples**) on the **Design** tab and then click **Multi-Arm Design: Pairwise Comparisons to Control - Difference of Means**. This will launch a new window.

There is a box at the top with the label **Number of Arms**. For our example, we have 3 treatment groups plus a placebo. So enter 4 for **Number of Arms**. Under the **Design Parameters** tab, there are several fields which we will fill in. First, there is a box with the label **Side**. Here you need to specify whether you want a one-sided or two-sided test. Currently, only one-sided tests are available. Under it you will see the box with label **Sample Size (n)**. For now skip this box and move to the next dropdown box with the label **Rejection Region**. If left tail is selected, the critical value for the test is located in the left tail of the distribution of the test statistic. Likewise, if right tail is selected the critical value for the test is located in the right tail of the distribution of the test statistic. For our example, we will select **Right Tail**. Under that, there is a box with the label **Type - 1 Error (α)**. This is where you need to specify the FWER. For our example, enter 0.025. Now go to the box with the label **Total Sample Size**. Here we input the total number of subjects, including those in the placebo arm. For this example, enter 200.

To the right, there will be a heading with the title **Multiple Comparison Procedures**. In the parametric grouping, check the box next to **Dunnett's single step**, as this is the multiple comparison procedure we are illustrating in this subsection. After entering these parameters your screen should now look like this:



| | |
|----------------------------|------------|
| Test Type: | 1-Sided |
| Rejection Region: | Right-Tail |
| Type I Error (α): | 0.025 |
| Sample Size (n): | 200 |

| Multiple Comparison Procedures | |
|---|--|
| Parametric | p-value |
| <input checked="" type="checkbox"/> Dunnett's single step | <input type="checkbox"/> Bonferroni |
| <input type="checkbox"/> Dunnett's step-down | <input type="checkbox"/> Sidak |
| <input type="checkbox"/> Dunnett's step-up | <input type="checkbox"/> Weighted Bonferroni |
| | <input type="checkbox"/> Holm's step down |
| | <input type="checkbox"/> Hochberg's step up |
| | <input type="checkbox"/> Hommel's step up |
| | <input type="checkbox"/> Fixed Sequence |
| | <input type="checkbox"/> Fallback |

Now click on **Response Generation Info** tab. You will see a table titled **Table of Proportions**. In this table we can specify the labels for treatment arms. Also you have to specify the dose level if you want to generate means through dose-response curve. Since we are comparing placebo and 3 dose groups, enter **Placebo**, **Dose1**, **Dose2** and **Dose3** in the 4 cells in first column labeled as **Arm**.

15 Multiple Comparison Procedures for Continuous Data

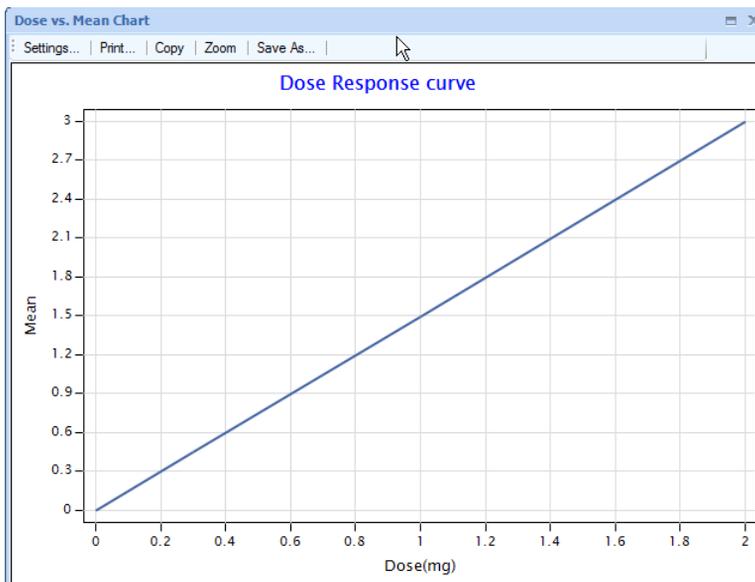
The table contains the default mean and standard deviation for each arm which we will change later. There are two check boxes in this tab above the table. The first is labeled **Generate Means through DR Curve**. There are two ways to specify the mean response for each arm: 1) generate means for each arm through a dose-response curve or 2) Specify the mean directly in the **Table of Proportions**. To specify the mean directly just enter the mean value for each arm in the table in **Mean** column. However, in this example, we will generate means through dose response curve. In order to do this, check **Generate Means through DR Curve** box. Once you check this box you will notice two things. First, an additional column with label **Dose** will appear in the table. Here you need to enter the dose levels for each arm. For this example, enter 0, 0.3, 1 and 2 for Placebo, Dose1, Dose2 and Dose3 arms, respectively. Secondly, you will notice an additional section will appear to the right which provides the option to generate the mean response from four families of parametric curves which are Four Parameter Logistic, Emax, Linear and Quadratic. The technical details about each curve can be found in the Appendix H.

Here you need to choose the appropriate parametric curve from the drop-down list under **Dose Response Curve** and then you have to specify the parameters associated with these curves. For the Alzheimer's disease example, suppose the dose response follows a linear curve with intercept 0 and slope 1.5. To do this, we would need to select "Linear" from the dropdown list. To right of this dropdown box, specify the parameter values of the selected curve family by inputting 0 for **Intercept(E0)** and 1.5 for **Slope(δ)**. After specifying this, the mean values in the table will be changed accordingly. Here we are generating the means using the following linear dose-response curve:

$$E(Y|Dose) = E0 + \delta \times Dose \quad (15.3)$$

For placebo, the mean can be obtained by specifying *Dose* as 0 in the above equation. This gives the mean for placebo arm as 0. For arm Dose1, mean would be $0 + 1.5 \times 0.3$ or 0.45. Similarly the means for the arm Dose2 and Dose3 will be obtained as 1.5 and 3. You can verify that the values in **Mean** column is changed to 0, 0.45, 1.5 and 3 for the four arms, respectively.

Now click **Plot DR Curve** to see the plot of means against the dose levels.



You will see the linear dose response curve that intersects the Y-axis at 0. Now close this window. The dose response curve generates means, but still we have to specify the standard deviation. Standard deviation for each arm could be either equal or different. To specify the common standard deviation, check the box with label **Common Standard Deviation** and specify the common standard deviation in the field next to it. When standard deviations for different arms are not all equal, the standard deviations need to be directly specified in the table in column labeled with **Std. Dev.**. In this example, we are considering a common standard deviation of 5. So check the box for **Common Standard Deviation** and specify 5 in the field next to it. Now the column **Std.Dev.** will be updated with 5 for all the four arms. As we have finished specifying all the fields in the **Response Generation Info** tab, this should appear as below.

Generate Means Through DR Curve

Common Standard Deviation

Table of Means

| Arm | Dose | Mean | Std.Dev. |
|---------|------|------|----------|
| Control | 0 | 0 | 5 |
| 1 | 0.3 | 0.45 | 5 |
| 2 | 1 | 1.5 | 5 |
| 3 | 2 | 3 | 5 |

Dose Response Curve:

Parameters

Intercept(E_0) Slope(β)

[Plot DR Curve](#)

15 Multiple Comparison Procedures for Continuous Data

Click on the **Include Options** button located in the right-upper corner in the Simulation window and check **Randomized Info**. This will add an additional tab - **Randomization Info**. Now click on the **Randomization Info** tab. Second column of the **Table of Allocation** table displays the allocation ratio of each treatment arm to that of control arm. The cell for control arm is always one and is not editable. Only those cells for treatment arms other than control need to be filled in. The default value for each treatment arm is one which represents a balanced design. For the Alzheimer's disease example, we consider a balanced design and leave the default values for the allocation ratios unchanged. Your screen should now look like this:

Randomization Method:

Table of Allocation

| Arm | $r_i = n_i/n_1$ |
|---------|-----------------|
| Control | 1.000 |
| 1 | 1.000 |
| 2 | 1.000 |
| 3 | 1.000 |

The last tab is **Simulation Control Info**. Specify 10000 as **Number of Simulations** and 1000 as **Refresh Frequency** in this tab. The box labeled **Random Number Generator** is where you can set the seed for the random number generator. You can either use the clock as the seed or choose a fixed seed (in order to replicate past simulations). The default is the **clock** and we will use that. The box on the right hand side is labeled **Output Options**. This is where you can choose to save summary statistics for each simulation run and/or to save subject level data for a specific number of simulation runs. To save the output for each simulation, check the box with label **Save summary statistics for every simulation run**. Now click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. Pairwise Comps. | Sample Size | Rejection Region | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|-------|-------------------------------|-----------|--------------------|--------------|---------------------|-------------|------------------|--------------|-------------------|-------------------|---------------------|
| Sim 1 | Dunnett's single step | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.75 | 0.019 | 0.75 | 200 |

Note that a simulation node Sim 1 is created in the library. Also note that another node is appended to the simulation node with label **SummaryStat** which contains detailed simulation summary statistics for each simulation run. Select Sim 1 in the **Output Preview** and click  icon to save the simulation in the library. Now double-click on

Sim 1 in the **Library**. The simulation output details will be displayed in the right pane.

Simulation: Continuous Endpoint: Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons Control - Differences of Means

Hypothesis:
 $H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

| Simulation Parameters | |
|-----------------------------------|-----------------------|
| Simulation ID | Sim1 |
| Test Type | 1-Sided |
| Rejection Region | Right-Tail |
| Nominal Type-1 Error (α) | 0.025 |
| Sample Size (n) | 200 |
| Multiple Comparison Procedure | Dunnett's single step |
| Total Number of Arms (k) | 4 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Overall Powers

| | |
|--|-------|
| Global (Reject any H_i) | 0.75 |
| Conjunctive (Reject all H_i where $\mu_i > \mu_0$) | 0.019 |
| Disjunctive (Reject at least one H_i where $\mu_i > \mu_0$) | 0.75 |
| FWER (Reject any H_i where $\mu_i \leq \mu_0$) | 0 |

Marginal Powers

| | |
|-------|-------|
| Arm 1 | 0.03 |
| Arm 2 | 0.197 |
| Arm 3 | 0.738 |

Treatment Parameters

| Arm | Dose | Mean | Standard Deviation | Sample Size Per Group | Allocation Ratio |
|---------|------|------|--------------------|-----------------------|------------------|
| Control | 0 | 0 | 5 | 50 | 1 |
| 1 | 0.3 | 0.45 | 5 | 50 | 1 |
| 2 | 1 | 1.5 | 5 | 50 | 1 |
| 3 | 2 | 3 | 5 | 50 | 1 |

Simulation Seed and Elapsed Time

Starting Seed: 73533547
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:05

Summary
 Using Dunnett's single step test, this study has 0.75 global power to detect at least one treatment arm which is different from control arm given a total sample size 200. This test has conjunctive power 0.019 to detect all treatment arms which are truly different from control arm. This test provides disjunctive power 0.75 to detect at least one treatment arm which is truly different from control arm.

The first section in the output is the **Hypothesis** section. In our situation, we are testing 3 hypotheses. We are comparing the mean score on the Alzheimer's disease assessment scale (13-item cognitive sub-scale) for each dose with that of placebo. That is, we are testing the 3 hypotheses:

$$H_1 : \mu_1 = \mu_0 \quad \text{vs} \quad K_1 : \mu_1 > \mu_0$$

$$H_2 : \mu_2 = \mu_0 \quad \text{vs} \quad K_2 : \mu_2 > \mu_0$$

$$H_3 : \mu_3 = \mu_0 \quad \text{vs} \quad K_3 : \mu_3 > \mu_0$$

Here, μ_P , μ_1 , μ_2 and μ_3 represent the population mean score on the Alzheimer's disease assessment scale for the placebo, 0.3 mg, 1 mg and 2 mg dose groups, respectively. Also, H_i and K_i are the null and alternative hypotheses, respectively, for the i -th test.

The **Input Parameters** section provides the design parameters that we specified earlier. The next section **Overall Power** gives us estimated power based on the

15 Multiple Comparison Procedures for Continuous Data

simulation. The second line gives us the global power, which is about 75%. Global power indicates the power to reject global null $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_0$. Thus, the global power indicates that 75% of times the global null will be rejected. In other words, at least one of the H_1 , H_2 and H_3 is rejected in about 75% of the occasion. Global power is useful to show the existence of dose-response relationship and dose-response may be claimed if any of the doses in the study is significantly different from placebo.

The next line displays the conjunctive power. Conjunctive power indicates the proportion of cases in the simulation where all the H_i 's, which are truly false, were rejected. In this example, all the H_i 's are false. Therefore, for this example, conjunctive power is the proportion of cases where all of the H_1 , H_2 and H_3 were rejected. For this simulation conjunctive power is only about 2.0% which means that only in 2.0% of time, all of the H_1 , H_2 and H_3 were rejected.

Disjunctive power indicates the proportion of rejecting at least one of those H_i 's where H_i is truly false. The main distinction between global and distinctive power is that the former finds any rejection whereas the latter look for rejection only among those H_i 's which are false. Since here all of the H_1 , H_2 and H_3 are false, therefore, global and disjunctive power ought to be the same.

The next section gives us the marginal power for each hypothesis. Marginal power finds the proportion of times when a particular hypothesis is rejected after applying multiplicity adjustment. Based on simulation results, H_1 is rejected about 3% of times, H_2 is rejected about 20% of times and H_3 is rejected a little more than 70% of times.

Recall that we have asked East to save the simulation results for each simulation run—. Open this file by clicking on **SummaryStat** in the library and you will see that it contains 10,000 rows - each rows represents results for a single simulation. Find the 3 columns with labels **Rej_Flag_1**, **Rej_Flag_2** and **Rej_Flag_3**, respectively. These columns represents the rejection status for H_1 , H_2 and H_3 , respectively. A value of 1 is indicator of rejection on that particular simulation, otherwise the null is not rejected. Now the proportion of 1's in **Rej_Flag_1** indicates the marginal power to reject H_1 . Similarly we can find out the marginal power for H_2 and H_3 from **Rej_Flag_2** and **Rej_Flag_3**, respectively. To obtain the global and disjunctive power, count the total number of cases where at least one of the H_1 , H_2 and H_3 have been rejected and then divide by the total number of simulations of 10,000. Similarly, to obtain the conjunctive power count the total number of cases where all of the H_1 , H_2 and H_3 have been rejected and then divide by the total number of simulations of 10,000.

Next we will consider an example to show how global and disjunctive power are different from each other. Select Sim 1 in **Library** and click . Now go to the **Response Generation Info** tab and uncheck the **Generate Means Through DR Curve** box. The table will now have only three columns. Specify **Dose1**, **Dose2** and **Dose3** in the 4 cells in first column labeled as **Arm** and enter 0, 0, 1 and 1.2 in the 4 cells in second column labeled as **Mean**.

Generate Means Through DR Curve

Common Standard Deviation

Table of Proportions

| Arm | Mean | Std.Dev. |
|---------|------|----------|
| Placebo | 0 | 5 |
| Dose1 | 0 | 5 |
| Dose2 | 1 | 5 |
| Dose3 | 1.2 | 5 |

Here we are generating response for placebo from distribution $N(0, 5^2)$, for Dose1 from distribution $N(0, 5^2)$, for Dose2 from distribution $N(1, 5^2)$ and for Dose3 from distribution $N(1.2, 5^2)$. Now click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 2.

| | ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. Pairwise Comps. | Sample Size | Rejection Region | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|--|------|-------------------------------|-----------|--------------------|--------------|---------------------|-------------|------------------|--------------|-------------------|-------------------|---------------------|
| | Sim1 | Dunnett's single step | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.75 | 0.019 | 0.75 | 200 |
| | Sim2 | Dunnett's single step | 1-Sided | 0.025 | 0.009 | 3 | 200 | Right-Tail | 0.179 | 0.029 | 0.176 | 200 |

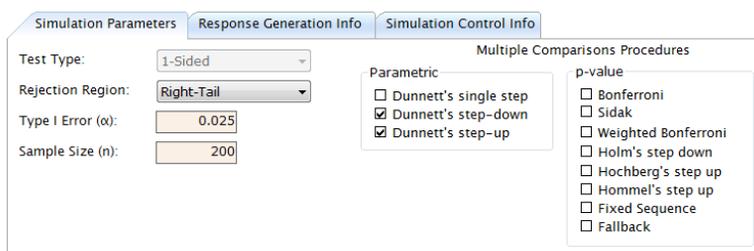
For Sim 2, the global power and disjunctive power are 17.9% and 17.6%, respectively. To understand why, we need to open the saved simulation data for Sim 2. The total number of cases where at least one of H_1 , H_2 and H_3 is rejected is 1790 and dividing this by total number of simulation 10,000 gives the global power of 17.9%. Again, the total number of cases where at least one of H_2 and H_3 are rejected is 1760 and dividing this by total number of simulation 10,000 gives the disjunctive power of 17.6%. The exact result of the simulations may differ slightly, depending on the seed.

15 Multiple Comparison Procedures for Continuous Data

15.1.2 Dunnett's step-down and step-up procedures

Dunnett's Step-Down procedure is described below using the same Alzheimer's Disease example from the previous section 15.1.1 on Dunnett's Single Step.

Since the other design specification remains same except that we are using Dunnett's step-down in place of single step Dunnett's test, we can design simulation in this section with only little effort. Select Sim 1 in **Library** and click . Now go to the **Design Parameters** tab. There in the **Multiple Comparison Procedures** box, uncheck the **Dunnett's single step** box and check the **Dunnett's step-down** and **Dunnett's step-up** box.



The screenshot shows a software interface with three tabs: "Simulation Parameters", "Response Generation Info", and "Simulation Control Info". The "Simulation Parameters" tab is active, showing "Test Type" as "1-Sided", "Rejection Region" as "Right-Tail", "Type I Error (α)" as "0.025", and "Sample Size (n)" as "200". The "Simulation Control Info" tab is also visible, showing the "Multiple Comparisons Procedures" section. Under "Parametric", "Dunnett's single step" is unchecked, while "Dunnett's step-down" and "Dunnett's step-up" are checked. Under "p-value", "Bonferroni", "Sidak", "Weighted Bonferroni", "Holm's step down", "Hochberg's step up", "Hommel's step up", "Fixed Sequence", and "Fallback" are all unchecked.

Now click **Simulate** to obtain power. Once the simulation run has completed, East will add two additional rows to the **Output Preview** labeled as Sim 3 and Sim 4.

Dunnett step-down procedure and step-down have global and disjunctive power of close to 75% and conjunctive power of close to 4%. To see the marginal power for each test, select Sim 3 and Sim 4 in the **Output Preview** and click  icon. Now, double-click on Sim 3 in the **Library**. The simulation output for Dunnett step-down

procedure details will be displayed in the right pane.

Simulation: Continuous Endpoint: Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Differences of Means

Hypothesis:
 $H_1: \mu_1 - \mu_0 \leq 0$ Vs. $K_1: \mu_1 - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

| Simulation Parameters | |
|-----------------------------------|----------------------|
| Simulation ID | Sim3 |
| Test Type | 1-Sided |
| Rejection Region | Right-Tail |
| Nominal Type-1 Error (α) | 0.025 |
| Sample Size (n) | 200 |
| Multiple Comparison Procedure | Dunnnett's step-down |
| Total Number of Arms (k) | 4 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Overall Powers

| | |
|--|-------|
| Global (Reject any H_i) | 0.746 |
| Conjunctive (Reject all H_i where $\mu_i > \mu_0$) | 0.042 |
| Disjunctive (Reject at least one H_i where $\mu_i > \mu_0$) | 0.746 |
| FWER (Reject any H_i where $\mu_i \leq \mu_0$) | 0 |

Marginal Powers

| | |
|-------|-------|
| Arm 1 | 0.052 |
| Arm 2 | 0.23 |
| Arm 3 | 0.738 |

Treatment Parameters

| Arm | Dose | Mean | Standard Deviation | Sample Size Per Group | Allocation Ratio |
|---------|------|------|--------------------|-----------------------|------------------|
| Control | 0 | 0 | 5 | 50 | 1 |
| 1 | 0.3 | 0.45 | 5 | 50 | 1 |
| 2 | 1 | 1.5 | 5 | 50 | 1 |
| 3 | 2 | 3 | 5 | 50 | 1 |

Simulation Seed and Elapsed Time

| | |
|-----------------------------|----------|
| Starting Seed | 85083994 |
| Total Number of Simulations | 10000 |
| Elapsed Time | 00:00:05 |

Summary
 Using Dunnnett's step-down test, this study has 0.746 global power to detect at least one treatment arm which is different from control arm given a total sample size 200. This test has conjunctive power 0.042 to detect all treatment arms which are truly different from control arm. This test provides disjunctive power 0.746 to detect at least one treatment arm which is truly different from control arm.

The marginal power for comparison of Dose1, Dose2 and Dose3 using Dunnnett step-down procedure are close to 5%, 23% and 74%, respectively. Similarly one can find the marginal power for individual tests in Dunnnett step-up procedure.

15.2 p-value based Procedures

15.2.1 Single step MC procedures

15.2.2 Data-driven step-down MC procedure

15.2.3 Data-driven step-up MC procedures

15.2.4 Fixed-sequence stepwise MC procedures

p-value based procedures strongly control the FWER regardless of the joint distribution of the raw p-values as long as the individual raw p-values are legitimate p-values. Assume that there are k arms including the placebo arm. Let n_i be the number of subjects for i -th treatment arm ($i = 0, 2, \dots, k-1$). Let $N = \sum_{i=0}^{k-1} n_i$ be the total sample size and the arm 0 refers to placebo. Let Y_{ij} be the response from subject j in treatment arm i and y_{ij} be the observed value of Y_{ij} ($i = 0, 2, \dots, k-1, j = 1, 2, \dots, n_i$). Suppose that

$$Y_{ij} = \mu_i + e_{ij} \tag{15.4}$$

where $e_{ij} \sim N(0, \sigma_i^2)$. We are interested in the following hypotheses:

- For the right tailed test: $H_i: \mu_i - \mu_0 \leq 0$ vs $K_i: \mu_i - \mu_0 > 0$
- For the left tailed test: $H_i: \mu_i - \mu_0 \geq 0$ vs $K_i: \mu_i - \mu_0 < 0$

15 Multiple Comparison Procedures for Continuous Data

For the global null hypothesis at least one of the H_i is rejected in favor of K_i after controlling for FWER. Here H_i and K_i refer to null and alternative hypotheses, respectively, for comparison of i -th arm with the placebo arm.

Let \bar{y}_i be the sample mean for treatment arm i , s_i^2 be the sample variance from i -th arm and s^2 be the pooled sample variance for all arms. For the unequal variance case, the test statistic for comparing treatment effect of arm i with placebo can be defined as

$$T_i = \frac{\bar{y}_i - \bar{y}_0}{\sqrt{\frac{1}{n_i} s_i^2 + \frac{1}{n_0} s_0^2}} \quad (15.5)$$

For the equal variance case, one need to replace s_i^2 and s_0^2 by the pooled sample variance s^2 . For both the case, T_i is distributed as Student's t distribution. However, the degrees of freedom varies for equal variance and unequal variance case. For equal variance case the degrees of freedom would be $N - k$. For the unequal variance case, the degrees of freedom is subject to Satterthwaite correction.

Let t_i be the observed value of T_i and these observed values for $K - 1$ treatment arms can be ordered as $t_{(1)} \geq t_{(2)} \geq \dots \geq t_{(k-1)}$. For the right tailed test the marginal p-value for comparing the i -th arm with placebo is calculated as $p_i = P(T > t_i)$ and for left tailed test $p_i = P(T < t_i)$, where T is distributed as Student's t distribution. Let $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k-1)}$ be the ordered p-values.

15.2.1 Single step MC procedures

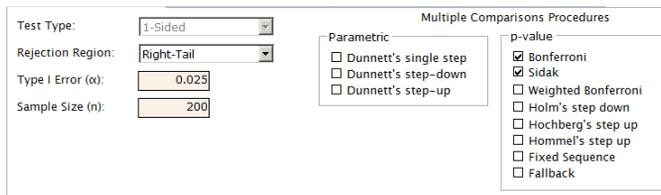
East supports three p-value based single step MC procedures - Bonferroni procedure, Sidak procedure and weighted Bonferroni procedure. For the Bonferroni procedure, H_i is rejected if $p_i < \frac{\alpha}{k-1}$ and the adjusted p-value is given as $\min(1, (k-1)p_i)$. For the Sidak procedure, H_i is rejected if $p_i < 1 - (1 - \alpha)^{\frac{1}{k-1}}$ and the adjusted p-value is given as $1 - (1 - p_i)^{k-1}$. For the weighted Bonferroni procedure, H_i is rejected if $p_i < w_i \alpha$ and the adjusted p-value is given as $\min(1, \frac{p_i}{w_i})$. Here w_i denotes the proportion of α allocated to the H_i such that $\sum_{i=1}^{k-1} w_i = 1$. Note that, if $w_i = \frac{1}{k-1}$, then the Bonferroni procedure is reduced to the regular Bonferroni procedure.

Bonferroni and Sidak procedures

Bonferroni and Sidak procedures are described below using the same Alzheimer's Disease example from the section 15.1.1 on Dunnett's Single Step.

Since the other design specification remains same except that we are using Bonferroni and Sidak in place of single step Dunnett's test, we can design simulation in this

section with only little effort. Select Sim 1 in **Library** and click . Now go to the **Design Parameters** tab. In the **Multiple Comparison Procedures** box, uncheck the **Dunnett's single step** box and check the **Bonferroni** and **Sidak** boxes.



Now click **Simulate** to obtain power. Once the simulation run has completed, East will add two additional rows to the **Output Preview**. Bonferroni and Sidak procedures have disjunctive and global powers of close to 73% and conjunctive power of about 1.8%. Now select Sim 5 and Sim 6 in the **Output Preview** using the Ctrl key and click  icon. This will save Sim 5 and Sim 6 in the Wbk1 in **Library**.

Weighted Bonferroni procedure

As before we will use the same Alzheimer's Disease example to illustrate weighted Bonferroni procedure. Select Sim 1 in **Library** and click . Now go to the **Design Parameters** tab. There in the **Multiple Comparison Procedures** box, uncheck the **Dunnett's single step** box and check the **Weighted Bonferroni** box.

Table of Means

| Arm | Dose | Mean | Std.Dev. | Proportion of Alpha |
|---------|------|------|----------|---------------------|
| Control | 0 | 0 | 5 | |
| 1 | 0.3 | 0.45 | 5 | 0.333 |
| 2 | 1 | 1.5 | 5 | 0.333 |
| 3 | 2 | 3 | 5 | 0.333 |

Next click on **Response Generation Info** tab and look at the **Table of Proportions**. You will see an additional column with label **Proportion of Alpha** is added. Here you have to specify the proportion of total alpha you want to spend in each test. Ideally, the values in this column should add up to 1; if not, then East will normalize it to add them up to 1. By default, East distributes the total alpha equally among all tests. Here we have 3 tests in total, therefore each of the tests have proportion of alpha as 1/3 or 0.333. You can specify other proportions as well. For this example, keep the equal

15 Multiple Comparison Procedures for Continuous Data

proportion of alpha for each test. Now click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 7. The weighted Bonferroni MC procedure has global and disjunctive power of 73.7% and conjunctive power of 1.6%. Note that, the powers in the weighted Bonferroni procedure is quite close to the Bonferroni procedure. This is because the weighted Bonferroni procedure with equal proportion is equivalent to the simple Bonferroni procedure. The exact result of the simulations may differ slightly, depending on the seed. Now select Sim 7 in the **Output Preview** and click  icon. This will save Sim 7 in Wbk1 in **Library**.

15.2.2 Data-driven step-down MC procedure

In the single step MC procedures, the decision to reject any hypothesis does not depend on the decision to reject other hypotheses. On the other hand, in the stepwise procedures decision of one hypothesis test can influence the decisions on the other tests of hypotheses. There are two types of stepwise procedures. One type of procedures proceed in data-driven order. The other type proceeds in a fixed order set a priori. Stepwise tests in a data-driven order can proceed in step-down or step-up manner. East supports Holm step-down MC procedure which start with the most significant comparison and continue as long as tests are significant until the test for certain hypothesis fails. The testing procedure stops at the first time a non-significant comparison occurs and all remaining hypotheses will be retained. In i -th step, $H_{(k-i)}$ is rejected if $p_{(k-i)} \leq \frac{\alpha}{i}$ and go to the next step.

Holm's step-down

As before we will use the same Alzheimer's Disease example to illustrate Holm's step-down procedure. Select Sim 1 in **Library** and click . Now go to the **Design Parameters** tab. In the **Multiple Comparison Procedures** box, uncheck the **Dunnett's single step** box and check the **Holm's Step-down** box.

| | | | |
|-------------------|------------|--|--|
| Test Type: | 1-Sided | Multiple Comparisons Procedures | |
| Rejection Region: | Right-Tail | Parametric | p-value |
| Type I Error (α): | 0.025 | <input type="checkbox"/> Dunnett's single step | <input type="checkbox"/> Bonferroni |
| Sample Size (n): | 200 | <input type="checkbox"/> Dunnett's step-down | <input type="checkbox"/> Sidak |
| | | <input type="checkbox"/> Dunnett's step-up | <input type="checkbox"/> Weighted Bonferroni |
| | | | <input checked="" type="checkbox"/> Holm's step down |
| | | | <input type="checkbox"/> Hochberg's step up |
| | | | <input type="checkbox"/> Hommel's step up |
| | | | <input type="checkbox"/> Fixed Sequence |
| | | | <input type="checkbox"/> Fallback |

Now click **Simulate** to obtain power. Once the simulation run has completed, East will

add an additional row to the **Output Preview** labeled as Sim 8.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. Pairwise Comps. | Sample Size | Rejection Region | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|------|-------------------------------|-----------|--------------------|--------------|---------------------|-------------|------------------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Dunnett's single step | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.75 | 0.019 | 0.75 | 200 |
| Sim3 | Dunnett's step-down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.746 | 0.042 | 0.746 | 200 |
| Sim4 | Dunnett's step-up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.741 | 0.045 | 0.741 | 200 |
| Sim5 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.729 | 0.017 | 0.729 | 200 |
| Sim6 | Sidak | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.732 | 0.018 | 0.732 | 200 |
| Sim7 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.016 | 0.737 | 200 |
| Sim8 | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.74 | 0.045 | 0.74 | 200 |

Holm's step-down procedure has global and disjunctive power of 74% and conjunctive power of 4.5%. The exact result of the simulations may differ slightly, depending on the seed. Now select Sim 8 in the **Output Preview** and click  icon. This will save Sim 8 in Wbk1 in **Library**.

Simulation: Continuous Endpoint: Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Differences of Means

Hypothesis:

$$H_1: \mu_i - \mu_0 \leq 0 \text{ Vs. } K_1: \mu_i - \mu_0 > 0 \text{ for } i = 1, 2, \dots, k-1 \text{ where } k \text{ is the total number of arms.}$$

| Simulation Parameters | |
|-----------------------------------|------------------|
| Simulation ID | Sim8 |
| Test Type | 1-Sided |
| Rejection Region | Right-Tail |
| Nominal Type-1 Error (α) | 0.025 |
| Sample Size (n) | 200 |
| Multiple Comparison Procedure | Holm's step down |
| Total Number of Arms (k) | 4 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Overall Powers

| | |
|---|-------|
| Global (Reject any H _i) | 0.74 |
| Conjunctive (Reject all H _i where $\mu_i > \mu_0$) | 0.045 |
| Disjunctive (Reject at least one H _i where $\mu_i > \mu_0$) | 0.74 |
| FWER (Reject any H _i where $\mu_i \leq \mu_0$) | 0 |

Marginal Powers

| | |
|-------|-------|
| Arm 1 | 0.053 |
| Arm 2 | 0.23 |
| Arm 3 | 0.73 |

Treatment Parameters

| Arm | Dose | Mean | Standard Deviation | Sample Size Per Group | Allocation Ratio |
|---------|------|------|--------------------|-----------------------|------------------|
| Control | 0 | 0 | 5 | 50 | 1 |
| 1 | 0.3 | 0.45 | 5 | 50 | 1 |
| 2 | 1 | 1.5 | 5 | 50 | 1 |
| 3 | 2 | 3 | 5 | 50 | 1 |

Simulation Seed and Elapsed Time

| | |
|-----------------------------|----------|
| Starting Seed | 86876095 |
| Total Number of Simulations | 10000 |
| Elapsed Time | 00:00:03 |

Summary

Using Holm's step down test, this study has 0.74 global power to detect at least one treatment arm which is different from control arm given a total sample size 200. This test has conjunctive power 0.045 to detect all treatment arms which are truly different from control arm. This test provides disjunctive power 0.74 to detect at least one treatment arm which is truly different from control arm.

15.2.3 Data-driven step-up MC procedures

Step-up tests start with the least significant comparison and continue as long as tests are not significant until the first time when a significant comparison occurs and all remaining hypotheses will be rejected. East supports two such MC procedures -

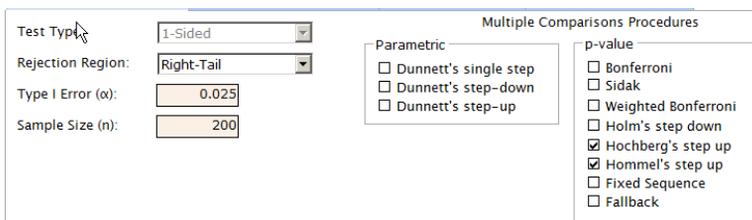
15 Multiple Comparison Procedures for Continuous Data

Hochberg step-up and Hommel step-up procedures. In the Hochberg step-up procedure, in i -th step $H_{(k-i)}$ is retained if $p_{(k-i)} > \frac{\alpha}{i}$. In the Hommel step-up procedure, in i -th step $H_{(k-i)}$ is retained if $p_{(k-j)} > \frac{i-j+1}{i} \alpha$ for $j = 1, \dots, i$. Fixed sequence test and fallback test are the types of tests which proceed in a prespecified order.

Hochberg's and Hommel's step-up procedures

Hochberg's and Hommel's step-up procedures are described below using the same Alzheimer's Disease example from the section 15.1.1 on Dunnett's Single Step.

Since the other design specification remains same except that we are using Hocheberg and Hommel step-up procedures in place of single step Dunnett's test we can design simulation in this section with only little effort. Select Sim 1 in **Library** and click . Now go to the **Design Parameters** tab. There in the **Multiple Comparison Procedures** box, uncheck the **Dunnett's single step** box and check the **Hochberg's step-up** and **Hommel's step-up** boxes.



Now click **Simulate** to obtain power. Once the simulation run has completed, East will add two additional rows to the **Output Preview** labeled as Sim 9 and Sim 10.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. Pairwise Comps. | Sample Size | Rejection Region | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|-------|-------------------------------|-----------|--------------------|--------------|---------------------|-------------|------------------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Dunnett's single step | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.75 | 0.019 | 0.75 | 200 |
| Sim3 | Dunnett's step-down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.746 | 0.042 | 0.746 | 200 |
| Sim4 | Dunnett's step-up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.741 | 0.045 | 0.741 | 200 |
| Sim5 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.729 | 0.017 | 0.729 | 200 |
| Sim6 | Sidak | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.732 | 0.018 | 0.732 | 200 |
| Sim7 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.016 | 0.737 | 200 |
| Sim8 | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.74 | 0.045 | 0.74 | 200 |
| Sim9 | Hochberg's step up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.739 | 0.051 | 0.739 | 200 |
| Sim10 | Hommel's step up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.048 | 0.737 | 200 |

Hocheberg and Hommel procedures have disjunctive and global powers of close to 74

15.2.4 Fixed-sequence stepwise MC procedures

In data-driven stepwise procedures, we don't have any control on the order of the hypotheses to be tested. However, sometimes based on our preference or prior knowledge we might want to fix the order of tests a priori. Fixed sequence test and fallback test are the types of tests which proceed in a pre-specified order. East supports both of these procedures.

Assume that H_1, H_2, \dots, H_{k-1} are ordered hypotheses and the order is pre-specified so that H_1 is tested first followed by H_2 and so on. Let p_1, p_2, \dots, p_{k-1} be the associated raw marginal p-values. In the fixed sequence testing procedure, for $i = 1, \dots, k - 1$, in i -th step, if $p_i < \alpha$, reject H_i and go to the next step; otherwise retain H_i, \dots, H_{k-1} and stop.

Fixed sequence testing strategy is optimal when early tests in the sequence have largest treatment effect and performs poorly when early hypotheses have small treatment effect or are nearly true (Westfall and Krishen 2001). The drawback of fixed sequence test is that once a hypothesis is not rejected no further testing is permitted. This will lead to lower power to reject hypotheses tested later in the sequence.

Fallback test alleviates the above undesirable feature for fixed sequence test. Let w_i be the proportion of α for testing H_i such that $\sum_{i=1}^{k-1} w_i = 1$. In the fixed sequence testing procedure, in i -th step ($i = 1, \dots, k - 1$), test H_i at $\alpha_i = \alpha_{i-1} + \alpha w_i$ if H_{i-1} is rejected and at $\alpha_i = \alpha w_i$ if H_{i-1} is retained. If $p_i < \alpha_i$, reject H_i ; otherwise retain it. Unlike the fixed sequence testing approach, the fallback procedure can continue testing even if a non-significant outcome is encountered by utilizing the fallback strategy. If a hypothesis in the sequence is retained, the next hypothesis in the sequence is tested at the level that would have been used by the weighted Bonferroni procedure. With $w_1 = 1$ and $w_2 = \dots = w_{k-1} = 0$, the fallback procedure simplifies to fixed sequence procedure.

Fixed sequence testing procedure

As before we will use the same Alzheimer's Disease example to illustrate fixed sequence testing procedure. Select Sim 1 in **Library** and click . Now go to the **Design Parameters** tab. There in the **Multiple Comparison Procedures** box,

15 Multiple Comparison Procedures for Continuous Data

uncheck the **Dunnett's single step** box and check the **Fixed Sequence** box.

Next click on **Response Generation Info** tab and look at the **Table of Proportions**. You will see an additional column with label **Test Sequence** is added. Here you have to specify the order in which the hypotheses will be tested. Specify 1 for the test that will be tested first, 2 for the test that will be tested next and so on. By default East specifies 1 to the first test, 2 to the second test and so on. For now we will keep the default which means that H_1 will be tested first followed by H_2 and finally H_3 will be tested.

Table of Means

| Arm | Dose | Mean | Std.Dev. | Test Sequence |
|---------|------|------|----------|---------------|
| Control | 0 | 0 | 5 | |
| 1 | 0.3 | 0.45 | 5 | 1 |
| 2 | 1 | 1.5 | 5 | 2 |
| 3 | 2 | 3 | 5 | 3 |

Now click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 11.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. Pairwise Comps. | Sample Size | Rejection Region | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|-------|-------------------------------|-----------|--------------------|--------------|---------------------|-------------|------------------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Dunnett's single step | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.75 | 0.019 | 0.75 | 200 |
| Sim3 | Dunnett's step-down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.746 | 0.042 | 0.746 | 200 |
| Sim4 | Dunnett's step-up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.741 | 0.045 | 0.741 | 200 |
| Sim5 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.729 | 0.017 | 0.729 | 200 |
| Sim6 | Sidak | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.732 | 0.018 | 0.732 | 200 |
| Sim7 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.016 | 0.737 | 200 |
| Sim8 | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.74 | 0.045 | 0.74 | 200 |
| Sim9 | Hochberg's step up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.739 | 0.051 | 0.739 | 200 |
| Sim10 | Hommel's step up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.048 | 0.737 | 200 |
| Sim11 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.068 | 0.05 | 0.068 | 200 |

The fixed sequence procedure with the specified sequence has global and disjunctive power of less than 7% and conjunctive power of 5%. The reason for small global and

disjunctive power is due to the smallest treatment effect is tested first and the magnitude of treatment effect increases gradually for the remaining tests. For optimal power in fixed sequence procedure, the early tests in the sequence should have larger treatment effects. In our case, Dose3 has largest treatment effect followed by Dose2 and Dose1. Therefore, to obtain optimal power, H_3 should be tested first followed by H_2 and H_1 .

Select Sim 11 in the **Output Preview** and click  icon. Select Sim 11 in **Library**, click  and go to the **Response Generation Info** tab. In **Test Sequence** column in the table, specify 3 for Dose1, 2 for Dose2 and 1 for Dose3.

Table of Means

| Arm | Dose | Mean | Std.Dev. | Test Sequence |
|---------|------|------|----------|---------------|
| Control | 0 | 0 | 5 | |
| 1 | 0.3 | 0.45 | 5 | 3 |
| 2 | 1 | 1.5 | 5 | 2 |
| 3 | 2 | 3 | 5 | 1 |

Now click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional rows to the **Output Preview** labeled as Sim 12.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. Pairwise Comparisons | Sample Size | Rejection Region | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|-------|-------------------------------|-----------|--------------------|--------------|--------------------------|-------------|------------------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Dunnett's single step | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.75 | 0.019 | 0.75 | 200 |
| Sim3 | Dunnett's step-down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.746 | 0.042 | 0.746 | 200 |
| Sim4 | Dunnett's step-up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.741 | 0.045 | 0.741 | 200 |
| Sim5 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.729 | 0.017 | 0.729 | 200 |
| Sim6 | Sidak | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.732 | 0.018 | 0.732 | 200 |
| Sim7 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.016 | 0.737 | 200 |
| Sim8 | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.74 | 0.045 | 0.74 | 200 |
| Sim9 | Hochberg's step up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.739 | 0.051 | 0.739 | 200 |
| Sim10 | Hommel's step up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.048 | 0.737 | 200 |
| Sim11 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.068 | 0.05 | 0.068 | 200 |
| Sim12 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.849 | 0.046 | 0.849 | 200 |

Now the fixed sequence procedure with the pre-specified sequence (H_3, H_2, H_1) has global and disjunctive power close to 85% and conjunctive power close to 5%. This example illustrates that fixed sequence procedure is powerful provided the hypotheses are tested in a sequence of descending treatment effects. Fixed sequence procedure controls the FWER because for each hypothesis, testing is conditional upon rejecting all hypotheses earlier in sequence. The exact result of the simulations may differ slightly, depending on the seed. Select Sim 12 in the **Output Preview** and click .

15 Multiple Comparison Procedures for Continuous Data

icon to save it in **Library**.

Fallback procedure

Again we will use the same Alzheimer’s Disease example to illustrate the fallback procedure. Select Sim 1 in **Library** and click . There in the **Multiple Comparison Procedures** box, uncheck the **Dunnett’s single step** box and check the **Fallback** box.

| | | | |
|-------------------|------------|--|--|
| Test Type: | 1-Sided | Multiple Comparisons Procedures | |
| Rejection Region: | Right-Tail | Parametric | p-value |
| Type I Error (α): | 0.025 | <input type="checkbox"/> Dunnett’s single step | <input type="checkbox"/> Bonferroni |
| Sample Size (n): | 200 | <input type="checkbox"/> Dunnett’s step-down | <input type="checkbox"/> Sidak |
| | | <input type="checkbox"/> Dunnett’s step-up | <input type="checkbox"/> Weighted Bonferroni |
| | | | <input type="checkbox"/> Holm’s step down |
| | | | <input type="checkbox"/> Hochberg’s step up |
| | | | <input type="checkbox"/> Hommel’s step up |
| | | | <input type="checkbox"/> Fixed Sequence |
| | | | <input checked="" type="checkbox"/> Fallback |

Next click on **Response Generation Info** tab and look at the **Table of Proportions**. You will see two additional columns with label **Test Sequence** and **Proportion of Alpha**. In the column **Test Sequence**, you have to specify the order in which the hypotheses will be tested. Specify 1 for the test that will be tested first, 2 for the test that will be tested next and so on. By default East specifies 1 to the first test, 2 to the second test and so on. For now we will keep the default which means that H_1 will be tested first followed by H_2 and finally H_3 will be tested.

In the column **Proportions of Alpha**, you have to specify the proportion of total alpha you want to spend in each test. Ideally, the values in this column should add up to 1; if not, then East will normalize it to add them up to 1. By default East distributes the total alpha equally among the all tests. Here we have 3 tests in total, therefore each of the tests have proportion of alpha as $1/3$ or 0.333 . You can specify other proportions as well. For this example, keep the equal proportion of alpha for each test.

Table of Means

| Dose | Mean | Std.Dev. | Test Sequence | Proportion of Alpha |
|------|------|----------|---------------|---------------------|
| 0 | 0 | 5 | | |
| 0.3 | 0.45 | 5 | 1 | 0.333 |
| 1 | 1.5 | 5 | 2 | 0.333 |
| 2 | 3 | 5 | 3 | 0.333 |

Now click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional rows to the **Output Preview** labeled as Sim 13.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. Pairwise Comps. | Sample Size | Rejection Region | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|-------|-------------------------------|-----------|--------------------|--------------|---------------------|-------------|------------------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Dunnett's single step | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.75 | 0.019 | 0.75 | 200 |
| Sim3 | Dunnett's step-down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.746 | 0.042 | 0.746 | 200 |
| Sim4 | Dunnett's step-up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.741 | 0.045 | 0.741 | 200 |
| Sim5 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.729 | 0.017 | 0.729 | 200 |
| Sim6 | Sidak | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.732 | 0.018 | 0.732 | 200 |
| Sim7 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.016 | 0.737 | 200 |
| Sim8 | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.74 | 0.045 | 0.74 | 200 |
| Sim9 | Hochberg's step up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.739 | 0.051 | 0.739 | 200 |
| Sim10 | Hommel's step up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.048 | 0.737 | 200 |
| Sim11 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.068 | 0.05 | 0.068 | 200 |
| Sim12 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.849 | 0.046 | 0.849 | 200 |
| Sim13 | Fallback | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.02 | 0.737 | 200 |

Now we will consider a sequence where H_3 will be tested first followed by H_2 and H_1 because in our case, Dose3 has largest treatment effect followed by Dose2 and Dose1.

Select Sim 13 in the **Output Preview** and click  icon. Select Sim 12 in **Library**, click  and go to the **Response Generation Info** tab. In **Test Sequence** column in the table, specify 3 for Dose1, 2 for Dose2 and 1 for Dose3.

Table of Means

| Dose | Mean | Std.Dev. | Test Sequence | Proportion of Alpha |
|------|------|----------|---------------|---------------------|
| 0 | 0 | 5 | | |
| 0.3 | 0.45 | 5 | 3 | 0.333 |
| 1 | 1.5 | 5 | 2 | 0.333 |
| 2 | 3 | 5 | 1 | 0.333 |

Now click **Simulate** to obtain power. Once the simulation run has completed, East will

15 Multiple Comparison Procedures for Continuous Data

add an additional rows to the **Output Preview** labeled as Sim 14.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. Pairwise Comps. | Sample Size | Rejection Region | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|-------|-------------------------------|-----------|--------------------|--------------|---------------------|-------------|------------------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Dunnett's single step | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.75 | 0.019 | 0.75 | 200 |
| Sim3 | Dunnett's step-down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.746 | 0.042 | 0.746 | 200 |
| Sim4 | Dunnett's step-up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.741 | 0.045 | 0.741 | 200 |
| Sim5 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.729 | 0.017 | 0.729 | 200 |
| Sim6 | Sidak | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.732 | 0.018 | 0.732 | 200 |
| Sim7 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.016 | 0.737 | 200 |
| Sim8 | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.74 | 0.045 | 0.74 | 200 |
| Sim9 | Hochberg's step up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.739 | 0.051 | 0.739 | 200 |
| Sim10 | Hommel's step up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.048 | 0.737 | 200 |
| Sim11 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.068 | 0.05 | 0.068 | 200 |
| Sim12 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.849 | 0.046 | 0.849 | 200 |
| Sim13 | Fallback | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.02 | 0.737 | 200 |
| Sim14 | Fallback | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.735 | 0.045 | 0.735 | 200 |

Note that the fallback test is more robust to the misspecification of the test sequence but fixed sequence test is very sensitive to the test sequence. If the test order is misspecified, fixed sequence test has very poor performance.

15.3 Comparison of MC procedures

We have obtained the power (based on the simulation) for different MC procedures for the Alzheimer's Disease example from the section 15.1.1. Now the obvious question is which MC procedure to choose. To compare all the MC procedure, we will perform simulation for all the MC procedures under the following scenario.

- Treatment arms: placebo, dose1 (dose=0.3 mg), dose2 (dose=1 mg) and dose3 (dose=2 mg) with respective groups means as 0, 0.45, 1.5 and 3, respectively.
- common standard deviation = 5
- Type I Error: 0.025 (right-tailed)
- Number of Simulations:10000
- Total Sample Size:200
- Allocation ratio: 1 : 1 : 1 : 1

For comparability of simulation results, we have used similar seed for simulation under all MC procedures. Following output displays the powers under different MC

procedures.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. Pairwise Comps. | Sample Size | Rejection Region | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|-------|-------------------------------|-----------|--------------------|--------------|---------------------|-------------|------------------|--------------|-------------------|-------------------|---------------------|
| Sim12 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.849 | 0.046 | 0.849 | 200 |
| Sim1 | Dunnett's single step | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.75 | 0.019 | 0.75 | 200 |
| Sim3 | Dunnett's step-down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.746 | 0.042 | 0.746 | 200 |
| Sim4 | Dunnett's step-up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.741 | 0.045 | 0.741 | 200 |
| Sim8 | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.74 | 0.045 | 0.74 | 200 |
| Sim9 | Hochberg's step up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.739 | 0.051 | 0.739 | 200 |
| Sim10 | Hommel's step up | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.048 | 0.737 | 200 |
| Sim7 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.016 | 0.737 | 200 |
| Sim13 | Fallback | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.737 | 0.02 | 0.737 | 200 |
| Sim14 | Fallback | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.735 | 0.045 | 0.735 | 200 |
| Sim6 | Sidak | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.732 | 0.018 | 0.732 | 200 |
| Sim5 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.729 | 0.017 | 0.729 | 200 |
| Sim11 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 200 | Right-Tail | 0.068 | 0.05 | 0.068 | 200 |

Here we have used equal proportions for weighted Bonferroni and Fallback procedures. For the two fixed sequence testing procedures (fixed sequence and fallback) two sequences have been used - (H_1, H_2, H_3) and (H_3, H_2, H_1) . As expected, Bonferroni and weighted Bonferroni procedures provides similar powers. It appears that fixed sequence procedure with the pre-specified sequence (H_3, H_2, H_1) provides the power of close to 85% which is the maximum among all the procedures. However, fixed sequence procedure with the pre-specified sequence (H_1, H_2, H_3) provides power of less than 7%. Therefore, power in fixed sequence procedure is largely dependent on the specification of sequence of testing and a mis-specification might result in huge drop in power. For this reason, fixed sequence procedure may not be considered as appropriate MC procedure to go with.

Dunnett's single step, step-down and step-up procedures are the next in order after fixed sequence procedure with the pre-specified sequence (H_3, H_2, H_1) . All the three procedures attain close to 75% of disjunctive power, respectively. However, all these three procedures assume that all the treatment arms have equal variance. Therefore, if homogeneity of variance between the treatment arms is a reasonable assumption, Dunnett's step-down or single step procedure should be the best option based on these simulation results. However, when the assumption of equal variance is not met, Dunnett's procedure may not be the appropriate procedure as the type I error might not be strongly controlled.

Next in the list are the fallback procedures and both of them provides a little more than 73% power which is very close to the power attained by Dunnett's procedures. Therefore, unlike fixed sequence procedure, fallback procedure does not depend much on the order of the hypotheses they are tested. Moreover, this does not require the

15 Multiple Comparison Procedures for Continuous Data

assumption of equal variance among the treatment arms to be met. For all these reasons, fallback procedure seems to be the most appropriate MC procedure for the design we are interested in.

Now, we will perform the comparison but this time with unequal variance between the treatment arms. Precisely, we simulate data under the following scenario to see the type I error rate control of different procedures.

- Treatment arms: placebo, dose1 (dose=0.3 mg), dose2 (dose=1 mg) and dose3 (dose=2 mg) with respective groups means as 0, 0, 0 and 0, respectively.
- standard deviation for placebo, dose1 and dose2 is 5; standard deviation for dose3 is 10
- Type I Error: 0.025 (right-tailed)
- Number of Simulations:1000000
- Total Sample Size:200
- Allocation ratio: 1 : 1 : 1 : 1

Following output displays the type I error rate under different MC procedures for the unequal variance case.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. Pairwise Comps. | Sample Size | Rejection Region | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|-------|-------------------------------|-----------|--------------------|--------------|---------------------|-------------|------------------|--------------|-------------------|-------------------|---------------------|
| Sim15 | Dunnett's single step | 1-Sided | 0.025 | 0.027 | 3 | 200 | Right-Tail | 0.027 | 0 | 0 | 200 |
| Sim16 | Dunnett's step-down | 1-Sided | 0.025 | 0.027 | 3 | 200 | Right-Tail | 0.027 | 0 | 0 | 200 |
| Sim17 | Dunnett's step-up | 1-Sided | 0.025 | 0.026 | 3 | 200 | Right-Tail | 0.026 | 0 | 0 | 200 |
| Sim13 | Fixed Sequence | 1-Sided | 0.025 | 0.025 | 3 | 200 | Right-Tail | 0.025 | 0 | 0 | 200 |
| Sim12 | Hommel's step up | 1-Sided | 0.025 | 0.024 | 3 | 200 | Right-Tail | 0.024 | 0 | 0 | 200 |
| Sim10 | Hochberg's step up | 1-Sided | 0.025 | 0.024 | 3 | 200 | Right-Tail | 0.024 | 0 | 0 | 200 |
| Sim6 | Sidak | 1-Sided | 0.025 | 0.023 | 3 | 200 | Right-Tail | 0.023 | 0 | 0 | 200 |
| Sim9 | Holm's step down | 1-Sided | 0.025 | 0.023 | 3 | 200 | Right-Tail | 0.023 | 0 | 0 | 200 |
| Sim7 | Weighted Bonferroni | 1-Sided | 0.025 | 0.023 | 3 | 200 | Right-Tail | 0.023 | 0 | 0 | 200 |
| Sim14 | Fallback | 1-Sided | 0.025 | 0.023 | 3 | 200 | Right-Tail | 0.023 | 0 | 0 | 200 |
| Sim5 | Bonferroni | 1-Sided | 0.025 | 0.023 | 3 | 200 | Right-Tail | 0.023 | 0 | 0 | 200 |

Note that the Dunnett tests slightly inflate type I error rate but all other procedures control the type I error rate below the nominal level 0.025.

16 *Multiple Endpoints-Gatekeeping Procedures*

16.1 *Introduction*

Clinical trials are often designed to assess benefits of a new treatment compared to a control treatment with respect to multiple clinical endpoints which are divided into hierarchically ordered families. Typically, the primary family of endpoints defines the overall outcome of the trial, provides the basis for regulatory claim and is included in the product label. The secondary families of endpoints play a supportive role and provide additional information for physicians, patients, payers and hence are useful for enhancing product label. Gatekeeping procedures are specifically designed to address this type of multiplicity problems by explicitly taking into account the hierarchical structure of the multiple objectives. The terminology-gatekeeping indicates the hierarchical decision structure where the higher ranked families serve as gatekeepers for the lower ranked family. The lower ranked families won't be tested if the higher ranked families are not passed. Two types of gatekeeping procedures are described in this chapter. One is serial gatekeeping procedure and the other one is parallel gatekeeping procedure. In the next few sections, specific examples will be provided to illustrate how to design trials with each type of gatekeeping procedures. For more information about applications of gatekeeping procedures in a clinical trial setting and literature review on this topic, please refer to Dmitrienko and Tamhane (2007).

16.2 *Simulate Serial Gatekeeping Design*

Serial gatekeeping procedures were studied by Maurer, Hothorn and Lehman (1995), Bauer et al. (1998) and Westfall and Krishen (2001). Serial gatekeepers are encountered in trials where endpoints are usually ordered from most important to least important. Reisberg et al. 2003 reported a study designed to investigate memantine, an N-methyl-D-aspartate (NMDA) antagonist, for the treatment of Alzheimer's disease in which patients with moderate-to-severe Alzheimer's disease were randomly assigned to receive placebo or 20 mg of memantine daily for 28 weeks. The two primary efficacy variables were: (1) the Clinician's Interview-Based Impression of Change Plus Caregiver Input (CIBIC-Plus) global score at 28 weeks, (2) the change from base line to week 28 in the Alzheimer's Disease Cooperative Study Activities of Daily Living Inventory modified for severe dementia (ADCS-ADLsev). The CIBIC-Plus measures overall global change relative to base line and is scored on a seven-point scale ranging from 1 (markedly improved) to 7 (markedly worse). For illustration purpose, we redefine the primary endpoint of clinician's global assessment score as 7 minus the CIBIC-Plus score so that a larger value indicates improvement (0 markedly worse and 6 markedly improved). The secondary efficacy endpoints included the Severe Impairment Battery and other measures of cognition, function, and behavior. Suppose that the trial is declared successful only if the treatment effect is demonstrated on both endpoints. If the trial is successful, it is of interest to assess the two secondary endpoints: (1) Severe Impairment Battery (SIB), (2) Mini-Mental State Examination

16 Multiple Endpoints-Gatekeeping Procedures

(MMSE). The SIB was designed to evaluate cognitive performance in advanced Alzheimer' disease. A 51-item scale, it assesses social interaction, memory, language, visuospatial ability, attention, praxis, and construction. The scores range from 0 (greatest impairment) to 100. The MMSE is a 30-point scale that measures cognitive function. The means of the endpoints for subjects in the control group and experimental group and the common covariance matrix are as follows

| | Mean Treatment | Mean Control |
|-------------|----------------|--------------|
| CIBIC-Plus | 2.6 | 2.3 |
| ADCS-ADLsev | -2.5 | -4.5 |
| SIB | -6.5 | -10 |
| MMSE | -0.4 | -1.2 |

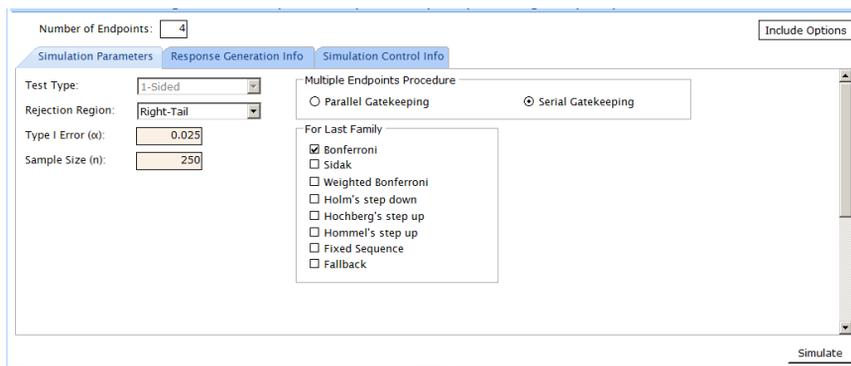
| CIBIC-Plus | ADCS-ADLsev | SIB | MMSE |
|------------|-------------|-----|------|
| 1.2 | 3.6 | 6.8 | 1.6 |
| 3.6 | 42 | 38 | 9.3 |
| 6.8 | 38 | 145 | 17 |
| 1.6 | 9.3 | 17 | 8 |

Typically there are no analytical ways to compute the power for gatekeeping procedures. Simulations can be used to assess the operating characteristics of different designs. For example, one could simulate the power for given sample sizes. To start the simulations, click [Two Samples](#) in the [Design](#) tab and select [Multiple Comparisons-Multiple Endpoints](#) to see the following input windows

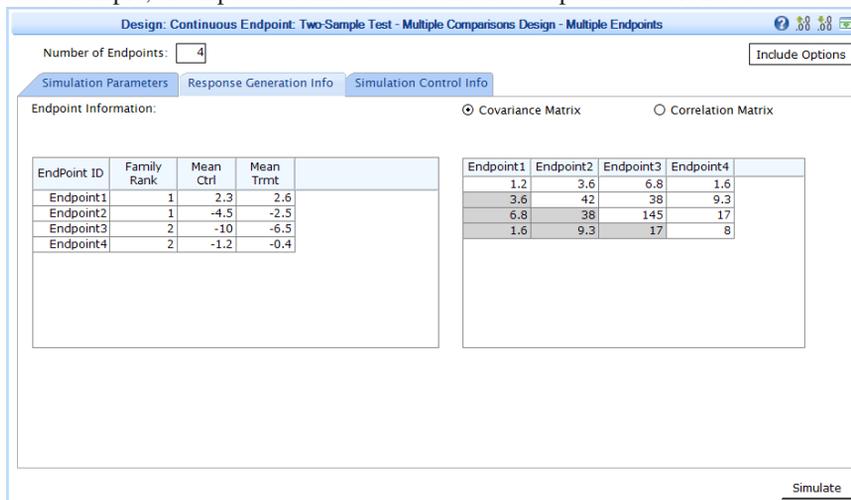
The screenshot displays the 'Design: Continuous Endpoint: Two-Sample Test - Multiple Comparisons Design - Multiple Endpoints' window. The 'Number of Endpoints' is set to 4. The 'Simulation Parameters' tab is active, showing 'Test Type' as 1-Sided, 'Rejection Region' as Right-Tail, 'Type I Error (α)' as 0.025, and 'Sample Size (n)' as 100. The 'Multiple Endpoints Procedure' section has 'Serial Gatekeeping' selected. The 'For Last Family' section has 'Bonferroni' checked, with other options like Sidak, Weighted Bonferroni, Holm's step down, Hochberg's step up, Hommel's step up, Fixed Sequence, and Fallback unchecked. A 'Simulate' button is located at the bottom right.

On the top of this input window, one needs to specify the total number of endpoints and other input parameters such as Rejection Region, Type I Error, Sample Size. One also needs to select the multiple comparison procedure which will be used to test the last family of endpoints. The type I error specified on this screen is the nominal level of the familywise error rate which is defined as the probability of falsely declaring the efficacy of the new treatment compared to control with respect to any endpoint. For the Alzheimer's disease example, CIBIC-Plus and ADCS-ADlsev form the primary family, and the other endpoints SIB and MMSE form the secondary family. Suppose that we would like to see the power for a sample size of 250 at a nominal type I error rate 0.025 using Bonferroni test for the secondary family, then the input window looks as follows

16 Multiple Endpoints-Gatekeeping Procedures



Behind the window for **Simulation Parameters**, there is another window tab labeled as **Response Generation Info**. The window for **Response Generation Info** tab shown below allows one to specify the underlying joint distribution among the multiple endpoints for control arm and for experimental arm. The joint distribution among the endpoints are assumed to be multivariate normal with common covariance matrix. One also needs to specify which family each endpoint belongs to in the column with label **Family Rank**. One can also customize the label for each endpoint. For the Alzheimer’s disease example, the inputs for this window should be specified as follows

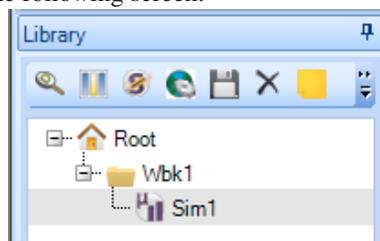


One can specify the number of simulations to be performed on the window with the

label **Simulation Control Info**. By default, 10000 simulations will be performed. One can also save the summary statistics for each simulated trial or save subject-level data by checking the appropriate box in the output option area. To simulate this design, click the **Simulate** button at the bottom right of the screen to see the preliminary output displayed in the output preview area as seen in the following screen. All the results displayed in the yellow cells are summary outputs generated from simulations. For example, the actually FWER, number of families, conjunctive power for the primary family, conjunctive power and disjunctive power for the last family.

| Output Preview | | | | | | | | | | | |
|----------------|--------------------------------|-----------|--------------------|--------------------|--------------|------------------|-----------------|-------------|------------------|-------------------------------|-------------------------------|
| ID | Multiple Comparisons Procedure | Test Type | Design Type | Specified α | Overall FWER | No. of Endpoints | No. of Families | Sample Size | Rejection Region | Conjunctive Power Last Family | Disjunctive Power Last Family |
| Sim1 | Bonferroni | 1-Sided | Serial Gatekeeping | 0.025 | 0 | 4 | 2 | 250 | Right-Tail | 0.259 | 0.417 |

To view the detailed output, first save the simulation into a workbook in the library by clicking on the tool button  and you will notice that a simulation node appears in the library as shown in the following screen.



Now double click on the simulation node **Sim1** to see the detailed output as shown in the following screen. The detailed output summarizes all the main input parameters such as the multiple comparison procedure used for the last family of endpoints, the nominal type I error level, total sample size, mean values for each endpoint in the control arm and that in the experimental arm etc. It also displays the attained overall FWER, conjunctive power, disjunctive power, the FWER and conjunctive power for each gatekeeper family, the FWER and conjunctive power and disjunctive power for the last family. The definitions of different types of power are as follows:

Overall Power and FWER:

Global: probability of declaring significance on any of the endpoints

Conjunctive: probability of declaring significance on all of the endpoints for which the

16 Multiple Endpoints-Gatekeeping Procedures

treatment arm is truly better than the control arm

Disjunctive: probability of declaring significance on any of the endpoints for which the treatment arm is truly better than the control arm

FWER: probability of making at least one type I error among all the endpoints

Power and FWER for Individual Gatekeeper Family except the Last Family:

Conjunctive Power: probability of declaring significance on all of the endpoints in the particular gatekeeper family for which the treatment arm is truly better than the control arm

FWER: probability of making at least one type I error when testing the endpoints in the particular gatekeeper family

Power and FWER for the Last Family:

Conjunctive Power: probability of declaring significance on all of the endpoints in the last family for which the treatment arm is truly better than the control arm

Disjunctive Power: probability of declaring significance on any of the endpoints in the last family for which the treatment arm is truly better than the control arm

FWER: probability of making at least one type I error when testing the endpoints in the last family

Marginal Power: probability of declaring significance on the particular endpoint

For the Alzheimer's disease example, the conjunctive power, which characterizes the power for the study, is 46.9% for a total sample size of 250. Using Bonferroni test for the last family, the design has 40.5% probability (disjunctive power for the last family) to detect the benefit of memantine with respect to at least one of the two secondary endpoints, SIB and MMSE. It has 25.1% chance (conjunctive power for the last family)

to declare the benefit of memantine with respect to both of the secondary endpoints.

Simulation: Continuous Endpoint: Two-Sample Test - Multiple Comparisons Design - Multiple Endpoints - Serial Gatekeeping

Hypothesis:
 $H_1: \mu_{i1} - \mu_{i0} \leq 0$ Vs. $K_i: \mu_{i1} - \mu_{i0} > 0$ for $i = 1, 2, \dots, n$ where n is the total number of endpoints.

| Simulation Parameters | |
|-----------------------------------|------------|
| Simulation ID | Sim2 |
| Test Type | 1-Sided |
| Rejection Region | Right-Tail |
| Nominal Type-1 Error (α) | 0.025 |
| Sample Size (n) | 250 |

| Multiple Endpoint Procedure | |
|---|------------|
| Total Number of Endpoints | 4 |
| Total Number of Families | 2 |
| Gatekeeping Procedure | Serial |
| Multiple Endpoints Procedure(Last Family) | Bonferroni |

| Simulation Control Parameters | |
|-------------------------------|-------|
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Overall Powers

| | |
|--|-------|
| Global (Reject any H_i) | 0.469 |
| Conjunctive (Reject all H_i where $\mu_{i1} > \mu_{i0}$) | 0.251 |
| Disjunctive (Reject at least one H_i where $\mu_{i1} > \mu_{i0}$) | 0.469 |
| FWER (Reject any H_i where $\mu_{i1} \leq \mu_{i0}$) | 0 |

Output for Gatekeeper Families

| Family Rank | Conjunctive Power | FWER |
|-------------|-------------------|------|
| 1 | 0.469 | 0 |

Output for Last Family

| Family Rank | Conjunctive Power | Disjunctive Power | FWER |
|-------------|-------------------|-------------------|------|
| 2 | 0.251 | 0.405 | 0 |

Marginal Powers for Endpoints with $\mu_{i1} > \mu_{i0}$

| Endpoint ID | Marginal Power |
|-------------|----------------|
| Endpoint1 | 0.469 |
| Endpoint2 | 0.469 |
| Endpoint3 | 0.333 |
| Endpoint4 | 0.323 |

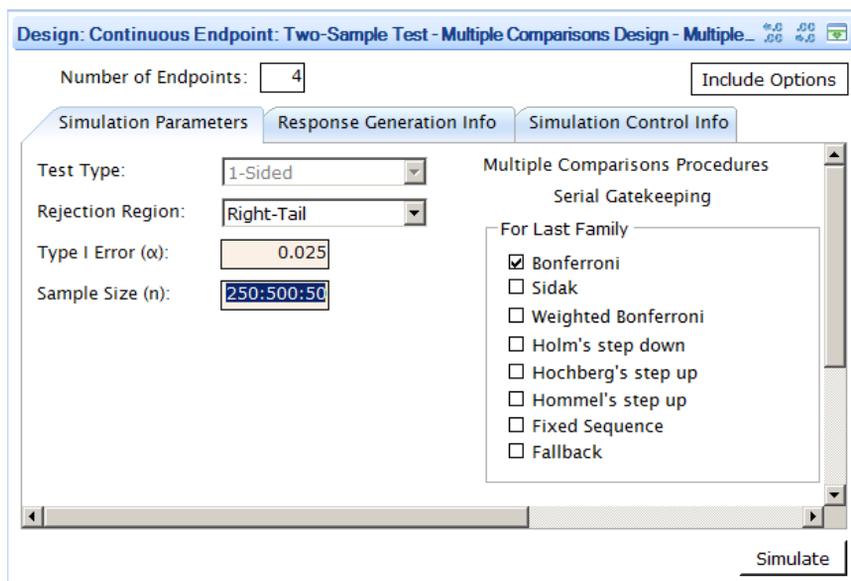
Response Parameters

| Endpoint ID | Family Rank | Mean Control | Mean Treatment | Standard Deviation |
|-------------|-------------|--------------|----------------|--------------------|
| Endpoint1 | 1 | 2.3 | 2.6 | 1.095 |

One can find the sample size to achieve a target power by simulating multiple designs in a batch mode. For example, one could simulate a batch of designs for a range of

16 Multiple Endpoints-Gatekeeping Procedures

sample size changing from 250 to 500 in step of 50 as shown in the following window.



Note that a total sample size somewhere between 450 to 500 provides 80% power to detect the mean differences for both primary endpoints CIBIC-Plus and ADCS-ADLsev as seen in the following window.

| ID | Multiple Endpoints Procedure for Last Family | Test Type | Design Type | Specified α | Overall FWER | Global Power | Conjunctive Power | Disjunctive Power | No. of Endpoints | No. of Families | Sample Size | Rejection Region | Conjunctive Power First Family |
|------|--|-----------|-------------|--------------------|--------------|--------------|-------------------|-------------------|------------------|-----------------|-------------|------------------|--------------------------------|
| Sim2 | Bonferroni | 1-Sided | Serial | 0.025 | 0 | 0.472 | 0.254 | 0.472 | 4 | 2 | 250 | Right-Tail | 0.472 |
| Sim3 | Bonferroni | 1-Sided | Serial | 0.025 | 0 | 0.563 | 0.336 | 0.563 | 4 | 2 | 300 | Right-Tail | 0.563 |
| Sim4 | Bonferroni | 1-Sided | Serial | 0.025 | 0 | 0.65 | 0.433 | 0.65 | 4 | 2 | 350 | Right-Tail | 0.65 |
| Sim5 | Bonferroni | 1-Sided | Serial | 0.025 | 0 | 0.715 | 0.511 | 0.715 | 4 | 2 | 400 | Right-Tail | 0.715 |
| Sim6 | Bonferroni | 1-Sided | Serial | 0.025 | 0 | 0.783 | 0.598 | 0.783 | 4 | 2 | 450 | Right-Tail | 0.783 |
| Sim7 | Bonferroni | 1-Sided | Serial | 0.025 | 0 | 0.828 | 0.668 | 0.828 | 4 | 2 | 500 | Right-Tail | 0.828 |

To get a more precise sample size to achieve 80% power, one could simulate a bunch of designs with the sample size ranging from 450 to 500 in step of 10. One will notice that a sample size of 480 provides over 80% power to claim the significant differences with respect to both primary endpoints.

| ID | Multiple Endpoints Procedure for Last Family | Test Type | Design Type | Specified α | Overall FWER | Global Power | Conjunctive Power | Disjunctive Power | No. of Endpoints | No. of Families | Sample Size | Rejection Region | Conjunctive Power First Family |
|------|--|-----------|-------------|--------------------|--------------|--------------|-------------------|-------------------|------------------|-----------------|-------------|------------------|--------------------------------|
| Sim2 | Bonferroni | 1-Sided | Serial | 0.025 | 0 | 0.783 | 0.597 | 0.783 | 4 | 2 | 450 | Right-Tail | 0.783 |
| Sim3 | Bonferroni | 1-Sided | Serial | 0.025 | 0 | 0.793 | 0.612 | 0.793 | 4 | 2 | 460 | Right-Tail | 0.793 |
| Sim4 | Bonferroni | 1-Sided | Serial | 0.025 | 0 | 0.803 | 0.627 | 0.803 | 4 | 2 | 470 | Right-Tail | 0.803 |
| Sim5 | Bonferroni | 1-Sided | Serial | 0.025 | 0 | 0.801 | 0.627 | 0.801 | 4 | 2 | 480 | Right-Tail | 0.801 |
| Sim6 | Bonferroni | 1-Sided | Serial | 0.025 | 0 | 0.818 | 0.648 | 0.818 | 4 | 2 | 490 | Right-Tail | 0.818 |
| Sim7 | Bonferroni | 1-Sided | Serial | 0.025 | 0 | 0.827 | 0.663 | 0.827 | 4 | 2 | 500 | Right-Tail | 0.827 |

One could compare the multiple designs side by side by clicking on the tool button  in the output preview area as follows:

| | Sim2 | Sim3 | Sim4 | Sim5 | Sim6 | Sim7 |
|--|------------|------------|------------|------------|------------|------------|
| Mnemonic | MN-25-ME | MN-25-ME | MN-25-ME | MN-25-ME | MN-25-ME | MN-25-ME |
| Test Parameters | | | | | | |
| Multiple Endpoints Procedure for Last Family | Bonferroni | Bonferroni | Bonferroni | Bonferroni | Bonferroni | Bonferroni |
| Test Type | 1-Sided | 1-Sided | 1-Sided | 1-Sided | 1-Sided | 1-Sided |
| Design Type | Serial | Serial | Serial | Serial | Serial | Serial |
| Specified α | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 |
| No. of Endpoints | 4 | 4 | 4 | 4 | 4 | 4 |
| No. of Families | 2 | 2 | 2 | 2 | 2 | 2 |
| Rejection Region | Right-Tail | Right-Tail | Right-Tail | Right-Tail | Right-Tail | Right-Tail |
| MCP Results | | | | | | |
| Overall FWER | 0 | 0 | 0 | 0 | 0 | 0 |
| Global Power | 0.783 | 0.793 | 0.803 | 0.801 | 0.818 | 0.827 |
| Conjunctive Power | 0.597 | 0.612 | 0.627 | 0.627 | 0.648 | 0.663 |
| Disjunctive Power | 0.783 | 0.793 | 0.803 | 0.801 | 0.818 | 0.827 |
| Conjunctive Power First Family | 0.783 | 0.793 | 0.803 | 0.801 | 0.818 | 0.827 |
| Sample Size | | | | | | |
| Maximum | 450 | 460 | 470 | 480 | 490 | 500 |
| Other Parameters | | | | | | |
| Starting Seed | 77592463 | 77592246 | 77600485 | 77604507 | 77607827 | 77611103 |
| Simulation Results (Overall) | | | | | | |
| Average Sample Size | 450 | 460 | 470 | 480 | 490 | 500 |

There is a special case where all the endpoints belong to one single family. The software handle this special case in a particular manner. Intersection-Union test will be applied to a single family of endpoints and the selected MCP for the last family in the Simulation Parameter tab is not applicable for this special case. For the Alzheimer disease example, if we are only interested in testing the two endpoints (CIBIC-Plus and ADCS-ADLsev) as co-primary endpoints as indicated by the family rank in the window for Response Generation Info, then the Intersection-Union test will be applied to the two endpoints so that each endpoint is tested at nominal level α . The detailed output window is slightly different in case of single family of endpoints as seen in the

16 Multiple Endpoints-Gatekeeping Procedures

following window.

Simulation: Continuous Endpoint: Two-Sample Test - Multiple Comparisons Design - Multiple Endpoints - Serial Gatekeeping

Hypothesis:
 $H_1: \mu_{iT} - \mu_{iC} \leq 0$ Vs. $K_i: \mu_{iT} - \mu_{iC} > 0$ for $i = 1, 2, \dots, n$ where n is the total number of endpoints.

| Simulation Parameters | |
|-----------------------------------|------------|
| Simulation ID | Sim8 |
| Test Type | 1-Sided |
| Rejection Region | Right-Tail |
| Nominal Type-1 Error (α) | 0.025 |
| Sample Size (n) | 480 |

| Multiple Endpoint Procedure | |
|------------------------------|-------------------------|
| Total Number of Endpoints | 2 |
| Total Number of Families | 1 |
| Multiple Endpoints Procedure | Intersection-Union test |

| Simulation Control Parameters | |
|-------------------------------|-------|
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Overall Powers

| | |
|--|-------|
| Global (Reject any H_i) | 0.803 |
| Conjunctive (Reject all H_i where $\mu_{iT} > \mu_{iC}$) | 0.803 |
| Disjunctive (Reject at least one H_i where $\mu_{iT} > \mu_{iC}$) | 0.803 |
| FWER (Reject any H_i where $\mu_{iT} \leq \mu_{iC}$) | 0 |

Output for Co-primary Endpoints

| Family Rank | Conjunctive Power | FWER |
|-------------|-------------------|------|
| 1 | 0.803 | 0 |

Marginal Powers for Endpoints with $\mu_{iT} > \mu_{iC}$

| Endpoint ID | Marginal Power |
|-------------|----------------|
| Endpoint1 | 0.803 |
| Endpoint2 | 0.803 |

Response Parameters

| Endpoint ID | Family Rank | Mean Control | Mean Treatment | Standard Deviation |
|-------------|-------------|--------------|----------------|--------------------|
| Endpoint1 | 1 | 2.3 | 2.6 | 1.095 |
| Endpoint2 | 1 | -4.5 | -2.5 | 6.481 |

Correlation Matrix

| | Endpoint1 | Endpoint2 |
|-----------|-----------|-----------|
| Endpoint1 | 1 | 0.507 |
| Endpoint2 | 0.507 | 1 |

16.3 Simulate Parallel Gatekeeping Design

Parallel gatekeeping procedures are often used in clinical trials with several primary objectives where each individual objective can characterize a successful trial outcome. In other words, the trial can be declared to be successful if at least one primary objective is met. Consider a randomized, double blinded and parallel group designed clinical trial to compare two vaccines against the human papilloma virus. Denote vaccine T the new vaccine and vaccine C the comparator. The primary objective of this study is to demonstrate that vaccine T is superior to vaccine C for the antigen type 16 or 18 which account for 70% of cervical cancer cases globally. If the new vaccine shows superiority over the comparator with respect to either antigen type 16 or 18, it is of interest to test the superiority of vaccine T to vaccine C for the antigen type 31 or 45. The two types of vaccines are compared based on the immunological response, i.e. the number of T-cell in the blood, seven months after the vaccination. Assume that the log transformed data is normally distributed with mean μ_{iT} or μ_{iC} ($i = 1, 2, 3, 4$) where the index 1, 2, 3, and 4 represent the four antigen types respectively. The null

Table 16.1: Mean response and Standard Deviation

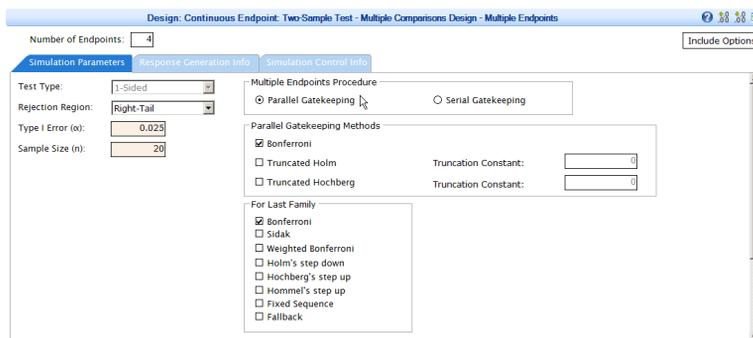
| Endpoints | Mean for Vaccine C | Mean for Vaccine T | Standard Deviation |
|-----------|--------------------|--------------------|--------------------|
| Type 16 | 4 | 4.57 | 0.5 |
| Type 18 | 3.35 | 4.22 | 0.5 |
| Type 31 | 2 | 2.34 | 0.6 |
| Type 45 | 1.42 | 2 | 0.3 |

hypotheses and alternative hypotheses can be formulated as

$$H_{i0} : \mu_{iT} - \mu_{iC} \leq 0 \text{ vs } H_{i1} : \mu_{iT} - \mu_{iC} > 0$$

The parallel gatekeeping test strategy is suitable for this example. The two null hypotheses H_{10} and H_{20} for antigen type 16 and 18 constitute the primary family which serves as the gatekeeper for the second family of hypotheses which contains H_{30} and H_{40} . Assume that the means and the standard deviations for all four antigen types are as follows:

Assume that the total sample size is 20 and one-sided significance level is 0.025. To assess the operating characteristics of the parallel gatekeeping procedures, we first need to open the simulation window for multiple endpoints. To this end, click on the **Design** menu, choose **Two Sample** for continuous endpoint and then select **Multiple Endpoints** from the drop-down list and the following screen will show up.



On the top of the above screen, one need to specify the total number of endpoints. The

16 Multiple Endpoints-Gatekeeping Procedures

lower part of the above screen is the **Simulation Parameters** tab which allows one to specify the important design parameters including the nominal type I error rate, total sample size, multiple comparison procedures. Now select **Parallel Gatekeeping** and choose **Bonferroni** for the parallel gatekeeping methods. For the last family, select **Bonferroni** as the multiple testing procedure. Next to the **Simulation Parameters** tab are two additional tabs: **Response Generation Info** and **Simulation Control Info**. We need to specify the mean responses for each endpoint for both treatment and control arm as well as the covariance structure among the endpoints. In addition, we need to specify which family each specific endpoint belongs to in the column with the label **Family Rank** in the same table for specifying the mean responses. There are two ways of specifying the covariance structure: **Covariance Matrix** or **Correlation Matrix**. If the **Correlation Matrix** option is selected, one needs to input the standard deviation for each endpoint in the same table for specifying the mean responses. There is a simpler way to input the standard deviation for each endpoint if all the endpoints share a common standard deviation. This can be done by checking the box for **Common Standard Deviation** and specify the value of the common standard deviation in the box to the right hand side. One also need to specify the correlations among the endpoints in the table to the right hand side. Similarly, if all the endpoints have a common correlation, then we can just check the box for **Common Correlation** and specify the value of the common correlation in the box to the right. For the vaccine example, assume the endpoints share a common mild correlation 0.3. Then the window with completed inputs for generating data looks like the following screen.

Design: Continuous Endpoint: Two-Sample Test - Multiple Comparisons: Design - Multiple Endpoints

Number of Endpoints: Include Options

Simulation Parameters | **Response Generation Info** | Simulation Control Info

Endpoint Information:

Common Standard Deviation:

| EndPoint ID | Family Rank | Mean Ctrl | Mean Trmt | Std. Dev. |
|-------------|-------------|-----------|-----------|-----------|
| Endpoint1 | 1 | 4 | 4.57 | 0.5 |
| Endpoint2 | 1 | 3.35 | 4.22 | 0.5 |
| Endpoint3 | 2 | 2 | 2.34 | 0.5 |
| Endpoint4 | 2 | 1.42 | 2 | 0.3 |

Covariance Matrix Correlation Matrix

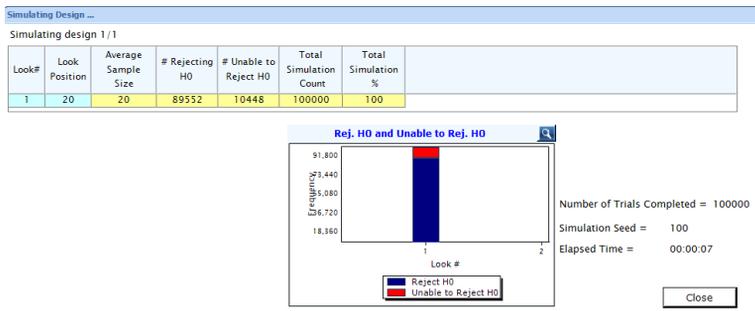
Common Correlation:

| Endpoint1 | Endpoint2 | Endpoint3 | Endpoint4 |
|-----------|-----------|-----------|-----------|
| 1 | 0.3 | 0.3 | 0.3 |
| 0.3 | 1 | 0.3 | 0.3 |
| 0.3 | 0.3 | 1 | 0.3 |
| 0.3 | 0.3 | 0.3 | 1 |

Simulate

In the window for **Simulation Control Info**, we can specify the total number of simulations, refresh frequency, type of random number seed. We can also choose to save the simulation data for more advanced analyses. After finishing specifying all the

input parameter values, click on the **Simulate** button on the bottom right of the window to run the simulations. The progress window will report how many simulations have been completed as seen in the following screen.



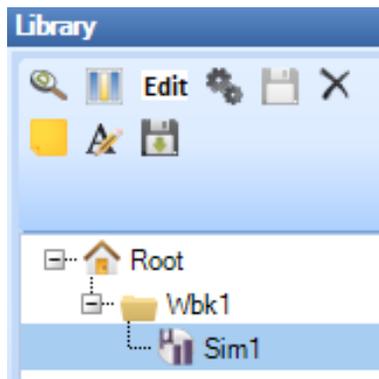
When all the requested simulations have been completed, click on the **Close** button at the right bottom of the progress report screen and the preliminary simulation summary will show up in the output preview window where one can see overall power summary and the power summary for the primary family as well as the attained overall FWER etc.

| ID | Multiple Endpoints Procedure for Last Family | Parallel Gatekeeping Method | Test Type | Design Type | Specified α | Overall FWER | Global Power | Conjunctive Power | Disjunctive Power | No. of Endpoints | No. of Families | Sample Size |
|------|--|-----------------------------|-----------|-------------|--------------------|--------------|--------------|-------------------|-------------------|------------------|-----------------|-------------|
| Sim1 | Bonferroni | Bonferroni | 1-Sided | Parallel | 0.025 | 0 | 0.945 | 0.101 | 0.945 | 4 | 2 | 20 |

To see the detailed output, we need to save the simulation in the workbook by clicking on the icon  on the top of the output preview window. A simulation node will be

16 Multiple Endpoints-Gatekeeping Procedures

appended in the corresponding workbook in the library as seen in the follow window.



Next double click on the simulation node in the library and the detailed outputs will be displayed accordingly.

Simulation: Continuous Endpoint: Two-Sample Test - Multiple Comparisons Design - Multiple Endpoints - Parallel Gatekeeping

Hypothesis:

$$H_1: \mu_{i_1} - \mu_{i_2} \leq 0 \text{ Vs. } K_i: \mu_{i_1} - \mu_{i_2} > 0 \text{ for } i = 1, 2, \dots, n \text{ where } n \text{ is the total number of endpoints.}$$

| Simulation Parameters | |
|---|------------------------|
| Simulation ID | Sim1cont03_BorF1_BorF2 |
| Test Type | 1-Sided |
| Rejection Region | Right-Tail |
| Nominal Type-1 Error (α) | 0.025 |
| Sample Size (n) | 20 |
| Multiple Endpoints Procedure | |
| Total Number of Endpoints | 4 |
| Total Number of Families | 2 |
| Gatekeeping Procedure | Parallel |
| Gatekeeping Method | Bonferroni |
| Common Correlation | 0.3 |
| Multiple Endpoints Procedure(Last Family) | Bonferroni |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 100000 |

Overall Powers

| | |
|---|--------|
| Global (Reject any H ₀) | 0.9449 |
| Conjunctive (Reject all H ₀ where μ ₁ > μ ₂) | 0.1012 |
| Disjunctive (Reject at least one H ₀ where μ ₁ > μ ₂) | 0.9449 |
| FWER (Reject any H ₀ where μ ₁ <= μ ₂) | 0 |

Output for Gatekeeper Families

| Family Rank | Conjunctive Power | Disjunctive Power | FWER |
|-------------|-------------------|-------------------|------|
| 1 | 0.5219 | 0.9449 | 0 |

Output for Last Family

| Family Rank | Conjunctive Power | Disjunctive Power | FWER |
|-------------|-------------------|-------------------|------|
| 2 | 0.1253 | 0.8955 | 0 |

Marginal Powers for Endpoints with μ₁ > μ₂

| Endpoint ID | Marginal Power |
|-------------|----------------|
| Endpoint1 | 0.5522 |
| Endpoint2 | 0.9146 |
| Endpoint3 | 0.127 |
| Endpoint4 | 0.8938 |

Response Parameters

| Endpoint ID | Family Rank | Mean Control | Mean Treatment | Standard Deviation |
|-------------|-------------|--------------|----------------|--------------------|
| Endpoint1 | 1 | 4 | 4.57 | 0.5 |
| Endpoint2 | 1 | 3.35 | 4.22 | 0.5 |
| Endpoint3 | 2 | 2 | 2.34 | 0.6 |
| Endpoint4 | 2 | 1.42 | 2 | 0.3 |

Correlation Matrix

| | Endpoint1 | Endpoint2 | Endpoint3 | Endpoint4 |
|-----------|-----------|-----------|-----------|-----------|
| Endpoint1 | 1 | 0.3 | 0.3 | 0.3 |
| Endpoint2 | 0.3 | 1 | 0.3 | 0.3 |
| Endpoint3 | 0.3 | 0.3 | 1 | 0.3 |
| Endpoint4 | 0.3 | 0.3 | 0.3 | 1 |

In case of testing multiple endpoints, the power definition is not unique. East provides the overall power summary and the power summary for each specific family. In the

overall power summary table, the following types of power are provided with the overall FWER: global power, conjunctive power and disjunctive power, which capture the overall performance of this gatekeeping procedure. The definitions of the powers are given below:

Overall Power and FWER:

Global: probability of declaring significance on any of the endpoints

Conjunctive: probability of declaring significance on all of the endpoints for which the treatment arm is truly better than the control arm

Disjunctive: probability of declaring significance on any of the endpoints for which the treatment arm is truly better than the control arm

FWER: probability of making at least one type I error among all the endpoints

Power and FWER for Individual Gatekeeper Families except the Last Family:

Conjunctive Power: probability of declaring significance on all of the endpoints in the particular gatekeeper family for which the treatment arm is truly better than the control arm

Disjunctive Power: probability of declaring significance on any of the endpoints in the particular gatekeeper family for which the treatment arm is truly better than the control arm

FWER: probability of making at least one type I error when testing the endpoints in the particular gatekeeper family

Power and FWER for the Last Gatekeeper Family:

Conjunctive Power: probability of declaring significance on all of the endpoints in the last family for which the treatment arm is truly better than the control arm

Disjunctive Power: probability of declaring significance on any of the endpoints in the last family for which the treatment arm is truly better than the control arm

FWER: probability of making at least one type I error when testing the endpoints in the last family

Marginal Power: probability of declaring significance on the particular endpoint

For the vaccine example, we see that the gatekeeping procedure using Bonferroni test for both the primary family and the secondary family provides 94.49% power to detect the difference in at least one of the two antigen types 16 and 18. It provides 52.19% power to detect the differences in both antigen types. Also note that this gatekeeping procedure only provides 89.55% power to detect the response difference in any of the other two antigen types 31 or 45 and only 12.53% to detect both antigen types 31 and 45. The marginal power table displays the probabilities of declaring significance on the

16 Multiple Endpoints-Gatekeeping Procedures

Table 16.2: Power Comparisons under Different Correlation Assumptions

| Correlation | Primary Family | | Secondary Family | | Overall Power | |
|-------------|----------------|-----------|------------------|-----------|---------------|-----------|
| | Disjunct. | Conjunct. | Disjunct. | Conjunct. | Disjunct. | Conjunct. |
| 0.3 | 0.9449 | 0.5219 | 0.8955 | 0.1253 | 0.9449 | 0.1012 |
| 0.5 | 0.9324 | 0.5344 | 0.8867 | 0.1327 | 0.9324 | 0.1192 |
| 0.8 | 0.9174 | 0.5497 | 0.8855 | 0.1413 | 0.9174 | 0.1402 |

particular endpoint after multiplicity adjustment. For example, the power of detecting antigen type 16 is 55.22%.

If it is of interest to assess the robustness of this procedure with respect to the correlation among the different endpoints, we can go back to the input window to change the correlations and run simulation again. To this end, right click on the Sim1 node in the library and select [Edit Simulation](#) from the dropdown list. Next click on the [Response Generation Info](#) tab, change the common correlation to 0.5 and click [Simulate](#) button. We can repeat this for a common correlation 0.8. The following table summarizes the power comparisons under different correlation assumptions. Note that the disjunctive power decreases as the correlation increases and conjunctive power increases as the correlation increases.

There are three available parallel gatekeeping methods: Bonferroni, Truncated Holm and Truncated Hochberg. The multiple comparison procedures applied to the gatekeeper families need to satisfy the so-called separable condition. A multiple comparison procedure is separable if the type I error rate under partial null configuration is strictly less than the nominal level α . Bonferroni is a separable procedure. However, the regular Holm and Hochberg procedure are not separable and can't be applied directly to the gatekeeper families. The truncated versions obtained by taking the convex combinations of the critical constants for the regular Holm/Hochberg procedure and Bonferroni procedure are separable and more powerful than Bonferroni test. The truncation constant leverages the degree of conservativeness. The larger value of the truncation constant results in more powerful procedure. If the truncation constant is set to be 1, it reduces to the regular Holm or Hochberg test. To see this, let's simulate the design using the truncated Holm procedure for the primary family and Bonferroni test for the second family for the vaccine example with common correlation 0.3. Table 3 compares the conjunctive power and disjunctive power for each family and the overall ones for different truncation parameter values. As the value of the truncation parameter increases, the conjunctive power for the primary family increases and the disjunctive power remain unchanged. Both the conjunctive power

Table 16.3: Impact of Truncation Constant in Truncated Holm Procedure on Overall Power and Power for Each Family

| Truncation Constant | Primary Family | | Secondary Family | | Overall Power | |
|---------------------|----------------|-----------|------------------|-----------|---------------|-----------|
| | Conjunct. | Disjunct. | Conjunct. | Disjunct. | Conjunct. | Disjunct. |
| 0 | 0.5219 | 0.9449 | 0.1253 | 0.8955 | 0.1012 | 0.9449 |
| 0.25 | 0.5647 | 0.9449 | 0.1229 | 0.8872 | 0.1065 | 0.9449 |
| 0.5 | 0.5988 | 0.9449 | 0.1212 | 0.8747 | 0.1108 | 0.9449 |
| 0.8 | 0.6327 | 0.9449 | 0.1188 | 0.84 | 0.115 | 0.9449 |

Table 16.4: Impact of Truncation Constant in Truncated Holm Procedure on Marginal Power

| Truncation Constant | Primary Family | | Secondary Family | |
|---------------------|----------------|---------|------------------|---------|
| | Type 16 | Type 18 | Type 31 | Type 45 |
| 0 | 0.5522 | 0.9146 | 0.127 | 0.8938 |
| 0.25 | 0.5886 | 0.921 | 0.1246 | 0.8855 |
| 0.5 | 0.6183 | 0.9254 | 0.1227 | 0.8731 |
| 0.8 | 0.6483 | 0.9293 | 0.1203 | 0.8385 |

and disjunctive power for the secondary family decrease as we increase the truncation parameter. The overall conjunctive power also increases but the overall disjunctive power remains the same with the increase of truncation parameter. Table 4 shows the marginal powers of this design for different truncation parameter values. The marginal powers for the two endpoints in the primary family increase. On the other hand, the marginal powers for the two endpoints in the secondary family decrease.

Table 5 and Table 6 displays the operating characteristics for truncation Hochberg test with different truncation constant values. Note that both the conjunctive and disjunctive powers for the primary family increase as the truncation parameter increases. However, the power for the secondary family decreases with the larger truncation parameter value. The marginal powers for the primary family and for the secondary family behave similarly. The overall conjunctive and disjunctive powers also increase as we increase the truncation parameter.

If all the endpoints belong to one single family, the selected multiple testing procedures for the last family (Bonferroni, Sidak, Weighted Bonferroni, Holm’s step

16 Multiple Endpoints-Gatekeeping Procedures

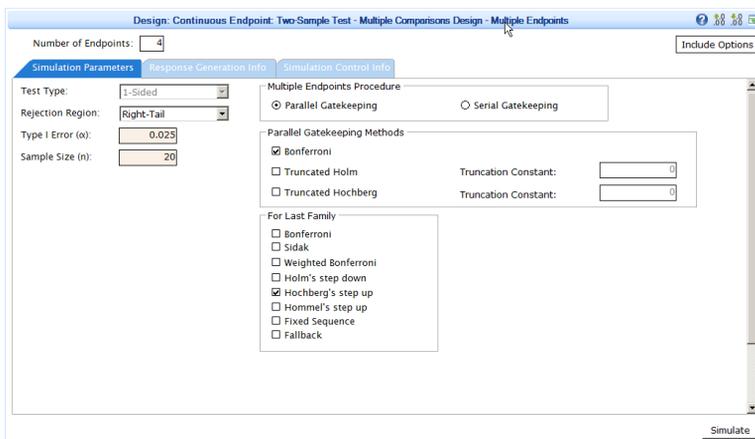
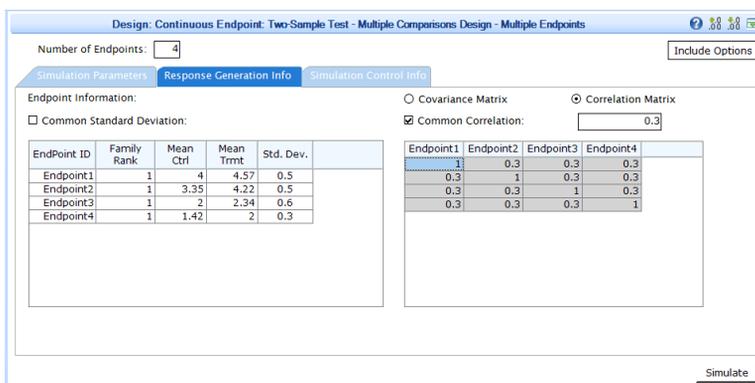
Table 16.5: Impact of Truncation Constant in Truncated Hochberg Procedure on Overall Power and Power for Each Family

| Truncation Constant | Primary Family | | Secondary Family | | Overall Power | |
|---------------------|----------------|-----------|------------------|-----------|---------------|-----------|
| | Conjunct. | Disjunct. | Conjunct. | Disjunct. | Conjunct. | Disjunct. |
| 0 | 0.5219 | 0.9449 | 0.1253 | 0.8955 | 0.1012 | 0.9449 |
| 0.25 | 0.5652 | 0.9455 | 0.1229 | 0.8877 | 0.1065 | 0.9455 |
| 0.5 | 0.6007 | 0.9468 | 0.1213 | 0.8764 | 0.1109 | 0.9468 |
| 0.8 | 0.6369 | 0.9491 | 0.119 | 0.8439 | 0.1152 | 0.9491 |

Table 16.6: Impact of Truncation Constant in Truncated Hochberg Procedure on Marginal Power

| Truncation Constant | Primary Family | | Secondary Family | |
|---------------------|----------------|---------|------------------|---------|
| | Type 16 | Type 18 | Type 31 | Type 45 |
| 0 | 0.5522 | 0.9146 | 0.127 | 0.8938 |
| 0.25 | 0.5892 | 0.9215 | 0.1246 | 0.886 |
| 0.5 | 0.6203 | 0.9273 | 0.1228 | 0.8749 |
| 0.8 | 0.6525 | 0.9335 | 0.1205 | 0.8424 |

down, Hochberg’s step up, Hommel’s step up, Fixed Sequence or Fallback) will be applied for multiplicity adjustment. For example, if all the four antigen types in the vaccine example are treated as primary endpoints as indicated by the family rank in the window for Response Generation Info and Hochberg’s step up test is selected for the last family in the window for Simulation Parameters, then the regular Hochberg test will be applied to the four endpoints for multiplicity adjustment. The detailed output window is slightly different in case of single family of endpoints as seen in the following window.



16 Multiple Endpoints-Gatekeeping Procedures

Simulation: Continuous Endpoint: Two-Sample Test - Multiple Comparisons Design - Multiple Endpoints - Parallel Gatekeeping

Hypothesis:
 $H_1: \mu_{i1} - \mu_{i0} < 0$ Vs $K_i: \mu_{i1} - \mu_{i0} > 0$ for $i = 1, 2, \dots, n$ where n is the total number of endpoints.

| Simulation Parameters | |
|-----------------------------------|------------|
| Simulation ID | Sim13 |
| Test Type | 1-Sided |
| Rejection Region | Right-Tail |
| Nominal Type-I Error (α) | 0.025 |
| Sample Size (n) | 20 |

| Multiple Endpoint Procedure | |
|------------------------------|--------------------|
| Total Number of Endpoints | 4 |
| Total Number of Families | 1 |
| Common Correlation | 0.3 |
| Multiple Endpoints Procedure | Hochberg's step up |

| Simulation Control Parameters | |
|-------------------------------|--------|
| Starting Seed | Fixed |
| Number of Simulations | 100000 |

Overall Powers

| | |
|---|--------|
| Global (Reject any H) | 0.9842 |
| Conjunctive (Reject all H where $\mu_i > \mu_{ic}$) | 0.1763 |
| Disjunctive (Reject at least one H where $\mu_i > \mu_{ic}$) | 0.9842 |
| FWER (Reject any H where $\mu_i \leq \mu_{ic}$) | 0 |

Output for Multiple Primary Endpoints

| Family Rank | Conjunctive Power | Disjunctive Power | FWER |
|-------------|-------------------|-------------------|------|
| 1 | 0.1763 | 0.9842 | 0 |

Marginal Powers for Endpoints with $\mu_i > \mu_{ic}$

| Endpoint ID | Marginal Power |
|-------------|----------------|
| Endpoint1 | 0.5687 |
| Endpoint2 | 0.8962 |
| Endpoint3 | 0.204 |
| Endpoint4 | 0.949 |

Response Parameters

| Endpoint ID | Family Rank | Mean Control | Mean Treatment | Standard Deviation |
|-------------|-------------|--------------|----------------|--------------------|
| Endpoint1 | 1 | 4 | 4.57 | 0.5 |
| Endpoint2 | 1 | 3.35 | 4.22 | 0.5 |
| Endpoint3 | 1 | 2 | 2.34 | 0.6 |
| Endpoint4 | 1 | 1.42 | 2 | 0.3 |

Correlation Matrix

| | Endpoint1 | Endpoint2 | Endpoint3 | Endpoint4 |
|-----------|-----------|-----------|-----------|-----------|
| Endpoint1 | 1 | 0.3 | 0.3 | 0.3 |
| Endpoint2 | 0.3 | 1 | 0.3 | 0.3 |
| Endpoint3 | 0.3 | 0.3 | 1 | 0.3 |
| Endpoint4 | 0.3 | 0.3 | 0.3 | 1 |

Simulation Seed and Elapsed Time

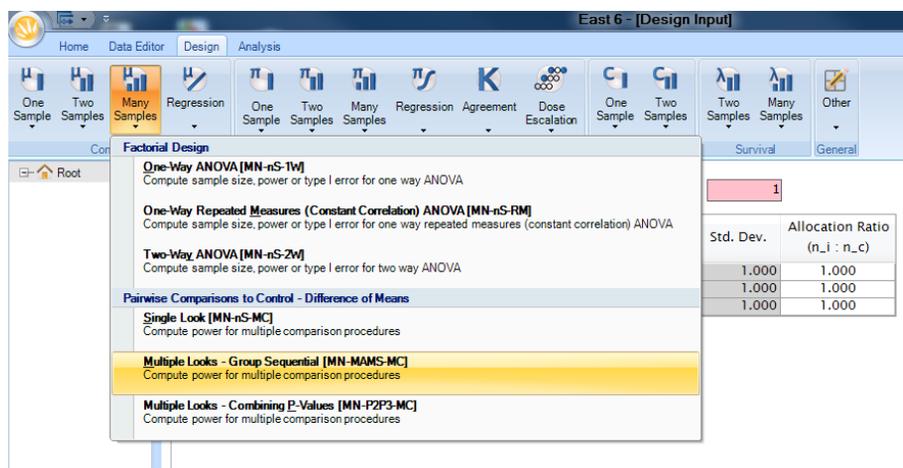
Starting Seed: 100
 Total Number of Simulations: 100000
 Planned Time: 00:00:09

17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs

17.1 Design

Consider designing a placebo controlled, double blind and randomized trial to evaluate the efficacy, pharmacokinetics, safety and tolerability of a new therapy given as multiple weekly infusions in subjects with a recent acute coronary syndrome. There are four dose regimens to be investigated. The treatment effect is assessed through the change in PAV (percent atheroma volume) from baseline to Day 36 post-randomization, as determined by IVUS (intravascular ultrasound). The expected change in PAV for placebo group and the four dose regimens are: 0, 1, 1.1, 1.2 and 1.3 and the common standard deviation is 3. The objective of the study is to find the optimal dose regimen based on the totality of the evidence including benefit-risk assessment and cost considerations.

To design such a study in EAST, we first need to invoke the design dialog window. To this end, one needs to click on the Design menu on the top of EAST window, select Many Samples for continuous type of response and then select Multiple Looks-Group Sequential in the drop-down list as shown in the following screen shot



After selecting the design, we will see a dialog window for the user to specify the main design parameters. On the top of the window, we need to specify the number of arms including the control arm and the number of looks. We also need to specify the

17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs

nominal significance level, power or sample size, mean response for each arm, standard deviation for each arm and allocation ratio of each arm to control arm. Suppose we would like to compute the sample size to achieve 90% power at one-sided 0.025 significance level. After filling in all the inputs, the design dialog window looks as follows:

Design: Continuous Endpoint: Many-Sample Test - Pairwise Comparisons to Control - Difference of Means - Multi-

Number of Arms: 5 Number of Looks: 1

Test Parameters

Type I Error (α): 0.025 Common Standard Deviation: 3

Power: 0.9 ○

Sample Size (n): Computed ⊕

| Arm | Mean | Std. Dev. | Allocation Ratio ($n_i : n_c$) |
|-------------|------|-----------|-------------------------------------|
| Control | 0 | 3.000 | 1.000 |
| Treatment 1 | 1 | 3.000 | 1.000 |
| Treatment 2 | 1.1 | 3.000 | 1.000 |
| Treatment 3 | 1.2 | 3.000 | 1.000 |
| Treatment 4 | 1.3 | 3.000 | 1.000 |

Compute

Now click on the compute button at the bottom right of the window to see the total sample size. Note that we need 519 subjects. Here the power is the probability of successfully detecting significant difference for at least one active treatment group

compared to control arm.

Design: Continuous Endpoint: Many-Sample Test - Pairwise Comparisons to Control - Difference of Means - Multiple Look-Group Sequential

Number of Arms: 5 Number of Looks: 1

Test Parameters

Type I Error (α): 0.025 Common Standard Deviation: 3

Power: 0.9 ○

Sample Size (n): 519 ⊕

| Arm | Mean | Std. Dev. | Allocation Ratio (n _i : n _c) |
|-------------|------|-----------|---|
| Control | 0 | 3.000 | 1.000 |
| Treatment 1 | 1 | 3.000 | 1.000 |
| Treatment 2 | 1.1 | 3.000 | 1.000 |
| Treatment 3 | 1.2 | 3.000 | 1.000 |
| Treatment 4 | 1.3 | 3.000 | 1.000 |

Suppose that now we would like to do a group sequential design with interim looks so that the trial can be terminated earlier if one or more of the treatment groups demonstrate overwhelming efficacy. To do this, we change the number of looks to 3. Note that there is another tab showing up beside the Test Parameter tab. This new tab with label Boundary is to specify efficacy boundary, futility boundary and the spacing of looks. Suppose we want to take two interim looks with equally spacing using O'Brien Fleming spending function from Lan-DeMats 1984. The input window looks

17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs

like the following

Design: Continuous Endpoint: Many-Sample Test - Pairwise Comparisons to Control - Difference of Means - Multip...   

Number of Arms: Number of Looks:

Test Parameters Boundary

Efficacy

Boundary Family: Futility (Non-Binding)

Spending Function: Boundary Family:

Parameter:

Type I Error (α): 0.025

Spacing of Looks Equal Unequal Efficacy Boundary:  

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.333 | 0.000 | 4.049 |
| 2 | 0.667 | 0.006 | 2.931 |
| 3 | 1.000 | 0.025 | 2.477 |

One can view the boundaries in terms of other scales including score, δ and p-value scale by clicking the drop-down box for boundary scale. For example, the δ scale boundary for this study is 2.904, 1.486 and 1.026.

Now click on the compute button on the bottom right of the window to create the design. Note that the total sample size to achieve 90% power is now 525 compared to 519 for the fixed sample design created earlier. The power definition here is the probability of successfully detecting any active treatment group which is significantly

different from control group at any look.

Design: Continuous Endpoint: Many-Sample Test - Pairwise Comparisons to Control - Difference of Means - Multiple Look-Group

5 3

Test Parameters Boundary

Type I Error (α): 0.025 Common Standard Deviation: 3

Power: 0.9 ○

Sample Size (n): 525 ○

| Arm | Mean | Std. Dev. | Allocation Ratio (n _i : n _c) |
|-------------|------|-----------|---|
| Control | 0 | 3.000 | 1.000 |
| Treatment 1 | 1 | 3.000 | 1.000 |
| Treatment 2 | 1.1 | 3.000 | 1.000 |
| Treatment 3 | 1.2 | 3.000 | 1.000 |
| Treatment 4 | 1.3 | 3.000 | 1.000 |

To view the detailed design output, keep the design in the library and double click the design node.

Design: Continuous Endpoint: Many-Sample Test - Pairwise Comparisons to Control - Difference of Means - Multiple Look-Group Sequential

| Test Parameters | |
|---------------------|-----------------|
| Design ID | DesignThreeLook |
| Number of Looks | 3 |
| Specified α | 0.025 |
| Power | 0.9 |
| Number of Arms | 5 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (3F) |

Sample Size Information

| | Control | Treatment 1 | Treatment 2 | Treatment 3 | Treatment 4 | Total |
|-------------------------|---------|-------------|-------------|-------------|-------------|---------|
| Sample Size (n) | | | | | | |
| Maximum | 105 | 105 | 105 | 105 | 105 | 525 |
| Expected H1 | 84.876 | 84.876 | 84.876 | 84.876 | 84.876 | 424.378 |
| Expected H0 | 104.771 | 104.771 | 104.771 | 104.771 | 104.771 | 523.856 |
| Maximum Information (I) | | | | | | 5.833 |

Note - H0 is Global Null and H1 is Global Alternative

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n _{max}) | Sample Size (n) | Cumulative α Spent | Boundaries Efficacy Score | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------------------|-----------------|---------------------------|---------------------------|---|----------|
| | | | | | Under H0 | Under H1 |
| | | | | | Efficacy | Efficacy |
| 1 | 0.333 | 175 | 1.035E-4 | 5.646 | 1.027E-4 | 0.027 |
| 2 | 0.667 | 350 | 0.006 | 5.78 | 0.006 | 0.521 |
| 3 | 1 | 525 | 0.025 | 5.983 | 0.019 | 0.352 |

Stopping Boundaries (Z Scale): Look by Look

| Look # | Info. Fraction (n/n _{max}) | Cumulative α Spent | Boundaries Z Scale |
|--------|--------------------------------------|---------------------------|--------------------|
| | | | Efficacy |
| 1 | 0.333 | 1.035E-4 | 4.049 |
| 2 | 0.667 | 0.006 | 2.931 |
| 3 | 1 | 0.025 | 2.477 |

Treatment Parameters

| Arm | Mean | Standard Deviation | Allocation Ratio (n _i : n _c) |
|-------------|------|--------------------|---|
| Control | 0 | 3 | 1 |
| Treatment 1 | 1 | 3 | 1 |
| Treatment 2 | 1.1 | 3 | 1 |
| Treatment 3 | 1.2 | 3 | 1 |
| Treatment 4 | 1.3 | 3 | 1 |

The first table shows the sample size information including the maximum sample size

17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs

if the trial goes all the way to the end and the sample size per arm. It also shows that the expected sample size under the global null where none of the active treatment group is different from control group and the expected sample size under the design alternative specified by the user. The secondary table displays the look-by-look information including sample size, cumulative type I error, boundaries, boundary crossing probability under the global null and under user-specified design alternative. The boundary crossing probability at each look shows the likelihood of at least one active treatment group crossing the boundary at that particular look. The third table shows the Z scale boundary.

One can also add a futility boundary to the design by clicking on the drop-down box for the futility boundary family. There are three families of boundary for futility: Spending Function, p value, δ which can be seen as in the following screen

Design: Continuous Endpoint: Many-Sample Test - Pairwise Comparisons to Control - Differen

Number of Arms: 5 Number of Looks: 3

Test Parameters **Boundary**

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.025

Futility (Non-Binding)

Boundary Family: Spending Functions

Spending Function: None

Parameter: p-value

Type II Error (β): 0.1

Spacing of Looks: Equal Unequal Boundary Scale: Z Scale

| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Cum. β Spent | Futility Boundary |
|--------|----------------|-------------------------------------|-------------------------------------|---------------------|-------------------|--------------------|-------------------|
| 1 | 0.333 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.000 | 4.049 | 0.005 | 0.178 |
| 2 | 0.667 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.006 | 2.931 | 0.045 | 1.642 |
| 3 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.025 | 2.477 | 0.102 | 2.477 |

Now click on recalc button to see the cumulative α , efficacy boundary, cumulative β and futility boundary displayed in the boundary table. The futility boundary is non-binding and the details on the computation of futility boundary is provided in Section J.2. The futility boundary is computed such that the probability for the best performed arm (compared to control arm) to cross the futility boundary at any look is equal to the incremental β . For example, the probability for the best performed treatment arm crossing 0.178 is 0.005 under the design alternative. The probability for the trial to stay in the continuous region at the first look but cross the futility boundary

1.647 at second look is 0.04 which is the incremental β spent.

Number of Arms: 5 Number of Looks: 3

Test Parameters Boundary

Efficacy
 Boundary Family: Spending Functions
 Spending Function: Lan-DeMets
 Parameter: OF
 Type I Error (α): 0.025

Futility (Non-Binding)
 Boundary Family: Spending Functions
 Spending Function: Lan-DeMets
 Parameter: OF
 Type II Error (β): 0.1

Spacing of Looks Boundary Scale: Z Scale

Equal Unequal

| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Cum. β Spent | Futility Boundary |
|--------|----------------|-------------------------------------|-------------------------------------|---------------------|-------------------|--------------------|-------------------|
| 1 | 0.333 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.000 | 4.062 | 0.004 | 0.144 |
| 2 | 0.667 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.006 | 2.925 | 0.044 | 1.647 |
| 3 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.025 | 2.478 | 0.099 | 2.478 |

Now click on Compute to see the required sample size to achieve 90% power. Note that we need a larger sample size 560 to achieve the same target power with futility boundary compared to the design without futility boundary. However, the expected sample size under H_0 with futility boundary is much smaller than the design without

17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs

futility.

Design: Continuous Endpoint: Many-Sample Test - Pairwise Comparisons to Control - Difference of Means - Multiple Look-Group Sequential

| Test Parameters | |
|---------------------|-------|
| Design ID | Des8 |
| Number of Looks | 3 |
| Specified α | 0.025 |
| Attained α | 0.022 |
| Power | 0.9 |
| Number of Arms | 5 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD |
| Futility Boundary | (OF) |
| | LD |
| | (OF) |
| | (NB) |

Note: Futility Boundary (Non-Binding)

Sample Size Information

| | Control | Treatment 1 | Treatment 2 | Treatment 3 | Treatment 4 | Total |
|-------------------------|---------|-------------|-------------|-------------|-------------|---------|
| Sample Size (n) | | | | | | |
| Maximum | 112 | 112 | 112 | 112 | 112 | 560 |
| Expected H1 | 88.675 | 88.675 | 88.675 | 88.675 | 88.675 | 443.373 |
| Expected H0 | 70.516 | 70.516 | 70.516 | 70.516 | 70.516 | 352.58 |
| Maximum Information (I) | | | | | | 6.222 |

Note - H0 is Global Null and H1 is Global Alternative.

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Cumulative β Spent | Boundaries | | Boundary Crossing Probability (Cumulative) | | | |
|--------|--------------------------|-----------------|---------------------------|--------------------------|----------------|----------------|--|----------|----------|----------|
| | | | | | Efficacy Score | Futility Score | Under H0 | | Under H1 | |
| | | | | | | | Efficacy | Futility | Efficacy | Futility |
| 1 | 0.33 | 187 | 9.632E-5 | 0.004 | 5.824 | 0.207 | 9.731E-5 | 0.251 | 0.029 | 0.004 |
| 2 | 0.67 | 373 | 0.006 | 0.044 | 5.971 | 3.363 | 0.006 | 0.858 | 0.586 | 0.044 |
| 3 | 1 | 560 | 0.025 | 0.099 | 6.18 | 6.18 | 0.022 | 0.978 | 0.901 | 0.099 |

Stopping Boundaries (Z Scale): Look by Look

| Look # | Info. Fraction (n/n_max) | Cumulative α Spent | Cumulative β Spent | Boundaries Z Scale | |
|--------|--------------------------|---------------------------|--------------------------|--------------------|----------|
| | | | | Efficacy | Futility |
| 1 | 0.33 | 9.632E-5 | 0.004 | 4.062 | 0.144 |
| 2 | 0.67 | 0.006 | 0.044 | 2.925 | 1.647 |
| 3 | 1 | 0.025 | 0.099 | 2.478 | 2.478 |

Treatment Parameters

| Arm | Mean | Standard Deviation | Allocation Ratio (n ₁ : n ₂ : n _c) |
|-------------|------|--------------------|--|
| Control | 0 | 3 | 1 |
| Treatment 1 | 1 | 3 | 1 |
| Treatment 2 | 1.1 | 3 | 1 |
| Treatment 3 | 1.2 | 3 | 1 |
| Treatment 4 | 1.3 | 3 | 1 |

One can also build a futility boundary based on δ . For example, one might want to terminate the study if negative δ is observed. It can be seen that such futility boundary is more conservative than the one constructed based on O'Brien-Fleming spending function in the sense that it terminates the trial earlier for futility with smaller

probability.

Design: Continuous Endpoint: Many-Sample Test - Pairwise Comparisons to Control - Difference of Means - Multiple Look-Group Sequential

| Test Parameters | |
|---------------------|-------------------------|
| Design ID | DesignThreeLookFutDelta |
| Number of Looks | 3 |
| Specified α | 0.025 |
| Attained α | 0.025 |
| Power | 0.9 |
| Number of Arms | 5 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Futility Boundary | δ (NB) |

Note: Futility Boundary (Non-Binding)

Sample Size Information

| | Control | Treatment 1 | Treatment 2 | Treatment 3 | Treatment 4 | Total |
|-------------------------|---------|-------------|-------------|-------------|-------------|---------|
| Sample Size (n) | | | | | | |
| Maximum | 106 | 106 | 106 | 106 | 106 | 530 |
| Expected H1 | 86.392 | 86.392 | 86.392 | 86.392 | 86.392 | 431.959 |
| Expected H0 | 88.31 | 88.31 | 88.31 | 88.31 | 88.31 | 441.549 |
| Maximum Information (I) | | | | | | 5.889 |

Note - H0 is Global Null and H1 is Global Alternative.

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Cumulative β Spent | Boundaries | | Boundary Crossing Probability (Cumulative) | | | |
|--------|--------------------------|-----------------|---------------------------|--------------------------|------------|-------------------|--|----------|----------|----------|
| | | | | | Score | Futility δ | Under H0 | | Under H1 | |
| | | | | | | | Efficacy | Futility | Efficacy | Futility |
| 1 | 0.33 | 177 | 9.592E-5 | 0.003 | 5.665 | 0 | 9.698E-5 | 0.2 | 0.026 | 0.003 |
| 2 | 0.67 | 353 | 0.006 | 0.003 | 5.808 | 0 | 0.006 | 0.294 | 0.556 | 0.003 |
| 3 | 1 | 530 | 0.025 | 0.1 | 6.013 | 1.021 | 0.025 | 0.975 | 0.9 | 0.1 |

Stopping Boundaries (Z Scale): Look by Look

| Look # | Info. Fraction (n/n_max) | Cumulative α Spent | Cumulative β Spent | Boundaries Z Scale | |
|--------|--------------------------|---------------------------|--------------------------|--------------------|----------|
| | | | | Efficacy | Futility |
| 1 | 0.33 | 9.592E-5 | 0.003 | 4.063 | 0 |
| 2 | 0.67 | 0.006 | 0.003 | 2.925 | 0 |
| 3 | 1 | 0.025 | 0.1 | 2.478 | 2.478 |

Treatment Parameters

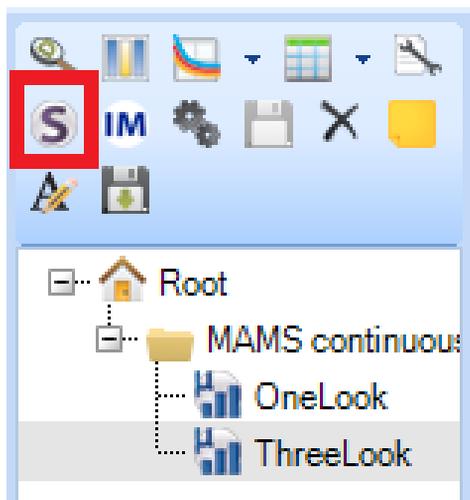
| Arm | Mean | Standard Deviation | Allocation Ratio (n1 : n_c) |
|-------------|------|--------------------|-----------------------------|
| Control | 0 | 3 | 1 |
| Treatment 1 | 1 | 3 | 1 |
| Treatment 2 | 1.1 | 3 | 1 |
| Treatment 3 | 1.2 | 3 | 1 |
| Treatment 4 | 1.3 | 3 | 1 |

17.2 Simulation

Multi-arm multi-stage design is complex study design with pros and cons. One of the pros is that it saves subjects compared to conducting separate studies to assess each treatment to control. It may also be advantageous in terms of enrolment. One of the cons is that the hurdle for demonstrating statistical significance is higher due to multiplicity. One needs to evaluate the operating characteristics of such designs through intensive simulations and to assess the pros and cons of using such design. To simulate a MAMS design, select the design node in the library and click on the

17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs

simulation icon located at the top of the library window



This will open the simulation dialog window. There are four windows for inputting values for simulation parameters: Test Parameters, Boundary, Response Generation and Simulation Controls. The Test Parameters window provides the total sample size, test statistics and variance type to be used in simulations. The boundary tab has similar inputs as that for design. The default inputs for boundary are carried from the design. One can modify the boundary in the simulation mode without having to go back to design. One can even add a futility boundary. The next screen is Response Generation tab where one needs to specify the underlying mean, standard deviation and allocation ratio for different treatment arm. The last tab, Simulation Control, allows one to specify the total number of simulations to be run and to save the intermediate simulation data for further analysis. For example, we can run simulation under the

design alternative where the mean differences from control are 1,1.1,1.2 and 1.3.

Trt Select **Simulation: Continuous Endpoint: Many-Sample Test -**

Number of Arms: 5 Number of Looks: 3

Test Parameters Boundary Response Generation Simulation Controls

Common Standard Deviation: 3

| Arm | Mean | Std. Dev. | Allocation Ratio (n _i : n _c) |
|-------------|------|-----------|---|
| Control | 0 | 3.000 | 1.000 |
| Treatment 1 | 1 | 3.000 | 1.000 |
| Treatment 2 | 1.1 | 3.000 | 1.000 |
| Treatment 3 | 1.2 | 3.000 | 1.000 |
| Treatment 4 | 1.3 | 3.000 | 1.000 |

After filling in all the inputs, click on the Simulation button on the right bottom of the window. After the simulation is completed, it will show up in the output preview area. To view the detailed simulation output, we can save it into the library and double click the simulation node.

Simulation: Continuous Endpoint: Many-Sample Test - Pairwise Comparisons to Control - Difference of Means - Multiple Look-Group Sequential

| Test Parameters | |
|-----------------|-------|
| Simulation ID | Sim26 |
| Number of Looks | 3 |
| Number of Arms | 5 |
| Sample Size (n) | 525 |
| Test Statistic | Z |

| Boundary Parameters | |
|---------------------|---------|
| Efficacy Boundary | LD (OF) |
| Spacing of Looks | Equal |

| Simulation Control Parameters | |
|-------------------------------|---------|
| Starting Seed | Clock |
| Number of Simulations | 1000000 |

Overall Powers

| | |
|---|--------|
| Global (Reject any H ₀) | 0.8995 |
| Conjunctive (Reject all H ₀ where $\mu > \mu_0$) | 0.0047 |
| Disjunctive (Reject at least one H ₀ where $\mu > \mu_0$) | 0.8995 |
| FWER (Reject any H ₀ where $\mu_i \leq \mu_0$) | 0 |

Probability of Trial Termination at Each Look

| Look # | Average Sample Size (n) | Info. Fraction (info_max) | Cumulative α Spent | Boundaries | | Probability (Cumulative) of Trial Termination | | | |
|---------|-------------------------|---------------------------|---------------------------|------------|----------|---|---------|---------|-------|
| | | | | Z | Efficacy | Efficacy | | | Total |
| | | | | | | Count | % | Count | |
| 1 | 175 | 0.333 | 1.035E-4 | 4.049 | 26855 | 2.686% | 26855 | 2.686% | |
| 2 | 350 | 0.667 | 0.006 | 2.931 | 547564 | 54.756% | 547564 | 54.756% | |
| 3 | 525 | 1 | 0.025 | 2.477 | 899496 | 89.950% | 1000000 | 89.950% | |
| Overall | 424.477 | | | | 899496 | 89.950% | 1000000 | | |

Note: The trial is terminated if any arm enters the efficacy region.

Additional Details of Probability of Trial Termination at Each Look

| Look # | Probability (Cumulative) Trial Terminates for Efficacy | | | | Trial Continues | | | |
|--------|--|---------|---------|--------|----------------------------|--------|--------|---------|
| | # of Treatments Claiming Efficacy | | | | # of Treatments Continuing | | | |
| | One | Two | Three | Four | One | Two | Three | Four |
| 1 | 2.330% | 0.259% | 0.049% | 0.008% | 0.000% | 0.000% | 0.000% | 97.315% |
| 2 | 26.986% | 15.146% | 8.717% | 4.007% | 0.000% | 0.000% | 0.000% | 45.244% |
| 3 | 38.560% | 25.509% | 17.006% | 8.475% | | | | |
| Total | 38.560% | 25.509% | 17.006% | 8.475% | | | | |

Note: The trial is terminated if any arm enters the efficacy region.

The first table in the detailed output shows the overall power including global power, conjunctive power, disjunctive power and FWER. The definitions for different powers

17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs

are as follows.

- Global Power: probability of demonstrating statistical significance on one or more treatment groups
- Conjunctive Power: probability of demonstrating statistical significance on all treatment groups which are truly effective
- Disjunctive Power: probability of demonstrating statistical significance on at least one treatment group which is truly effective
- FWER: probability of incorrectly demonstrating statistical significance on at least one treatment group which is truly ineffective

For this example, the global power is about 90% which confirms the design power. The conjunctive power is about 8%.

The second table for probability of trial termination at each look displays the average sample size, information fraction, cumulative α spent, boundary information, probability of trial termination at each look. For this example, the chance of terminating the trial at the very first look is less than 3%. The trial has about 55% chance to stop early by the second look. It can be seen that the average sample size for the trial is about 424 which is shown in the last entry of the average sample size column.

In MAMS design, when the trial stops for efficacy, there might be one or more treatments crossing the efficacy boundary. Such information is valuable in some situations. For example, when multiple dose options are desired for patients with different demographic characteristics, it might be beneficial to approve multiple doses on the product label which will give physicians the options to prescribe the appropriate dose for a specific patient. In this case, we are not only interested in the overall power of the study but also interested in the power of claiming efficacy on more than one dose groups. Such information is summarized in the third table. This table shows the probability of demonstrating significance on specific number of treatments at each look and across all looks. For example, the trial has about 90% overall power. With 39% probability out of 90%, it successfully shows significance on only one treatment, 26% probability on two treatments, 17% on three treatments and about 8.5% for all four treatments. It also shows such breakdown look by look.

The fourth table summarizes the marginal power for each treatment group look by look and across all looks. For example, the trial has a marginal power of 29% successfully demonstrating efficacy for Treatment 1, 38% for Treatment 2, 49% for Treatment 3 and 60% for Treatment 4. The detailed efficacy outcome table as seen in the following screen provides further efficacy details pertinent to treatment identities. For example,

the trial has about 3.77% probability of demonstrating efficacy only on Treatment 1, 1.34% for both treatment 1 and 2, 1.7% for treatment 1, 2 and 3. It has 8.5% probability of showing significance on all four treatments.

⊖ Detailed Efficacy Outcomes

| Efficacy Outcome | Count | % |
|--|---------|----------|
| Treatment 1 only | 37715 | 3.772% |
| Treatment 2 only | 64134 | 6.413% |
| Treatment 3 only | 107546 | 10.755% |
| Treatment 4 only | 176206 | 17.621% |
| Treatment 1 & Treatment 2 only | 13397 | 1.340% |
| Treatment 1 & Treatment 3 only | 21936 | 2.194% |
| Treatment 1 & Treatment 4 only | 35303 | 3.530% |
| Treatment 2 & Treatment 3 only | 35945 | 3.595% |
| Treatment 2 & Treatment 4 only | 58110 | 5.811% |
| Treatment 3 & Treatment 4 only | 94395 | 9.440% |
| Treatment 1 & Treatment 2 & Treatment 3 only | 17644 | 1.764% |
| Treatment 1 & Treatment 2 & Treatment 4 only | 28899 | 2.890% |
| Treatment 1 & Treatment 3 & Treatment 4 only | 46918 | 4.692% |
| Treatment 2 & Treatment 3 & Treatment 4 only | 76601 | 7.660% |
| All 4 treatments | 84747 | 8.475% |
| None | 100504 | 10.050% |
| Total | 1000000 | 100.000% |

17.2.1 Futility Stopping and Dropping the Losers

In the simulation mode, the futility boundary can be utilized in two different manners. Futility boundary can be used to terminate the trial earlier if the best performing treatment isn't doing well. It can also be used to drop arms which are futile along the way and only continue those treatments which are performing well. The two options can be accessed through the two radio buttons below the boundary table as seen in the

17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs

following screen.

Number of Arms: 5 Number of Looks: 3

Test Parameters Boundary Response Generation Simulation Controls

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.025

Futility (Non-Binding)

Boundary Family: δ

Spacing of Looks Efficacy Boundary: Z Scale

Equal Unequal

| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Fut. δ |
|--------|----------------|-------------------------------------|-------------------------------------|---------------------|-------------------|---------------|
| 1 | 0.333 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.000 | 4.049 | |
| 2 | 0.667 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.006 | 2.931 | |
| 3 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.025 | 2.477 | |

Terminate Trial if at least one treatment crosses efficacy boundary

Futility Options: Terminate Trial if all treatments cross futility boundary
 Drop treatment arm if it crosses futility boundary

Restore Original Design

Suppose that we would like to incorporate a conservative futility boundary so that we will terminate the trial if all δ s are negative at any interim look. We would specify the

futility boundary as in the following screen.

Number of Arms: 5 Number of Looks: 3

Test Parameters Boundary Response Generation Simulation Controls

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.025

Futility (Non-Binding)

Boundary Family: δ

Spacing of Looks Equal Unequal Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Fut. δ |
|--------|----------------|-------------------------------------|-------------------------------------|---------------------|-------------------|---------------|
| 1 | 0.333 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.000 | 4.049 | 0 |
| 2 | 0.667 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.006 | 2.931 | 0 |
| 3 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.025 | 2.477 | 1.026 |

Terminate Trial if at least one treatment crosses efficacy boundary

Futility Options: Terminate Trial if all treatments cross futility boundary
 Drop treatment arm if it crosses futility boundary

Restore Original Design

Suppose we want to see how often the trial will be terminated early for futility if none of the treatments are effective. Click on the Simulate button on the right bottom of the window to start simulation. The detailed output is shown below. Note that the trial will have about 20% probability of stopping early for futility at the very first look and a little more than 9% chance of stopping for futility at the second look. The average

17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs

sample size is about 437 compared to 523 for the design without futility boundary.

Simulation: Continuous Endpoint: Many-Sample Test - Pairwise Comparisons to Control - Difference of Means - Multiple Look-Group Sequential

| Test Parameters | |
|-----------------|-------|
| Simulation ID | Sim12 |
| Number of Looks | 3 |
| Number of Arms | 5 |
| Sample Size (n) | 525 |
| Variance | Equal |
| Test Statistic | t |

| Boundary Parameters | |
|---------------------|---------|
| Efficacy Boundary | LD (OF) |
| Futility Boundary | δ (NB) |
| Spacing of Looks | Equal |

| Simulation Control Parameters | |
|-------------------------------|---------|
| Starting Seed | Clock |
| Number of Simulations | 1000000 |

Note: Terminate trial if all treatments cross futility boundary.

| Overall Powers | |
|---|--------|
| Global (Reject any H ₁) | 0.0265 |
| Conjunctive (Reject all H ₁ where $\mu_i > \mu_0$) | 0 |
| Disjunctive (Reject at least one H ₁ where $\mu_i > \mu_0$) | 0 |
| POWER (Reject any H ₁ where $\mu_i \leq \mu_0$) | 0.0265 |

| Probability of Trial Termination at Each Look | | | | | | | | | | | | |
|---|-------------------------|--------------------------------------|--------------------|--------------------|------------|------------|--|----------------|--------|-------------|---------|---------|
| Look # | Average Sample Size (n) | Info. Fraction (n/n _{max}) | Cumulative α Spent | Cumulative β Spent | Boundaries | | Probability (Incremental) of Trial Termination | | | | | |
| | | | | | Efficacy Z | Futility δ | Efficacy Count | Futility Count | % | Total Count | % | |
| 1 | 175 | 0.333 | 1.035E-4 | 0.003 | 4.049 | 0 | 276 | 0.028% | 199670 | 19.965% | 199954 | 19.995% |
| 2 | 350 | 0.667 | 0.006 | 0.003 | 2.931 | 0 | 6920 | 0.693% | 92574 | 9.257% | 99502 | 9.950% |
| 3 | 525 | 1 | 0.025 | 0.103 | 2.477 | 1.026 | 19336 | 1.934% | 681208 | 68.121% | 700544 | 70.054% |
| Overall | 437.603 | | | | | | 26540 | | 973460 | 97.346% | 1000000 | |

Note: The trial is terminated if any arm enters the efficacy region or all arms enter the futility region.

Simulation: Continuous Endpoint: Many-Sample Test - Pairwise Comparisons to Control - Difference of Means - Multiple Look-Group Sequential

| Test Parameters | |
|-----------------|-------|
| Simulation ID | Sim13 |
| Number of Looks | 3 |
| Number of Arms | 5 |
| Sample Size (n) | 525 |
| Variance | Equal |
| Test Statistic | t |

| Boundary Parameters | |
|---------------------|---------|
| Efficacy Boundary | LD (OF) |
| Futility Boundary | δ (NB) |
| Spacing of Looks | Equal |

| Simulation Control Parameters | |
|-------------------------------|---------|
| Starting Seed | Clock |
| Number of Simulations | 1000000 |

Note: Terminate trial if all treatments cross futility boundary.

| Overall Powers | |
|---|--------|
| Global (Reject any H ₁) | 0.0272 |
| Conjunctive (Reject all H ₁ where $\mu_i > \mu_0$) | 0 |
| Disjunctive (Reject at least one H ₁ where $\mu_i > \mu_0$) | 0 |
| POWER (Reject any H ₁ where $\mu_i \leq \mu_0$) | 0.0272 |

| Probability of Trial Termination at Each Look | | | | | | | | | | | | |
|---|-------------------------|--------------------------------------|--------------------|--------------------|------------|------------|--|----------------|--------|-------------|---------|---------|
| Look # | Average Sample Size (n) | Info. Fraction (n/n _{max}) | Cumulative α Spent | Cumulative β Spent | Boundaries | | Probability (Incremental) of Trial Termination | | | | | |
| | | | | | Efficacy Z | Futility δ | Efficacy Count | Futility Count | % | Total Count | % | |
| 1 | 175 | 0.333 | 1.035E-4 | 0 | 4.049 | -100 | 250 | 0.025% | 0 | 0.000% | 250 | 0.025% |
| 2 | 350 | 0.667 | 0.006 | 0.003 | 2.931 | -100 | 7104 | 0.710% | 0 | 0.000% | 7104 | 0.710% |
| 3 | 525 | 1 | 0.025 | 0.102 | 2.477 | 1.026 | 19862 | 1.986% | 972784 | 97.278% | 992646 | 99.265% |
| Overall | 523.859 | | | | | | 27216 | 2.722% | 972784 | 97.278% | 1000000 | |

Note: The trial is terminated if any arm enters the efficacy region or all arms enter the futility region.

Under the design alternative, there is a very small probability (less than 0.5%) to terminate the trial early for futility as seen from the following table.

Probability of Trial Termination at Each Look

| Look # | Average Sample Size (n) | Info. Fraction (n/n _{max}) | Cumulative α Spent | Cumulative β Spent | Boundaries | | Probability (Incremental) of Trial Termination | | | | | |
|---------|-------------------------|--------------------------------------|--------------------|--------------------|------------|------------|--|----------------|--------|-------------|---------|---------|
| | | | | | Efficacy Z | Futility δ | Efficacy Count | Futility Count | % | Total Count | % | |
| 1 | 175 | 0.333 | 1.035E-4 | 0.003 | 4.049 | 0 | 38650 | 3.865% | 2844 | 0.284% | 41494 | 4.149% |
| 2 | 350 | 0.667 | 0.006 | 0.003 | 2.931 | 0 | 516605 | 51.661% | 106 | 0.011% | 516711 | 51.671% |
| 3 | 525 | 1 | 0.025 | 0.103 | 2.477 | 1.026 | 344240 | 34.424% | 97555 | 9.756% | 441795 | 44.180% |
| Overall | 420.053 | | | | | | 899495 | 89.950% | 100505 | 10.051% | 1000000 | |

Note: The trial is terminated if any arm enters the efficacy region or all arms enter the futility region.

For the big companies, a more aggressive futility boundary might be desirable so that trials for treatments with small effect can be terminated early and resources can be deployed to other programs. Suppose that a futility boundary based on $\delta = 0.5$ to be used. Under the global null hypothesis, there is almost 70% chance for the trial to stop early for futility. The average sample size for the study is about 316 compared to 437

for the design with futility based δ of zero.

⊖ Probability of Trial Termination at Each Look

| Look # | Average Sample Size (n) | Info. Fraction (n/n_max) | Cumulative α Spent | Cumulative β Spent | Boundaries | | Probability (Incremental) of Trial Termination | | | | | |
|---------|-------------------------|--------------------------|---------------------------|--------------------------|------------|----------|--|--------|----------|---------|--------|---------|
| | | | | | Efficacy | Futility | Efficacy | | Futility | | Total | |
| | | | | | Z | δ | Count | % | Count | % | Count | % |
| 1 | 175 | 0.333 | 1.035E-4 | 0.027 | 4.049 | 0.5 | 34 | 0.034% | 48645 | 48.645% | 48679 | 48.679% |
| 2 | 350 | 0.667 | 0.006 | 0.03 | 2.931 | 0.5 | 694 | 0.694% | 20992 | 20.992% | 21686 | 21.686% |
| 3 | 525 | 1 | 0.025 | 0.116 | 2.477 | 1.026 | 1805 | 1.805% | 27830 | 27.830% | 29635 | 29.635% |
| Overall | 316.673 | | | | | | 2533 | 2.533% | 97467 | 97.467% | 100000 | |

Note: The trial is terminated if any arm enters the efficacy region or all arms enter the futility region.

⊖ Probability of Trial Termination at Each Look

| Look # | Average Sample Size (n) | Info. Fraction (n/n_max) | Cumulative α Spent | Cumulative β Spent | Boundaries | | Probability (Incremental) of Trial Termination | | | | | |
|---------|-------------------------|--------------------------|---------------------------|--------------------------|------------|----------|--|--------|----------|---------|---------|---------|
| | | | | | Efficacy | Futility | Efficacy | | Futility | | Total | |
| | | | | | Z | δ | Count | % | Count | % | Count | % |
| 1 | 175 | 0.333 | 1.035E-4 | 0.003 | 4.049 | 0 | 276 | 0.028% | 199678 | 19.968% | 199954 | 19.995% |
| 2 | 350 | 0.667 | 0.006 | 0.003 | 2.931 | 0 | 6928 | 0.693% | 92574 | 9.257% | 99502 | 9.950% |
| 3 | 525 | 1 | 0.025 | 0.103 | 2.477 | 1.026 | 19336 | 1.934% | 681208 | 68.121% | 700544 | 70.054% |
| Overall | 437.603 | | | | | | 26540 | 2.654% | 973460 | 97.346% | 1000000 | |

Note: The trial is terminated if any arm enters the efficacy region or all arms enter the futility region.

The other use of the futility boundary is to drop those arms which are ineffective along the way. Such design would be more efficient if it is anticipated that there is a strong heterogeneity among different treatment arms. Suppose that two of the four treatment regimens have relative smaller treatment effect. For example, the mean difference from control might be 0.1, 0.1, 1.2, 1.3. Without applying any futility, the trial has about 85% and average sample size of 437. If we drop those doses which cross the futility boundary based on δ of 0.5, the trial has about 82% power and average sample size 328. From the table for probability of trial termination at each look, we can see that the trial has about 8% chance stopping early at the first interim look of which a little more than 2% for efficacy and about 5% chance for futility. The trial has 46% chance stopping earlier at the second look with about 45% for efficacy and less than 2% for futility. From the table for additional details of probability of trial termination at each look, we can see that the trial has 2.78% chance stopping for efficacy at the first look of which 2.55% probability the trial demonstrates significance on only one treatment. At the second look, the trial has about 45% probability stopping early for efficacy of which 29% probability it demonstrates significance on one treatment, 15% probability on two treatments and less than 1% probability on three or four treatments. This design has marginal power about 50% to detect significance on Treatment 3 and more than 60% probability on Treatment 4. Treatment 1 and Treatment 2 each has 70% chance being terminated at look 1 for futility. The marginal probability for futility stopping for each treatment counts those simulated trials for which the particular treatment crosses the futility boundary but it doesn't counts those trials for which the particular treatment

17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs

falls into the continuous region.

Probability of Trial Termination at Each Look

| Look # | Average Sample Size (n) | Info. Fraction (n/n_max) | Cumulative α Spent | Cumulative β Spent | Boundaries | | Probability (Incremental) of Trial Termination | | | | | | | | | |
|---------|-------------------------|--------------------------|---------------------------|--------------------------|------------|----------|--|---------|-------|---------|----------|---------|-------|---|-------|--|
| | | | | | Z | δ | Efficacy | | | | Futility | | | | Total | |
| | | | | | | | Count | % | Count | % | Count | % | Count | % | | |
| 1 | 175 | 0.333 | 1.035E-4 | 0.027 | 4.049 | 0.5 | 2779 | 2.779% | 5315 | 5.315% | 8094 | 8.094% | | | | |
| 2 | 293.258 | 0.667 | 0.006 | 0.03 | 2.931 | 0.5 | 44767 | 44.767% | 1406 | 1.406% | 46173 | 46.173% | | | | |
| 3 | 381.323 | 1 | 0.025 | 0.116 | 2.477 | 1.026 | 34159 | 34.159% | 11574 | 11.574% | 45733 | 45.733% | | | | |
| Overall | 328.019 | | | | | | 81705 | 81.705% | 18295 | 18.295% | 100000 | | | | | |

Note: The trial is terminated if any arm enters the efficacy region and an arm is dropped when it enters futility region.

Additional Details of Probability of Trial Termination at Each Look

| Look # | Probability (Incremental) Trial Terminates for Efficacy | | | | Probability (Incremental) Trial Terminates for Futility | Trial Continues | | | |
|--------|---|---------|--------|--------|---|----------------------------|---------|---------|---------|
| | # of Treatments Claiming Efficacy | | | | | # of Treatments Continuing | | | |
| | One | Two | Three | Four | | One | Two | Three | Four |
| 1 | 2.552% | 0.218% | 0.009% | 0.000% | 5.315% | 15.558% | 38.785% | 24.754% | 12.809% |
| 2 | 29.107% | 15.259% | 0.385% | 0.015% | 1.406% | 15.590% | 25.556% | 4.119% | 0.468% |
| 3 | 20.127% | 13.891% | 0.136% | 0.005% | 11.574% | | | | |
| Total | 51.786% | 29.368% | 0.530% | 0.021% | 18.295% | | | | |

Note: The trial is terminated if any arm enters the efficacy region and an arm is dropped when it enters futility region.

Marginal Probability of Trial Termination by Treatment at Each Look

| Look # | Treatment 1 | | Treatment 2 | | Treatment 3 | | Treatment 4 | |
|--------|-------------|----------|-------------|----------|-------------|----------|-------------|----------|
| | Efficacy | Futility | Efficacy | Futility | Efficacy | Futility | Efficacy | Futility |
| 1 | 0.013% | 71.088% | 0.012% | 70.983% | 1.246% | 16.517% | 1.744% | 13.122% |
| 2 | 0.264% | 13.289% | 0.291% | 13.362% | 26.803% | 2.921% | 33.486% | 2.035% |
| 3 | 0.108% | 2.942% | 0.094% | 2.942% | 21.710% | 14.305% | 26.425% | 12.405% |
| Total | 0.385% | 87.319% | 0.397% | 87.287% | 49.759% | 33.743% | 61.655% | 27.562% |

Note: The trial is terminated if any arm enters the efficacy region and an arm is dropped when it enters futility region.

The second table in the above screen shows the probability of demonstrating significance on specific number of treatments. However it doesn't provide information on the likelihood of showing efficacy on specific treatment combinations. Such information is provided in the table for detailed efficacy outcomes. For example, the trial has about 20% probability of success with Treatment 3 only, 32% with Treatment

4 only, 30% with both Treatment 3 and Treatment 4.

 Detailed Efficacy Outcomes

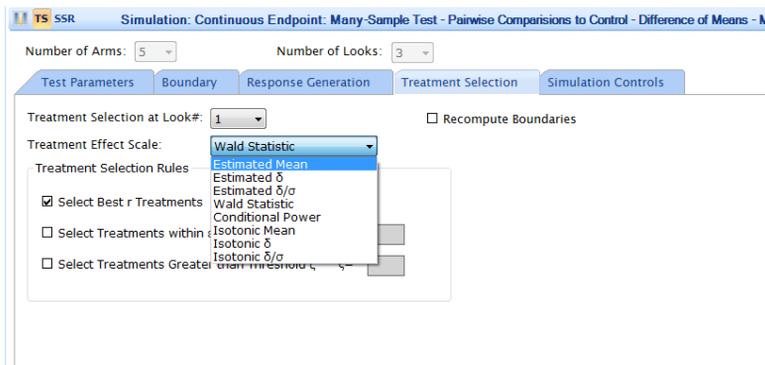
| Efficacy Outcome | Count | % |
|--|--------|----------|
| Treatment 1 only | 16 | 0.016% |
| Treatment 2 only | 15 | 0.015% |
| Treatment 3 only | 19896 | 19.896% |
| Treatment 4 only | 31566 | 31.566% |
| Treatment 1 & Treatment 2 only | 0 | 0.000% |
| Treatment 1 & Treatment 3 only | 29 | 0.029% |
| Treatment 1 & Treatment 4 only | 44 | 0.044% |
| Treatment 2 & Treatment 3 only | 35 | 0.035% |
| Treatment 2 & Treatment 4 only | 49 | 0.049% |
| Treatment 3 & Treatment 4 only | 29354 | 29.354% |
| Treatment 1 & Treatment 2 & Treatment 3 only | 0 | 0.000% |
| Treatment 1 & Treatment 2 & Treatment 4 only | 4 | 0.004% |
| Treatment 1 & Treatment 3 & Treatment 4 only | 239 | 0.239% |
| Treatment 2 & Treatment 3 & Treatment 4 only | 259 | 0.259% |
| All 4 treatments | 27 | 0.027% |
| None | 18467 | 18.467% |
| Total | 100000 | 100.000% |

17.2.2 Interim Treatment Selection

It might be desirable to select promising dose/treatment groups and drop those ineffective or unsafe groups after reviewing the interim data. In general, there are no analytical approach to evaluate such complex design. EAST provides the option to evaluate such adaptive design through intensive simulations. The treatment selection option can be incorporated by clicking on the  icon located on the top bar of the main simulation dialog window. The treatment selection window screen looks as follows. It takes several inputs from the user. The first input is the drop-down box for the user to specify the look position for performing treatment selection. The next input is drop-down box for the treatment effect scale. There is a list of treatment effect scale available as seen in the following screen including Wald Statistic, Estimated Mean,

17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs

Estimated δ etc.



EAST provides three different dose/treatment selection rules: (1) Select best r treatment, (2) Select treatments within ϵ of the best treatment, (3) Select treatments greater than threshold ζ where r, ϵ, ζ accept inputs from the user. For the same example, suppose we select two best treatments at the second interim look. The inputs

are as follows:

Number of Arms: 5 Number of Looks: 3

Test Parameters Boundary Response Generation Simulation Controls

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.025

Futility (Non-Binding)

Boundary Family: None

Spacing of Looks Equal Unequal

Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.333 | 0.000 | 4.049 |
| 2 | 0.667 | 0.006 | 2.931 |
| 3 | 1.000 | 0.025 | 2.477 |

Terminate Trial if at least one treatment crosses efficacy boundary

Restore Original Design

Number of Arms: 5 Number of Looks: 3

Test Parameters Boundary Response Generation Treatment Selection

Common Standard Deviation: 3

| Arm | Mean | Std. Dev. | Allocation Ratio (n _i : n _c) |
|-------------|------|-----------|---|
| Control | 0 | 3.000 | 1.000 |
| Treatment 1 | 0.1 | 3.000 | 1.000 |
| Treatment 2 | 0.1 | 3.000 | 1.000 |
| Treatment 3 | 1.2 | 3.000 | 1.000 |
| Treatment 4 | 1.3 | 3.000 | 1.000 |

17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs

Number of Arms: 5 Number of Looks: 3

Test Parameters Boundary Response Generation **Treatment Selection** Simulation Controls

Treatment Selection at Look#: 2 Reallocate remaining sample size to selected arm

Treatment Effect Scale: Wald Statistic

Treatment Selection Rules

- Select Best r Treatments $r =$ 2
- Select Treatments within ϵ of Best Treatment $\epsilon =$
- Select Treatments Greater than Threshold ζ $\zeta =$

Now click on simulation button to run simulations. When the simulation is done, save it into the library and view the detailed output as in the following screen. We can see that the trial has about 85% overall power to detect significance on at least one treatment group with an average sample size of 400 (Overall Powers). It has about 50% probability of stopping early by the second look (Prabability of Trial Termination at Each Look). From the third table (Additional Details of Probability of Trial Termination at Each Look), it can be seen that the trial has about 52% power to show significance on only one treatment and 33% probability on two treatments, less than 1% probability on three or four treatments. Marginally Treatment 3 has 53% chance of

success and Treatment 4 has 66% chance of success.

Simulation: Continuous Endpoint: Many-Sample Test - Pairwise Comparisons to Control - Difference of Means - Multiple Look-Group Sequential

| Test Parameters | |
|--------------------------------|----------------|
| Simulation ID | Sim30 |
| Number of Looks | 3 |
| Number of Arms | 5 |
| Sample Size (n) | 525 |
| Variance | Equal |
| Test Statistic | t |
| Boundary Parameters | |
| Efficacy Boundary | LD (OF) |
| Spacing of Looks | Equal |
| Treatment Selection Parameters | |
| Treatment Effect Scale | Wald Statistic |
| Treatment Selection Rule | Best 2 |
| Treatment Selection at Look No | 2 |
| Reallocate Sample Size | No |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 100000 |

Overall Powers

| | |
|---|--------|
| Global (Reject any H ₀) | 0.8549 |
| Conjunctive (Reject all H ₀ where $\mu_i > \mu_0$) | 0.0001 |
| Disjunctive (Reject at least one H ₀ where $\mu_i > \mu_0$) | 0.8549 |
| FWER (Reject any H ₀ where $\mu_i \leq \mu_0$) | 0 |

Probability of Trial Termination at Each Look

| Look # | Average Sample Size (n) | Info. Fraction (Info. max) | Cumulative α Spent | Boundaries | | Probability (Cumulative) of Trial Termination | | | |
|---------|-------------------------|----------------------------|---------------------------|------------|-------|---|--------|---------|---|
| | | | | Z Efficacy | Z | Count | % | Count | % |
| 1 | 175 | 0.333 | 1.035E-4 | 4.049 | 2947 | 2.947% | 2947 | 2.947% | |
| 2 | 350 | 0.667 | 0.006 | 2.931 | 47807 | 47.807% | 47807 | 47.807% | |
| 3 | 465 | 1 | 0.025 | 2.477 | 85487 | 85.487% | 100000 | 85.487% | |
| Overall | 399.82 | | | | 85487 | 85.487% | 100000 | | |

Note: The trial is terminated if any arm enters the efficacy region.

Additional Details of Probability of Trial Termination at Each Look

| Look # | Probability (Cumulative) Total Terminates for Efficacy | | | | Trial Continues | | | |
|--------|--|---------|--------|--------|----------------------------|---------|--------|---------|
| | # of Treatments Claiming Efficacy | | | | # of Treatments Continuing | | | |
| | One | Two | Three | Four | One | Two | Three | Four |
| 1 | 2.605% | 0.239% | 0.006% | 0.000% | 0.000% | 0.000% | 0.000% | 97.153% |
| 2 | 31.925% | 15.481% | 0.387% | 0.014% | 0.000% | 52.193% | 0.000% | 0.000% |
| 3 | 51.764% | 33.322% | 0.387% | 0.014% | | | | |
| Total | 51.764% | 33.322% | 0.387% | 0.014% | | | | |

Note: The trial is terminated if any arm enters the efficacy region.

Marginal Probability of Trial Termination by Treatment at Each Look

| Look # | Efficacy Boundary | | | |
|--------|-------------------|-------------|-------------|-------------|
| | Treatment 1 | Treatment 2 | Treatment 3 | Treatment 4 |
| 1 | 0.018% | 0.016% | 1.272% | 1.789% |
| 2 | 0.289% | 0.290% | 28.076% | 35.449% |
| 3 | 0.344% | 0.338% | 53.374% | 65.569% |
| Total | 0.344% | 0.338% | 53.374% | 65.569% |

Note: The trial is terminated if any arm enters the efficacy region.

When we select two best treatments, the sample size for the selected two treatments remains the same as the designed one. However we can reallocate the remaining sample size from the dropped groups to the selected arm to gain more power. If the sample size for the dropped arms are reallocated to the selected arms, the efficacy stopping boundary for the remaining looks will have to be recomputed in order to preserve the type I error. This can be achieved by checking the box for Reallocating remaining sample size to selected arm on the Treatment Selection tab as seen in the

17 Continuous Endpoint: Multi-arm Multi-stage (MaMs) Designs

following window.

Number of Arms: 5 Number of Looks: 3

Test Parameters Boundary Response Generation Treatment Selection Simulation Controls

Treatment Selection at Look#: 2

Treatment Effect Scale: Estimated δ

Reallocate remaining sample size to selected arm

Treatment Selection Rules

Select Best r Treatments $r =$ 2

Select Treatments within ϵ of Best Treatment $\epsilon =$

Select Treatments Greater than Threshold ζ $\zeta =$

The simulation output is shown in the following screen. Note that the power of the study is almost 92% in exchange of a higher average sample size 436 compared to the design without sample size reallocation (85% power and 400 average sample size). Also with sample size reallocation, the study has a higher power 43% of demonstrating significance on both Treatment 3 and Treatment 4 compared to the design without sample size reallocation which has 33% power.

Simulation: Continuous Endpoint: Many-Sample Test - Pairwise Comparisons to Control - Difference of Means - Multiple Look-Group Sequential

| Test Parameters | |
|--------------------------------|--------------------|
| Simulation ID | Sim31 |
| Number of Looks | 3 |
| Number of Arms | 5 |
| Sample Size (n) | 525 |
| Variance | Equal |
| Test Statistic | t |
| Boundary Parameters | |
| Efficacy Boundary | LD (OF) |
| Spacing of Looks | Equal |
| Treatment Selection Parameters | |
| Treatment Effect Scale | Estimated δ |
| Treatment Selection Rule | Best 2 |
| Treatment Selection at Look No | 2 |
| Reallocate Sample Size | Yes |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 100000 |

Overall Powers

| | |
|---|--------|
| Global (Reject any H ₀) | 0.915 |
| Conjunctive (Reject all H ₀ where $\mu_i > \mu_0$) | 0.0001 |
| Disjunctive (Reject at least one H ₀ where $\mu_i > \mu_0$) | 0.915 |
| FWER (Reject any H ₀ where $\mu_i \leq \mu_0$) | 0 |

Probability of Trial Termination at Each Look

| Look # | Average Sample Size (n) | Info. Fraction (n/n _{max}) | Cumulative μ Spent | Probability (Cumulative) of Trial Termination | | | |
|---------|-------------------------|--------------------------------------|------------------------|---|------------|---------|---------|
| | | | | Boundaries Efficacy | Efficacy % | Count | Total % |
| 1 | 175 | 0.333 | 1.035E-4 | 4.049 | 2.911% | 2911 | 2.911% |
| 2 | 350 | 0.667 | 0.006 | 2.931 | 47967 | 47.967% | 47967 |
| 3 | 525 | 1 | 0.025 | 2.477 | 91497 | 91.497% | 100000 |
| Overall | 436.964 | | | | 91497 | 91.497% | 100000 |

Note: The trial is terminated if any arm enters the efficacy region.

Additional Details of Probability of Trial Termination at Each Look

| Look # | Probability (Cumulative) Trial Terminates for Efficacy | | | | Trial Continues | | | |
|--------|--|---------|--------|--------|----------------------------|---------|--------|---------|
| | # of Treatments Claiming Efficacy | | | | # of Treatments Continuing | | | |
| | One | Two | Three | Four | One | Two | Three | Four |
| 1 | 2.660% | 0.246% | 0.005% | 0.000% | 0.000% | 0.000% | 0.000% | 97.089% |
| 2 | 31.978% | 15.578% | 0.397% | 0.014% | 0.000% | 52.033% | 0.000% | 0.000% |
| 3 | 47.830% | 43.256% | 0.397% | 0.014% | | | | |
| Total | 47.830% | 43.256% | 0.397% | 0.014% | | | | |

Note: The trial is terminated if any arm enters the efficacy region.

Marginal Probability of Trial Termination by Treatment at Each Look

| Look # | Efficacy Boundary | | | |
|--------|-------------------|-------------|-------------|-------------|
| | Treatment 1 | Treatment 2 | Treatment 3 | Treatment 4 |
| 1 | 0.009% | 0.013% | 1.294% | 1.851% |
| 2 | 0.299% | 0.297% | 28.245% | 35.549% |
| 3 | 0.365% | 0.370% | 61.449% | 73.405% |
| Total | 0.365% | 0.370% | 61.449% | 73.405% |

Note: The trial is terminated if any arm enters the efficacy region.

18 Two-Stage Multi-arm Designs using p -value combination

18.1 Introduction

In the drug development process, identification of promising therapies and inference on selected treatments are usually performed in two or more stages. The procedure we will be discussing here is an adaptive two-stage design that can be used for the situation of multiple treatments to be compared with a control. This will allow integration of both the stages within a single confirmatory trial controlling the multiple level type-I error. After the interim analysis in the first stage, the trial may be terminated early or continued with a second stage, where the set of treatments may be reduced due to lack of efficacy or presence of safety problems with some of the treatments. This procedure in East is highly flexible with respect to stopping rules and selection criteria and also allows re-estimation of the sample size for the second stage. Simulations show that the method may be substantially more powerful than classical one-stage multiple treatment designs with the same total sample size because second stage sample size is focused on evaluating only the promising treatments identified in the first stage. This procedure is available for continuous as well discrete endpoint studies. The current chapter deals with the continuous endpoint studies only; discrete endpoint studies are handled similarly.

18.2 Study Design

18.2.1 Introduction to the Study

18.2.2 Methodology

18.2.3 Study Design Inputs

18.2.4 Simulating under Different Alternatives

This section will explore different design options available in East with the help of an example.

18.2.1 Introduction to the Study

Consider designing a placebo controlled, double blind, randomized trial to evaluate the efficacy, pharmacokinetics, safety and tolerability of a New Chemical Entity (NCE) given as multiple weekly infusions in subjects with a recent acute coronary syndrome. There are four dose regimens to be investigated. The treatment effect is assessed through the change in PAV (percent atheroma volume) from baseline to Day 36 post-randomization, as determined by IVUS (intravascular ultrasound). The expected change in PAV for placebo group and the four dose regimens are: 0, 1, 1.1, 1.2, 1.3 and the common standard deviation is 3. The objective of the study is to find the optimal dose regimen based on the totality of the evidence including benefit-risk assessment and cost considerations.

18.2.2 Methodology

This is a randomized, double-blind, placebo-controlled study conducted in two parts using a 2-stage adaptive design. In Stage 1, approximately 250 eligible subjects will be randomized equally to one of four treatment arms (NCE [doses: 1, 2.5, 5 or 10 mg]) and matching placebo (which is 50 subjects/dose group) After all subjects in Stage 1

18 Two-Stage Multi-arm Designs using p-value combination

have completed treatment period or discontinued earlier, an interim analysis will be conducted to

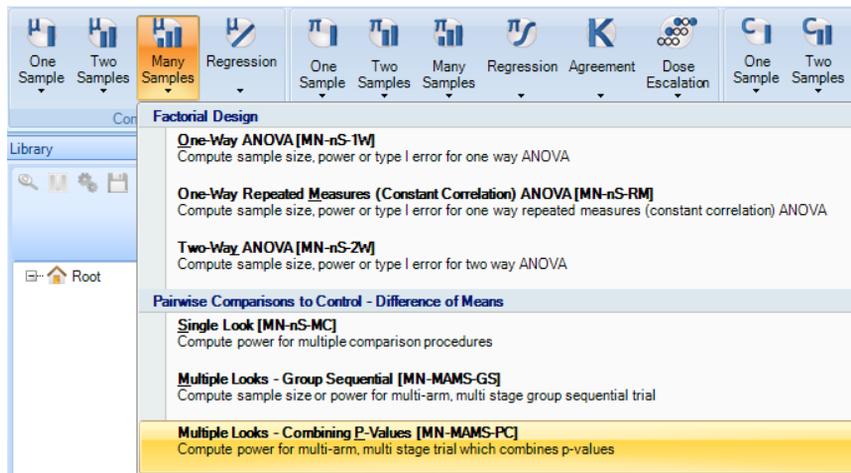
1. compare the means each dose group
2. assess safety within each dose group and
3. drop the less efficient doses

Based on the interim analysis, Stage 2 of the study will either continue with additional subjects enrolling into 1/2/3 arms (placebo and 1/2/3 favorable, active doses) or the study will be halted completely if unacceptable toxicity has been observed.

In this example, we will have the following workflow to cover different options available in East:

1. Start with four arms (4 doses + Placebo)
2. Evaluate the four doses at the interim analysis and based on the **Treatment Selection Rules** carry forward some of the doses to the next stage
3. While we select the doses, also increase the sample size of the trial by using **Sample Size Re-estimation (SSR)** tool to improve conditional power if necessary
In a real trial, both the above actions (early stopping as well as sample size re-estimation) will be performed after observing the interim data.
4. See the final design output in terms of different powers, probabilities of selecting particular dose combinations
5. See the early stopping boundaries for efficacy and futility on adjusted p-value scale
6. Monitor the actual trial using the **Interim Monitoring** tool in East.

Start East. Click Design tab, then click **Many Samples** in the **Continuous** category, and then click **Multiple Looks- Combining p-values** test.



This will bring up the input window of the design with some default values. Enter the inputs as discussed below.

18.2.3 Study Design Inputs

The four doses of the treatment- *1mg*, *2.5mg*, *5mg*, *10mg* will be compared with the Placebo arm based on their treatment means. Preliminary sample size estimates are provided to achieve an overall study power of at least 90% at an overall, adequately adjusted 1-sided type-1 or alpha level of 2.5%, after taking into account all interim and final hypothesis tests. Note that we always use 1-sided alpha since dose-selection rules are usually 1-sided.

In Stage 1, 250 subjects are initially planned for enrollment (5 arms with 50 subjects each). Following an interim analysis conducted after all subjects in Stage 1 have completed treatment period or discontinued earlier, an additional 225 subjects will be enrolled into three doses for Stage 2 (placebo and two active doses). So we start with the total of $250+225 = 475$ subjects.

The multiplicity adjustment methods available in East to compute the adjusted p-value (p-value corresponding to global NULL) are *Bonferroni*, *Sidak*, *Simes*. For discrete endpoint test, *Dunnnett Single Step* is not available since we will be using Z-statistic. Let us use the *Bonferroni* method for this example. The p-values obtained from both the stages can be combined by using the “Inverse Normal” method. In the “Inverse

18 Two-Stage Multi-arm Designs using p-value combination

Normal” method, East first computes the weights as follows:

$$w^{(1)} = \sqrt{\frac{n^{(1)}}{n}} \tag{18.1}$$

And

$$w^{(2)} = \sqrt{\frac{n^{(2)}}{n}} \tag{18.2}$$

where $n^{(1)}$ and $n^{(2)}$ are the total sample sizes corresponding to Stage 1 and stage 2 respectively and n is the total sample size.

EAST displays these weights by default but they are editable and user can specify any other weights as long as

$$w^{(1)^2} + w^{(2)^2} = 1 \tag{18.3}$$

Final p-value is given by

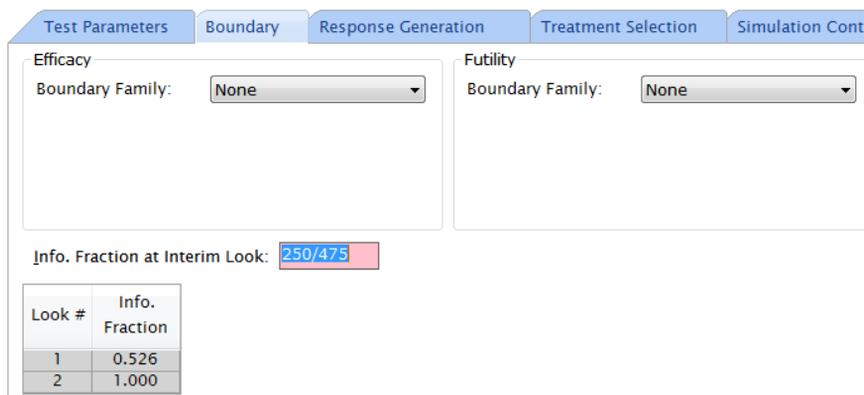
$$p = 1 - \Phi\left(w^{(1)}\Phi^{-1}(1 - p^{(1)}) + w^{(2)}\Phi^{-1}(1 - p^{(2)})\right) \tag{18.4}$$

The weights specified on this tab will be used for p-value computation. $w^{(1)}$ will be used for data before interim look and $w^{(2)}$ will be used for data after interim look. Thus, according to the samples sizes planned for the two stages in this example, the weights are calculated as $\sqrt{(250/475)}$ and $\sqrt{(225/475)}$. **Note :** These weights are updated by East once we specify the first look position as 250/475 in the **Boundary** tab. So leave these as default values for now. Set the *Number of Arms* as 5 and enter the rest of the inputs as shown below:

| | | | |
|---|------------|--------------------------|----------------|
| Number of Arms: | 5 | Number of Looks: | 2 |
| <div style="display: flex; justify-content: space-between;"> Test Parameters Boundary Response Generation Treatment Selection Simulation Controls </div> | | | |
| Test Type: | 1-Sided | Multiplicity Adjustment: | Bonferroni |
| Rejection Region: | Right-Tail | p-value Combination: | Inverse Normal |
| Type I Error (α): | 0.025 | Weight1: | 0.707 |
| Total Sample Size (n): | 475 | Weight2: | 0.707 |
| | | Test Statistic: | t |
| | | Variance: | Equal |

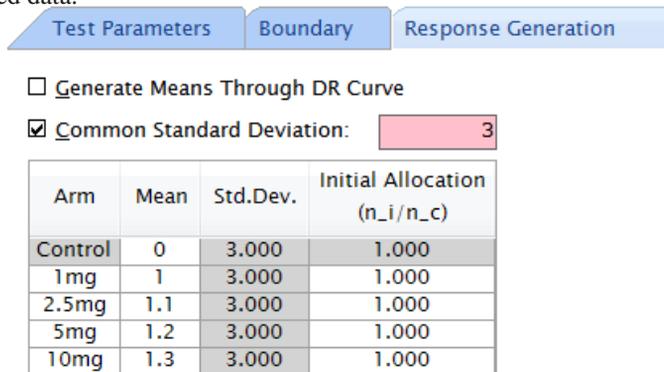
We can certainly have early stopping boundaries for efficacy and/or futility. But generally, in designs like this, the objective is to select the best dose(s) and not stop early. So for now, select the **Boundary** tab and set both the boundary families to “None”. Also, set the timing of the interim analysis as 0.526 which will be after

observing the data on 250 subjects out of 475. Enter 250/475 as shown below. Notice the updated weights on the bf Test Parameters tab.



| Look # | Info. Fraction |
|--------|----------------|
| 1 | 0.526 |
| 2 | 1.000 |

The next tab is **Response Generation** which is used to specify the true underlying means on the individual dose groups and the initial allocation from which to generate the simulated data.



| Arm | Mean | Std.Dev. | Initial Allocation (n _i /n _c) |
|---------|------|----------|--|
| Control | 0 | 3.000 | 1.000 |
| 1mg | 1 | 3.000 | 1.000 |
| 2.5mg | 1.1 | 3.000 | 1.000 |
| 5mg | 1.2 | 3.000 | 1.000 |
| 10mg | 1.3 | 3.000 | 1.000 |

One can also generate the mean response for all the arms using a dose-response curve like 4PL or Emax or Linear or Quadratic. It can be done by checking the box for **Generate Means through DR Curve** and entering appropriate parameters for DR

18 Two-Stage Multi-arm Designs using p -value combination

model selected.

The screenshot shows the 'Response Generation' tab with the following table and settings:

| Arm | Dose | Mean | Std.Dev. | Initial Allocation (n ₁ /n _c) |
|---------|-------|-------|----------|--|
| Control | 1.000 | 5.005 | 3.000 | 1.000 |
| 1mg | 2.000 | 5.037 | 3.000 | 1.000 |
| 2.5mg | 3.000 | 5.27 | 3.000 | 1.000 |
| 5mg | 4.000 | 6.788 | 3.000 | 1.000 |
| 10mg | 5.000 | 12.5 | 3.000 | 1.000 |

Settings: Generate Means Through DR Curve, Common Standard Deviation: 3.
 Dose Response Curve: Four Parameter Logistic (selected), Four Parameter Logistic, Emax, Linear, Quadratic.
 Parameters: $\theta = 5$, $\tau = 0.5$.
 Button: Plot DR Curve

For this example, we will use the given means and standard deviation and not generate them using a DR curve. Make sure the means are 0, 1, 1.1, 1.2, 1.3 and SD is 3.

Before we update the **Treatment Selection** tab, go to the **Simulation Control Parameters** tab where we can specify the number of simulations to run, the random number seed and also to save the intermediate simulation data. For now, enter the inputs as shown below and keep all other inputs as default.

The screenshot shows the 'Simulation Controls' tab with the following settings:

Number of Simulations: 10000
 Refresh Frequency: 1000
 Random Number Seed: Clock, Fixed (100)
 Suppress All Intermediate Output
 Pause after Refresh
 Pause at End

Output Options:
 Output Type: Case Data
 Save summary statistics for every simulation run
 Save subject level data for 1 simulation runs
 Note: Max. 100,000 records will be saved.

Click on the **Treatment Selection** tab. This tab is to select the scale to compute the treatment-wise effects. For selecting treatments for the second stage, the treatment effect scale will be required, but the control treatment will not be considered for selection. It will always be there in the second stage. The list under **Treatment Effect Scale** allows you to set the selection rules on different scales. Select **Estimated δ** from this list. It means that all the selection rules we specify on this tab will be in terms of the estimated value of treatment effect, δ , i.e., difference from

placebo. Here is a list of all available treatment effect scales: Estimated Mean, Estimated δ , Estimated δ/σ , Test Statistic, Conditional Power, Isotonic Mean, Isotonic δ , Isotonic δ/σ .

For more details on these scales, refer to the Appendix **K** chapter on this method.

The next step is to set the treatment selection rules for the second stage.

Select Best r Treatments: The best treatment is defined as the treatment having the highest or lowest mean effect. The decision is based on the rejection region. If it is “Right-Tail” then the highest should be taken as best. If it is “Left-Tail” then the lowest is taken as best. Note that the rejection region does not affect the choice of treatment based on conditional power.

Select treatments within ϵ of Best Treatment: Suppose the treatment effect scale is **Estimated δ** . If the best treatment has a treatment effect of δ_b and ϵ is specified as 0.1 then all the treatments which have a δ as $\delta_b - 0.1$ or more are chosen for Stage 2.

Select treatments greater than threshold ζ : The treatments which have the treatment effect scale greater or less than the threshold (ζ) specified by the user according to the rejection region. But if the treatment effect scale is chosen as the conditional power then it will be greater than all the time.

Use R for Treatment Selection: If you wish to define any customized treatment selection rules, it can be done by writing an R function for those rules to be used within East. This is possible due to the R Integration feature in East. Refer to the appendix chapter on **R Functions** for more details on syntax and use of this feature. A template file for defining treatment selection rules is also available in the subfolder **RSamples** under your East installation directory.

| Tasks | File Name | Function Name |
|--------------------------|-----------|---------------|
| Treatment Selection | | |
| Initialize R Environment | | |

For more details on using R to define Treatment selection rules, refer to section **O.10**.

18 Two-Stage Multi-arm Designs using p-value combination

Selecting multiple doses (arms) for Stage 2 would be more effective than selecting just the best one. For this example, select the first rule **Select Best r treatments** and set $r = 2$ which indicates that East will select the best two doses for Stage 2 out the four. We will leave the Allocation Ratio after Selection as 1 to yield equal allocation between the control and selected doses in Stage 2.

Click the **Simulate** button to run the simulations. When the simulations are over, a row gets added in the **Output Preview** area. Save this row to the **Library** by clicking the  icon in the toolbar. Rename this scenario as **Best2**. Double click it to see the detailed output.

| Test Parameters | |
|--------------------------------|--------------------|
| Simulation ID | Sim1 |
| Number of Arms | 5 |
| Number of Looks | 2 |
| Test Type | 1-Sided |
| Rejection Region | Right-Tail |
| Specified α | 0.025 |
| Total Sample Size (n) | 475 |
| Multiplicity Adjustment | Bonferroni |
| p-value Combination | Inverse Normal |
| Weight1 | 0.725 |
| Weight2 | 0.688 |
| Variance | Equal |
| Model Parameters | |
| Test Statistic | t |
| Boundary Parameters | |
| Info. Fraction at Interim Look | 0.526 |
| Treatment Selection Parameters | |
| Treatment Effect Scale | Estimated δ |
| Treatment Selection Rule | Best 2 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |
| Elapsed Time: | 00:00:03 |

Overall Powers

| | |
|---|-------|
| Global (Reject any Hi) | 0.885 |
| Conjunctive (Reject all Hi where $\mu_i > \mu_0$) | 0 |
| Disjunctive (Reject at least one Hi where $\mu_i > \mu_0$) | 0.885 |
| FWER (Reject any Hi where $\mu_i \leq \mu_0$) | 0 |

Lookwise Summary

| Look # | Average Total Sample Size (n) | Trial Termination | |
|---------|-------------------------------|-------------------|---------|
| | | Count | % |
| 1 | 250 | 0 | 0.000% |
| 2 | 475 | 8845 | 88.450% |
| OverAll | | 8845 | 88.450% |

Detailed Efficacy Outcomes for all 10000 Simulations

| Efficacy Outcome | Average Sample Size | Count (n) | % (n/10000) |
|-------------------|---------------------|-----------|-------------|
| 1mg only | 475 | 109 | 1.090% |
| 2.5mg only | 475 | 222 | 2.220% |
| 5mg only | 475 | 393 | 3.930% |
| 10mg only | 475 | 631 | 6.310% |
| 1mg & 2.5mg only | 475 | 529 | 5.290% |
| 1mg & 5mg only | 475 | 829 | 8.290% |
| 1mg & 10mg only | 475 | 1144 | 11.440% |
| 2.5mg & 5mg only | 475 | 1158 | 11.580% |
| 2.5mg & 10mg only | 475 | 1576 | 15.760% |
| 5mg & 10mg only | 475 | 2254 | 22.540% |
| None | 475 | 1155 | 11.550% |
| Total | | 10000 | 100.000% |

Marginal Probabilities of Selection and Efficacy

| Label | Average Sample Size | Marginal Probability of Selection | | Marginal Probability of Efficacy | |
|-------|---------------------|-----------------------------------|---------|----------------------------------|---------|
| | | Count | % | Count | % |
| 1mg | 76.805 | 3574 | 35.740% | 2611 | 26.110% |
| 2.5mg | 82.933 | 4391 | 43.910% | 3485 | 34.850% |
| 5mg | 91.648 | 5553 | 55.530% | 4634 | 46.340% |
| 10mg | 98.615 | 6482 | 64.820% | 5605 | 56.050% |

The first table in the detailed output shows the overall power including global power, conjunctive power, disjunctive power and FWER. The definitions for different powers are as follows:

- Global Power: probability of demonstrating statistical significance on one or

- more treatment groups
- Conjunctive Power: probability of demonstrating statistical significance on all treatment groups which are truly effective
- Disjunctive Power: probability of demonstrating statistical significance on at least one treatment group which is truly effective
- FWER: probability of incorrectly demonstrating statistical significance on at least one treatment group which is truly ineffective

For our example, there is 88% global power which is the probability of this design to reject any null hypothesis, where the set of null hypothesis are the TRUE proportion of responders at each dose equals that of control. Also shown is conjunctive and disjunctive power, as well as Family Wise Error Rate (FWER).

The **Lookwise Summary** table summarizes the number of simulated trials that ended with a conclusion of efficacy, i.e., rejected any null hypothesis, at each look. In this example, no simulated trial stopped at the interim analysis with an efficacy conclusion since there were no stopping boundaries, but 8845 simulations yielded an efficacy conclusion via the selected dose after Stage 2. This is consistent with the global power.

The table **Detailed Efficacy Outcomes for all 10000 Simulations** summarizes the number of simulations for which each individual dose group or pairs of doses were selected for Stage 2 and yielded an efficacy conclusion. For example, the pair (2.5mg, 10mg only) was observed to be efficacious in approximately 16% of the trials (1576/10000).

The next table **Marginal Probabilities of Selection and Efficacy**, summarizes the number and percent of simulations in which each dose was selected for Stage 2, regardless of whether it was found significant at end of Stage 2 or not, as well as the number and percent of simulations in which each dose was selected and found significant. Average sample size is also shown. It tells us how frequently the dose (either alone or with some other dose) was selected and efficacious. For example, dose 10mg was selected in approximately 65% trials and was efficacious in approximately 56% trials. (which is the sum of 631, 1144, 1576, 2254 simulations from previous table.)

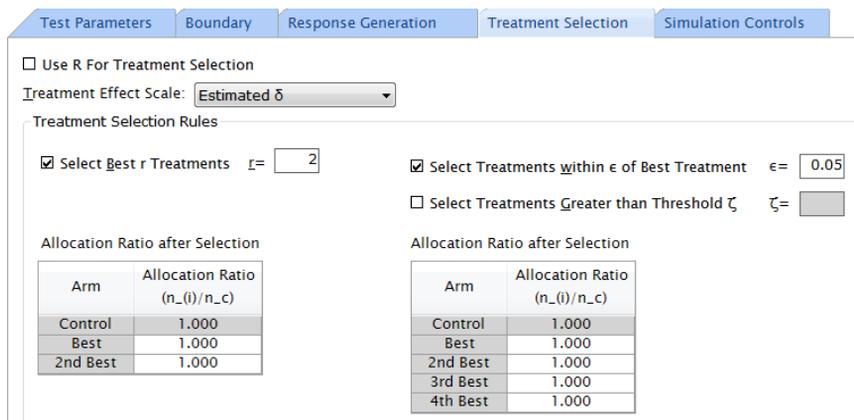
The advantage of 2-stage “treatment selection design” or “drop-the-loser” design is that it allows to drop the less performing/futile arms based on the interim data and still preserves the type-1 error as well as achieve the desired power.

In the **Best2** scenario, we dropped two doses ($r = 2$). Suppose, we had decided to proceed to stage 2 without dropping any doses. In this case, Power would have

18 Two-Stage Multi-arm Designs using p -value combination

dropped significantly. To verify this in East, click the  button on the bottom left corner of the screen. This will take us back to the input window of the last simulation scenario. Go to **Treatment Selection** tab and set $r = 4$ and save it to **Library**. Rename this scenario as **All4**. Double click it to see the detailed output. We can observe that the power drops from 88% to 78%. That is because the sample size of 225 is being shared among five arms as against three arms in the **Best2** case.

Now go back to **Treatment Selection** tab, set $r = 2$ as before. Select one more rule, **Select Treatments within ϵ of Best Treatment** and set the ϵ value as 0.05. The tab should look as shown below.



Use R For Treatment Selection
 Treatment Effect Scale: Estimated δ

Treatment Selection Rules

Select Best r Treatments $r =$

Select Treatments within ϵ of Best Treatment $\epsilon =$

Select Treatments Greater than Threshold ζ $\zeta =$

Allocation Ratio after Selection

| Arm | Allocation Ratio (n _(i) /n _c) |
|----------|--|
| Control | 1.000 |
| Best | 1.000 |
| 2nd Best | 1.000 |

Allocation Ratio after Selection

| Arm | Allocation Ratio (n _(i) /n _c) |
|----------|--|
| Control | 1.000 |
| Best | 1.000 |
| 2nd Best | 1.000 |
| 3rd Best | 1.000 |
| 4th Best | 1.000 |

Also set the Starting Seed on **Simulation Controls** tab to 100. **Note that since we have selected two treatment selection rules, East will simulate two different scenarios, one for each rule.** As we want to compare the results from these two scenarios, we use the same starting seed. That will ensure same random number generation and the only difference in results will be the effect of the two rules.

Save these two scenarios in the **Library** as **r=2** and **epsilon=0.05**, select them and click the  icon in the toolbar to see them side-by-side.

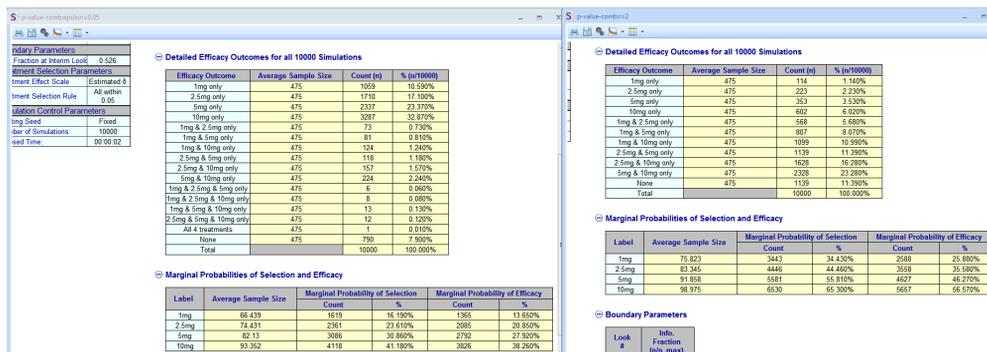
| | p-value-comb:r=2 | p-value-comb:epsilon=... |
|---------------------------------------|--------------------|--------------------------|
| Mnemonic | MN-MAMS-PC | MN-MAMS-PC |
| Test Parameters | | |
| Number of Arms | 5 | 5 |
| No. of Looks | 2 | 2 |
| Test Type | 1-Sided | 1-Sided |
| Rejection Region | Right-Tail | Right-Tail |
| Specified α | 0.025 | 0.025 |
| Multiplicity Adjustment | Bonferroni | Bonferroni |
| p-value Combination | Inverse Normal | Inverse Normal |
| Weight1 | 0.725 | 0.725 |
| Weight2 | 0.688 | 0.688 |
| Variance | Equal | Equal |
| Model Parameters | | |
| Test Statistic | t | t |
| Boundary Parameters | | |
| Info. Fraction at Interim Look | 0.526 | 0.526 |
| MCP Results | | |
| Global Power | 0.886 | 0.921 |
| Conjunctive Power | 0 | 0 |
| Disjunctive Power | 0.886 | 0.921 |
| Overall FWER | 0 | 0 |
| Sample Size | | |
| Maximum | 475 | 475 |
| Average Total Sample Size | 475 | 475 |
| Response Generation Parameters | | |
| Common Standard Deviation | 3 | 3 |
| Treatment Selection Parameters | | |
| Treatment Effect Scale | Estimated δ | Estimated δ |
| Treatment Selection Rule | Best 2 | All within 0.05 |

Notice the powers for the two scenarios. The scenario with the rule of $\delta_b - 0.05$ yields more power than the **Best2 Scenario**. Note that δ_b is the highest value among the simulated of δ values for the four doses at the interim look.

You can also view the **Output Details** of these two scenarios. Select the two nodes as

18 Two-Stage Multi-arm Designs using p -value combination

before but this time, click the  icon in the toolbar.



Left Window: Detailed Efficacy Outcomes for all 10000 Simulations

| Efficacy Outcome | Average Sample Size | Count (n) | % (n/10000) |
|-------------------------|---------------------|-----------|-------------|
| 1mg only | 475 | 1055 | 10.550% |
| 2.5mg only | 475 | 1710 | 17.100% |
| 5mg only | 475 | 2337 | 23.370% |
| 10mg only | 475 | 3207 | 32.070% |
| 1mg & 2.5mg only | 475 | 73 | 0.730% |
| 1mg & 5mg only | 475 | 81 | 0.810% |
| 1mg & 10mg only | 475 | 128 | 1.280% |
| 2.5mg & 5mg only | 475 | 118 | 1.180% |
| 2.5mg & 10mg only | 475 | 157 | 1.570% |
| 5mg & 10mg only | 475 | 224 | 2.240% |
| 1mg & 2.5mg & 5mg only | 475 | 6 | 0.060% |
| 1mg & 2.5mg & 10mg only | 475 | 8 | 0.080% |
| 1mg & 5mg & 10mg only | 475 | 13 | 0.130% |
| 2.5mg & 5mg & 10mg only | 475 | 12 | 0.120% |
| All 4 treatments | 475 | 1 | 0.010% |
| None | 475 | 790 | 7.900% |
| Total | | 10000 | 100.000% |

Marginal Probabilities of Selection and Efficacy

| Label | Average Sample Size | Count | % | Count | % |
|-------|---------------------|-------|---------|-------|---------|
| 1mg | 66.439 | 1619 | 16.190% | 1365 | 13.650% |
| 2.5mg | 74.571 | 2361 | 23.610% | 2083 | 20.830% |
| 5mg | 82.13 | 3086 | 30.860% | 2752 | 27.520% |
| 10mg | 93.362 | 4118 | 41.180% | 3626 | 36.260% |

Right Window: Detailed Efficacy Outcomes for all 10000 Simulations

| Efficacy Outcome | Average Sample Size | Count (n) | % (n/10000) |
|-------------------|---------------------|-----------|-------------|
| 1mg only | 475 | 114 | 1.140% |
| 2.5mg only | 475 | 223 | 2.230% |
| 5mg only | 475 | 263 | 2.630% |
| 10mg only | 475 | 692 | 6.920% |
| 1mg & 2.5mg only | 475 | 568 | 5.680% |
| 1mg & 5mg only | 475 | 807 | 8.070% |
| 1mg & 10mg only | 475 | 1099 | 10.990% |
| 2.5mg & 5mg only | 475 | 1139 | 11.390% |
| 2.5mg & 10mg only | 475 | 1628 | 16.280% |
| 5mg & 10mg only | 475 | 2328 | 23.280% |
| None | 475 | 1139 | 11.390% |
| Total | | 10000 | 100.000% |

Marginal Probabilities of Selection and Efficacy

| Label | Average Sample Size | Count | % | Count | % |
|-------|---------------------|-------|---------|-------|---------|
| 1mg | 75.823 | 3443 | 34.430% | 2988 | 29.880% |
| 2.5mg | 83.345 | 4446 | 44.460% | 3658 | 36.580% |
| 5mg | 91.858 | 5881 | 58.810% | 4827 | 48.270% |
| 10mg | 98.976 | 6930 | 69.300% | 5667 | 56.670% |

Boundary Parameters

| Look # | Info. Fraction (info_max) |
|--------|---------------------------|
| 1 | |

Notice from this comparison, due to a more general rule based on ϵ , we can select multiple doses and not just two. At the same time, the marginal probability of selection as well as efficacy for each dose drops significantly.

18.2.4 Simulating under Different Alternatives

Since this is a simulation based design, we can perform sensitivity analyses by changing some of the inputs and observing effects on the overall power and other output. Let us first make sure that this design preserves the total type I error. It can be done by running the simulations under “Null” hypothesis.

Select the last design created which would be **epsilon = 0.05** in the Library and click the  icon. This will take you to the input window of that design. Go to **Response Generation** tab and enter the inputs as shown below. Notice that all the means are 0

which means the simulations will be run under NULL assumption.

Test Parameters
Boundary
Response Generation

Generate Means Through DR Curve

Common Standard Deviation:

| Arm | Mean | Std.Dev. | Initial Allocation (n _i /n _c) |
|---------|-------|----------|---|
| Control | 0.000 | 3.000 | 1.000 |
| 1mg | 0.000 | 3.000 | 1.000 |
| 2.5mg | 0.000 | 3.000 | 1.000 |
| 5mg | 0.000 | 3.000 | 1.000 |
| 10mg | 0.000 | 3.000 | 1.000 |

Run the simulations and go to the detailed output by saving the row from **Output Preview** to the **Library**. Notice the global power and the simulated FWER is less than 0.025 which means the overall type1 error is preserved.

18.3 Sample Size Re-estimation

As seen in the previous scenario, the desired power of approximately 92% is achieved with the sample size of 475 if the initial assumptions ($\mu_c = 0$, $\mu_{1mg} = 1$, $\mu_{2.5mg} = 1.1$, $\mu_{5mg} = 1.2$ and $\mu_{10mg} = 1.3$) hold true. But if they do not, then the original sample size of 475 may be insufficient to achieve 92% power. The adaptive sample size re-estimation is suited to this purpose. In this approach we start out with a sample size of 475 subjects, but take an interim look after data are available on 250 subjects. The purpose of the interim look is not to stop the trial early but rather to examine the interim data and continue enrolling past the planned 475 subjects if the interim results are promising enough to warrant the additional investment of sample size. This strategy has the advantage that the sample size is finalized only after a thorough examination of data from the actual study rather than through making a large up-front sample size commitment before any data are available. Furthermore, if the sample size may only be increased but never decreased from the originally planned 475 subjects, there is no loss of efficiency due to overruns. Suppose the mean responses on the five doses are as shown below. Update the **Response Generation** tab

18 Two-Stage Multi-arm Designs using p -value combination

accordingly and also set the seed as 100 in the **Simulation Controls** tab.

Test Parameters Boundary Response Generation

Generate Means Through DR Curve

Common Standard Deviation:

| Arm | Mean | Std.Dev. | Initial Allocation (n _i /n _c) |
|---------|------|----------|---|
| Control | 0 | 3.000 | 1.000 |
| 1mg | 0.8 | 3.000 | 1.000 |
| 2.5mg | 0.95 | 3.000 | 1.000 |
| 5mg | 1 | 3.000 | 1.000 |
| 10mg | 1 | 3.000 | 1.000 |

Run 10000 simulations and save the simulation row to the **Library** by clicking the



icon in the toolbar. See the details.

Overall Powers

| | |
|---|-------|
| Global (Reject any Hi) | 0.779 |
| Conjunctive (Reject all Hi where $\mu_i > \mu_0$) | 2E-4 |
| Disjunctive (Reject at least one Hi where $\mu_i > \mu_0$) | 0.779 |
| FWER (Reject any Hi where $\mu_i \leq \mu_0$) | 0 |

Lookwise Summary

| Look # | Average Total Sample Size (n) | Trial Termination | |
|---------|-------------------------------|-------------------|---------|
| | | Efficacy | |
| | | Count | % |
| 1 | 250 | 0 | 0.000% |
| 2 | 475 | 7785 | 77.850% |
| OverAll | | 7785 | 77.850% |

Notice that the global power has dropped from 92% to 78%. Let us re-estimate the sample size to achieve the desired power. Add the **Sample Size Re-estimation** tab by clicking the button . A new tab gets added as shown below.

SSR At: For a K -look group sequential design, one can decide the time at which conditions for adaptations are to be checked and actual adaptation is to be

18 Two-Stage Multi-arm Designs using p -value combination

carried out. This can be done either at some intermediate look or after some specified information fraction. The possible values of this parameter depend upon the user choice. The default choice for this design is always the **Look #**. and is fixed to 1 since it is always a 2-look design.

Target CP for Re-estimating Sample Size: The primary driver for increasing the sample size at the interim look is the desired (or target) conditional power or probability of obtaining a positive outcome at the end of the trial, given the data already observed. For this example we have set the conditional power at the end of the trial to be 92%. East then computes the sample size that would be required to achieve this conditional power.

Maximum Sample Size if Adapt (multiplier; total): As just stated, a new sample size is computed at the interim analysis on the basis of the observed data so as to achieve some target conditional power. However the sample size so obtained will be overruled unless it falls between pre-specified minimum and maximum values. For this example, let us use the multiplier as 2 indicating that we intend to double the original sample size if the results are promising. The range of allowable sample sizes is [475, 950]. If the newly computed sample size falls outside this range, it will be reset to the appropriate boundary of the range. For example, if the sample size needed to achieve the desired 90% conditional power is less than 475, the new sample size will be reset to 475. In other words we will not decrease the sample size from what was specified initially. On the other hand, the upper bound of 950 subjects demonstrates that the sponsor is prepared to double the sample size in order to achieve the desired 90% conditional power. But if 90% conditional power requires more than 950 subjects, the sample size will be reset to 950, the maximum allowed.

Promising Zone Scale: One can define the promising zone as an interval based on conditional power, test statistic, or estimated δ/σ . The input fields change according to this choice. The decision of altering the sample size is taken based on whether the interim value of conditional power / test statistic / δ/σ lies in this interval or not. Let us keep the default scale which is Conditional Power.

Promising Zone: Minimum/Maximum Conditional Power (CP): The sample size will only be altered if the estimate of CP at the interim analysis lies in a pre-specified range, referred to as the “Promising Zone”. Here the promising zone is 0.30 – 0.90. The idea is to invest in the trial in stages. Prior to the interim analysis the sponsor is only committed to a sample size of 475 subjects. If, however, the results at the interim analysis appear reasonably promising, the

sponsor would be willing to make a larger investment in the trial and thereby improve the chances of success. Here we have somewhat arbitrarily set the lower bound for a promising interim outcome to be $CP = 0.30$. An estimate $CP < 0.30$ at the interim analysis is not considered promising enough to warrant a sample size increase. It might sometimes be desirable to also specify an upper bound beyond which no sample size change will be made. Here we have set that upper bound of the promising zone at $CP = 0.90$. In effect we have partitioned the range of possible values for conditional power at the interim analysis into three zones; *unfavorable* ($CP < 0.3$), *promising* ($0.3 \leq CP < 0.9$), and *favorable* ($CP \geq 0.9$). Sample size adaptations are made only if the interim CP falls in the promising zone at the interim analysis. The promising zone defined on the Test Statistic scale or the Estimated δ/σ scale works similarly.

SSR Function in Promising Zone: The behavior in the promising zone can either be defined by a continuous function or a step function. The default is continuous where East accepts the two quantities - (Multiplier, Target CP) and re-estimates the sample size depending upon the interim value of CP/test statistic/effect size. The SSR function can be defined as a step-function as well. This can be done with a single piece or with multiple pieces. For each piece, define the step function in terms of:

- the interval of CP/test statistic/ δ/σ . This depends upon the choice of promising zone scale.
- the value of re-estimated sample size in that interval.
- for single piece, just the total re-estimated sample size is required as an input.

If the interim value of CP/ test statistic/ δ/σ lies in the promising zone then the re-estimation will be done using this step function.

Let us set the inputs on **Sample Size Re-estimation** tab as shown below. Just for the comparison purpose, also run the simulations without adaptation. Both the scenarios

18 Two-Stage Multi-arm Designs using p -value combination

can also be run together by entering two values 1, 2 in the cell for Multiplier.

| | | | | | | | | | |
|--|--|--------------------|--|---------------------|--|---------------------|--|---------------------------|--|
| Number of Arms: 5 | | Number of Looks: 2 | | | | | | | |
| Test Parameters | | Boundary | | Response Generation | | Treatment Selection | | Sample Size Re-estimation | |
| SSR At: | | Look # | | 1 | | | | | |
| Max. Sample Size if Adapt (multiplier; total #): | | 1, 2 | | Computed | | | | | |
| Target CP for Re-estimating Sample Size: | | 0.92 | | | | | | | |
| Promising Zone Scale: (Based on Current Winner) | | Cond. Power | | CP | | | | | |
| Promising Zone: | | Min. CP: | | 0.3 | | | | | |
| | | Max. CP: | | 0.9 | | | | | |
| SSR Function in Promising Zone: | | Continuous | | | | | | | |

Run 10000 simulations and see the Details.

With Sample Size Re-estimation

Overall Powers

| | |
|---|-------|
| Global (Reject any H ₀) | 0.853 |
| Conjunctive (Reject all H ₀ where $\mu_i > \mu_0$) | 1E-4 |
| Disjunctive (Reject at least one H ₀ where $\mu_i > \mu_0$) | 0.853 |
| FWER (Reject any H ₀ where $\mu_i \leq \mu_0$) | 0 |

Zone-wise Averages

| Zone | Global Power | | Conjunctive Power | | Disjunctive Power | | FWER | | Not Rejecting Any H ₀ | | Total Simulations | | Average Sample Size |
|-------------|--------------|---------|-------------------|--------|-------------------|---------|-------|--------|----------------------------------|---------|-------------------|----------|---------------------|
| | Count | Row % | Count | Row % | Count | Row % | Count | Row % | Count | Row % | Count | Column % | |
| Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| Unfavorable | 11 | 2.056% | 0 | 0.000% | 11 | 2.056% | 0 | 0.000% | 524 | 97.944% | 535 | 5.350% | 475 |
| Promising | 1442 | 89.565% | 0 | 0.000% | 1442 | 89.565% | 0 | 0.000% | 168 | 10.435% | 1610 | 16.100% | 881.88 |
| Favorable | 7077 | 90.095% | 1 | 0.013% | 7077 | 90.095% | 0 | 0.000% | 778 | 9.905% | 7855 | 78.550% | 475 |
| Efficacy | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| All Trials | 8530 | 85.300% | 1 | 0.010% | 8530 | 85.300% | 0 | 0.000% | 1470 | 14.700% | 10000 | 100.000% | 540.508 |

Promising Zone defined as 0.3 <= CP < 0.9

Without Sample Size Re-estimation

Overall Powers

| | |
|---|-------|
| Global (Reject any H ₀) | 0.779 |
| Conjunctive (Reject all H ₀ where $\mu_i > \mu_0$) | 2E-4 |
| Disjunctive (Reject at least one H ₀ where $\mu_i > \mu_0$) | 0.779 |
| FWER (Reject any H ₀ where $\mu_i \leq \mu_0$) | 0 |

Zone-wise Averages

| Zone | Global Power | | Conjunctive Power | | Disjunctive Power | | FWER | | Not Rejecting Any H ₀ | | Total Simulations | | Average Sample Size |
|-------------|--------------|---------|-------------------|--------|-------------------|---------|-------|--------|----------------------------------|---------|-------------------|----------|---------------------|
| | Count | Row % | Count | Row % | Count | Row % | Count | Row % | Count | Row % | Count | Column % | |
| Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| Unfavorable | 6 | 1.230% | 0 | 0.000% | 6 | 1.230% | 0 | 0.000% | 482 | 98.770% | 488 | 4.880% | 475 |
| Promising | 641 | 40.289% | 0 | 0.000% | 641 | 40.289% | 0 | 0.000% | 950 | 59.711% | 1591 | 15.910% | 475 |
| Favorable | 7138 | 90.115% | 2 | 0.025% | 7138 | 90.115% | 0 | 0.000% | 783 | 9.885% | 7921 | 79.210% | 475 |
| Efficacy | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| All Trials | 7785 | 77.850% | 2 | 0.020% | 7785 | 77.850% | 0 | 0.000% | 2215 | 22.150% | 10000 | 100.000% | 475 |

Promising Zone defined as 0.3 <= CP < 0.9

We observe from the table the power of adaptive implementation is approximately

85% which is almost 8% improvement over the non-adaptive design. This increase in power has come at an average cost of 540-475 = 65 additional subjects. Next we observe from the **Zone-wise Averages** table that 1610 of 10000 trials (16%) underwent sample size re-estimation (Total Simulation Count in the “Promising Zone”) and of those 1610 trials, 89% were able to reject the Global null hypothesis. The average sample size, conditional on adaptation is 882.

18.4 Adding Early Stopping Boundaries

One can also incorporate stopping boundaries to stop at the interim early for efficacy or futility. The efficacy boundary can be defined based on **Adjusted p-value** scale whereas futility boundary can be on **Adjusted p-value** or δ/σ scale.



Click the  button on the bottom left corner of the screen. This will take you back to the input window of the last simulation scenario. Go to **Boundary** tab and set Efficacy and Futility boundaries to “Adjusted p-value”. These boundaries are for

18 Two-Stage Multi-arm Designs using p-value combination

early stopping at look1. As the note on this tab says:

- If any one adjusted p-value is \leq efficacy p-value boundary then stop the trial for efficacy
- If only all the adjusted p-values are $>$ futility p-value then stop the trial for futility. Else carry forward all the treatments to the next step of treatment selection.

Stopping early for efficacy or futility is step which is carried out before applying the treatment selection rules. The simulation output has the same explanation as above except the **Lookwise Summary** table may have some trials stopped at the first look due to efficacy or futility.

18.5 Interim Monitoring with Treatment Selection

Select the simulation node with **SSR** implementation and click the **IM** icon. It will invoke the **Interim Monitoring** dashboard. Click the **Enter Interim Data** icon to open the **Test Statistic Calculator**. The “Sample Size” column is filled out according to the originally planned design (50/arm). Enter the data as shown below:

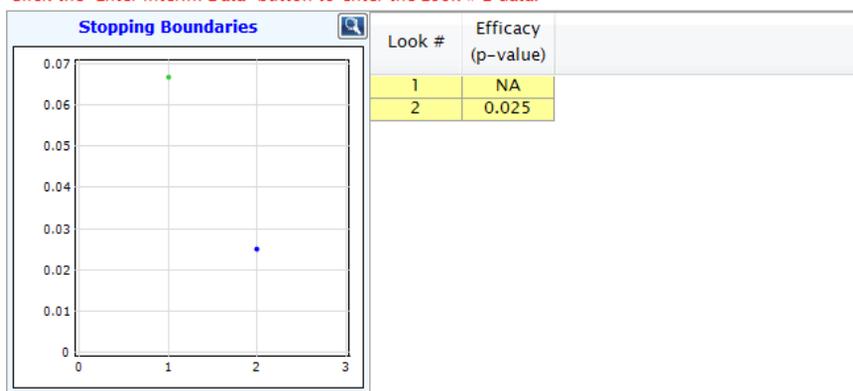
| Arm | Sample Size | Mean | Std. Deviation | Test Statistic | Raw p-values |
|---------|-------------|-------|----------------|----------------|--------------|
| Control | 50 | 0.005 | 3.000 | NA | NA |
| 1mg | 50 | 0.9 | 3.000 | 1.492 | 0.069 |
| 2.5mg | 50 | 1.15 | 3.000 | 1.908 | 0.030 |
| 5mg | 50 | 1.22 | 3.000 | 2.025 | 0.023 |
| 10mg | 50 | 1.3 | 3.000 | 2.158 | 0.017 |

Click **Recalc** to calculate the test statistic as well as the raw p-values. Notice that the

p-values for *1mg* and *2.5mg* are 0.069 and 0.030 respectively which are greater than 0.025. We will drop these doses in the second stage. On clicking **OK**, it updates the dashboard. The overall adjusted p-value is 0.067.

| Look # | Number of Arms | Adjusted p-value for Best Treatment | Prespecified Weights | Prespecified Efficacy Boundary (p-value) |
|--------|----------------|-------------------------------------|----------------------|--|
| 1 | 5 | 0.067 | 0.725 | NA |
| 2 | | | 0.688 | 0.025 |

Click the "Enter Interim Data" button to enter the Look # 2 data.



Open the test statistic calculator for the second look and enter the following information and also drop the two doses *1mg* and *2.5mg* using the dropdown of

18 Two-Stage Multi-arm Designs using p-value combination

“Action”. Click **Recalc** to calculate the test statistic as well as the raw p-values.

Test Statistic Calculator

Editing Look #2 (Incremental Data)

Terminate Trial for Futility

| Arm | Action | Sample Size | Mean | Std. Deviation | Test Statistic | Raw p-values |
|---------|----------|-------------|-------|----------------|----------------|--------------|
| Control | NA | 75 | 0.009 | 3.000 | NA | NA |
| 1mg | Drop | | | | | |
| 2.5mg | Drop | | | | | |
| 5mg | Continue | 75 | 1.23 | 3.000 | 2.492 | 0.007 |
| 10mg | Continue | 75 | 1.29 | 3.000 | 2.615 | 0.005 |

Recalc OK Cancel

On clicking **OK**, it updates the dashboard. Observe that the adjusted p-value for *10mg* crosses the efficacy boundary. It can also be observed in the **Stopping Boundaries** chart.

| Look # | Number of Arms | Adjusted p-value for Best Treatment | Prespecified Weights | Prespecified Efficacy Boundary (p-value) |
|--------|----------------|-------------------------------------|----------------------|--|
| 1 | 5 | 0.067 | 0.725 | NA |
| 2 | 3 | 0.004 | 0.688 | 0.025 |

Click the "Edit Interim Data" button to edit the Look # 2 data.

Stopping Boundaries

| Look # | Efficacy (p-value) |
|--------|--------------------|
| 1 | NA |
| 2 | 0.025 |

The final p-value adjusted for multiple treatments is 0.00353.

19 Normal Superiority Regression

Linear regression models are used to examine the relationship between a response variable and one or more explanatory variables assuming that the relationship is linear. In this chapter, we discuss the design of three types of linear regression models. In Section 19.1, we examine the problem of testing a single slope in a simple linear regression model involving one continuous covariate. In Section 19.2, we examine the problem of testing the equality of two slopes in a linear regression model with only one observation per subject. Finally, in Section 19.3, we examine the problem of testing the equality of two slopes in a linear regression repeated measures model, applied to a longitudinal setting.

19.1 Linear Regression, Single Slope

19.1.1 Trial Design

We assume that the observed value of a response variable Y is a linear function of an explanatory variable X plus random noise. For each of the $i = 1, \dots, n$ subjects in a study

$$Y_i = \gamma + \theta X_i + \epsilon_i$$

Here the ϵ_i are independent normal random variables with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma_\epsilon^2$. We follow Dupont et al. (1998) and emphasize a slight distinction between observational and experimental studies. In an observational study, the values X_i are attributes of randomly chosen subjects and their possible values are not known to the investigator at the time of a study design. In an experimental study, a subject is randomly assigned (with possibly different probabilities) to one of the predefined experimental conditions. Each of these conditions is characterized by a certain value of explanatory variable X that is completely defined at the time of the study design. In both cases the value X_i characterizing either an attribute or experimental exposure of subject i is a random variable with a variance σ_x^2 .

We are interested in testing that the slope θ is equal to a specified value θ_0 . Thus we test the null hypothesis $H_0: \theta = \theta_0$ against the two-sided alternative $H_1: \theta \neq \theta_0$ or a one-sided alternative hypothesis $H_1: \theta < \theta_0$ or $H_1: \theta > \theta_0$.

Let $\hat{\theta}$ denote the estimate of θ , and let $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_x^2$ denote the estimates of σ_ϵ^2 and σ_x^2 based on n observations. The variance of $\hat{\theta}$ is

$$\sigma^2 = \frac{\sigma_\epsilon^2}{n\sigma_x^2}. \tag{19.1}$$

The test statistic is defined as

$$Z = (\hat{\theta} - \theta_0) / \hat{\sigma}, \tag{19.2}$$

where

$$\hat{\sigma}^2 = \frac{\hat{\sigma}_\epsilon^2}{n\hat{\sigma}_x^2}$$

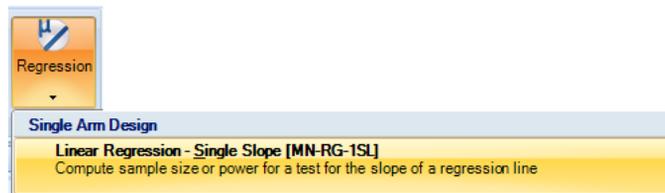
is the estimate of the variance of $\hat{\theta}$ based on n observations. Notice that the test statistic is centered so as to have a mean of zero under the null hypothesis.

We want to design the study so the power is attained when $\theta = \theta_1$. The power depends on θ_0 , θ_1 , σ_x , and σ_ϵ through $\theta_0 - \theta_1$ and σ_x/σ_ϵ .

19.1.1 Trial Design

During the development of medications, we often want to model the dose-response relationship, which may be done by estimating the slope of the regression, where Y is the appropriate response variable and the explanatory variable X is a set of specified doses. Consider a clinical trial involving four doses of a medication. The doses and randomization of subjects across the doses have been chosen so that the standard deviation $\sigma_x = 9$. Based on prior studies, it is assumed that $\sigma_\epsilon = 15$. If there is no dose response, the slope is equal to 0. Thus we will test the null hypothesis $H_0: \theta = 0$ against a two-sided alternative $H_1: \theta \neq 0$. The study is to be designed to have 90% power at the alternative $\theta_1 = 0.5$ with a type-1 error rate of 5%.

Start East afresh. Click **Continuous: Regression** on the **Design** tab and then click **Single-Arm Design: Linear Regression - Single Slope**.



This will launch a new input window. Select the **2-Sided** for **Test Type**. Enter 0.05 and 0.9 for **Type I Error (α)** and **Power**, respectively. Enter the values of $\theta_0 = 0$,

19 Normal Superiority Regression

$$\theta_1 = 0.5, \sigma_x = 9, \text{ and } \sigma_\epsilon = 15.$$

Click **Compute**. This will calculate the sample size for this design and the output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed sample size (119 subjects) is highlighted in yellow.

| ID | Design Type | Test Type | Specified α | Power | Sample Size | Std. Dev. | Std. Dev. of Residual | Null Hypothesis Slope | Alternative Hypothesis Slope |
|-------|-------------|-----------|--------------------|-------|-------------|-----------|-----------------------|-----------------------|------------------------------|
| Des 1 | Superiority | 2-Sided | 0.05 | 0.901 | 119 | 9 | 15 | 0 | 0.5 |

Des 1 requires 119 subjects in order to attain 90% power. Select this design by clicking anywhere along the row in the **Output Preview** and click . Some of the design

details will be displayed in the upper pane, labeled as **Output Summary**.

| Des 1 | |
|------------------------------|-------------|
| Mnemonic | MN-RG-1SL |
| Test Parameters | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.901 |
| Sample Size | |
| Maximum | 119 |
| Other Parameters | |
| Std. Dev. | 9 |
| Std. Dev. of Residual | 15 |
| Null Hypothesis Slope | 0 |
| Alternative Hypothesis Slope | 0.5 |

In the **Output Preview** toolbar, click  to save this design to Wbk1 in the **Library**. Now double-click on Des 1 in **Library**. You will see a summary of the design.

Design: Continuous Endpoint: Regression Model - Single Arm Design - Linear Regression - Single Slope

Model : $Y_i = \gamma + \theta X_i + \epsilon_i$

| Test Parameters | |
|--|----------------|
| Design ID | Des1 |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.901 |
| Sample Size (n) | To be Computed |
| Model Parameters | |
| Slope under Null Hypothesis(θ_0) | 0 |
| Slope under Alternative Hypothesis(θ_1) | 0.5 |
| Std. Dev. of Covariates (σ) | 9 |
| Std. Dev. of Residual (σ_2) | 15 |

Sample Size Information

Sample Size (n) 119

Critical Points

Lower Critical Point: -1.98
Upper Critical Point: 1.98

19.2 Linear Regression for Comparing Two Slopes

19.2.1 Trial Design

In some experimental situations, we are interested in comparing the slopes of two regression lines. The regression model relates the response variable Y to the explanatory variable X using the model $Y_{il} = \gamma + \theta_i X_{il} + \epsilon_{il}$, where the error ϵ_{il} has a normal distribution with mean zero and an unknown variance σ_ϵ^2 for Subject l in Treatment i , $i = c, t$ and $l = 1, \dots, n_i$. Let σ_{xc}^2 and σ_{xt}^2 denote the variance of the explanatory variable X for control (c) and treatment (t), respectively. We are interested in testing the equality of the slopes θ_c and θ_t . Thus we test the null hypothesis

19 Normal Superiority Regression

$H_0: \theta_c = \theta_t$ against the two-sided alternative $H_1: \theta_c \neq \theta_t$ or a one-sided alternative hypothesis $H_1: \theta_c < \theta_t$ or $H_1: \theta_c > \theta_t$.

Let $\hat{\theta}_c$ and $\hat{\theta}_t$ denote the estimates of θ_c and θ_t , and let $\hat{\sigma}_\epsilon^2$, $\hat{\sigma}_{xc}^2$, and $\hat{\sigma}_{xt}^2$, denote the estimates of σ_ϵ^2 , σ_{xc}^2 , and σ_{xt}^2 , based on n_c and n_t observations, respectively. The variance of $\hat{\theta}_i$ is

$$\sigma_i^2 = \frac{\sigma_\epsilon^2}{n_i \sigma_{xi}^2}.$$

Let $n = n_c + n_t$ and let $r = n_t/n$. Then, the test statistic is

$$Z_j = \frac{n^{1/2}(\hat{\theta}_t - \hat{\theta}_c)}{\hat{\sigma}_\epsilon \left(\frac{1}{(1-r)\hat{\sigma}_{xc}^2} + \frac{1}{r\hat{\sigma}_{xt}^2} \right)^{1/2}}. \quad (19.3)$$

19.2.1 Trial Design

We want to design the study so the power is attained for specified values of θ_c and θ_t . The power depends on $\theta_t, \theta_c, \sigma_{xc}, \sigma_{xt}$, and σ_ϵ through $\theta_t - \theta_c, \sigma_{xc}/\sigma_\epsilon$, and $\sigma_{xt}/\sigma_\epsilon$.

Suppose that a medication was found to have a response that depends on the level of a certain laboratory parameter. It was decided to develop a new formulation for which this interaction is decreased. The explanatory variable is the baseline value of the laboratory parameter. The study is designed with $\theta_t = 0.5, \theta_c = 1, \sigma_{xc} = \sigma_{xt} = 6$, and $\sigma_\epsilon = 10$. We examine the slopes of the two regressions by testing the null hypothesis $H_0: \theta_t = \theta_c$. Although we hope to decrease the slope, we test the null hypothesis against the two-sided alternative $H_1: \theta_t \neq \theta_c$.

Start East afresh. Click **Continuous: Regression** on the **Design** tab and then click **Parallel Design: Linear Regression - Difference of Slopes**.

This will launch a new input window. Select **2-Sided** for **Test Type**. Enter 0.05 and 0.9 for **Type I Error (α)** and **Power**, respectively. Select **Individual Slopes** for **Input Method**, and enter the values of $\theta_c = 1, \theta_t = 0.5, \sigma_{xc} = 6, \sigma_{xt} = 6$, and

$$\sigma_{\epsilon} = 10.$$

Click **Compute**. This will calculate the sample size for this design and the output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed sample size (469) is highlighted in yellow.

| ID | Design Type | Test Type | Specified α | Power | Sample Size | Slope under Control | Slope under Treatment | nt/nc | Std. Dev. of Covariates under Control | Std. Dev. of Covariates under Treatment | Std. Dev. of Residual |
|------|-------------|-----------|--------------------|-------|-------------|---------------------|-----------------------|-------|---------------------------------------|---|-----------------------|
| Des1 | Superiority | 2-Sided | 0.05 | 0.9 | 469 | 1 | 0.5 | 1 | 6 | 6 | 10 |

Des 1 requires 469 subjects in order to attain 90% power. Select this design by clicking anywhere along the row in the **Output Preview** and click . Some of the design

19 Normal Superiority Regression

details will be displayed in the upper pane, labeled as **Output Summary**.

| | Des 1 |
|---|-------------|
| Mnemonic | MN-RG-2SL |
| Test Parameters | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Sample Size | |
| Maximum | 469 |
| Other Parameters | |
| Slope under Control | 1 |
| Slope under Treatment | 0.5 |
| Std. Dev. of Covariates under Control | 6 |
| Std. Dev. of Covariates under Treatment | 6 |
| Std. Dev. of Residual | 10 |

19.3 Repeated Measures for Comparing Two Slopes

In many clinical trials, each subject is randomized to one of two groups, and responses are collected at various timepoints on the same individual over the course of the trial. In these “longitudinal” trials, we are interested in testing the equality of slopes, or mean response changes per unit time, between the treatment group (t) and the control group (c). A major difficulty associated with designing such studies is the fact that the data are independent across individuals, but the repeated measurements on the same individual are correlated. The sample size computations then depend on within – and between – subject variance components that are often unknown at the design stage. One way to tackle this problem is to use prior estimates of these variance components (also known as nuisance parameters) from other studies, or from pilot data.

Suppose each patient is randomized to either group c or group t . The data consist of a series of repeated measurements on the response variable for each patient over time. Let M denote the total number of measurements, inclusive of the initial baseline measurement, intended to be taken on each subject. These M measurements will be taken at times $v_m, m = 1, 2, \dots, M$, relative to the time of randomization, where $v_1 = 0$. A linear mixed effects model is usually adopted for analyzing such data. Let Y_{ilm} denote the response of subject l , belonging to group i , at time point v_m . Then the model asserts that

$$Y_{clm} = \gamma_c + \theta_c v_m + a_l + b_l v_m + e_{ilm} \tag{19.4}$$

for the control group, and

$$Y_{tlm} = \gamma_t + \theta_t v_m + a_l + b_l v_m + e_{ilm} \tag{19.5}$$

for the treatment group, where the random effect $(a_l, b_l)'$ is multivariate normal with mean $(0, 0)'$ and variance-covariance matrix

$$G = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix},$$

and the e_{lm} 's are all iid $N(0, \sigma_w^2)$. In this model, σ_w^2 denotes the “within – subject” variability, attributable to repeated measurements on the same subject, while G denotes the “between – subjects” variability, attributable to the heterogeneity of the population being studied.

Define

$$\delta = \theta_t - \theta_c$$

We are interested in testing

$$H_0: \delta = 0$$

against the two-sided alternative

$$H_1: \delta \neq 0$$

or against one-sided alternative hypotheses of the form

$$H_1: \delta > 0 \text{ or } H_1: \delta < 0$$

Let $(\hat{\theta}_C, \hat{\theta}_T)$ be the maximum likelihood estimates of (θ_C, θ_T) , based on an enrollment of (n_C, n_T) , respectively. The estimate of the difference of slopes is

$$\hat{\delta} = \hat{\theta}_T - \hat{\theta}_C \tag{19.6}$$

and its standard error is denoted by $se(\hat{\delta})$. The test statistic is the familiar Wald statistic

$$Z = \frac{\hat{\delta}}{se(\hat{\delta})} \tag{19.7}$$

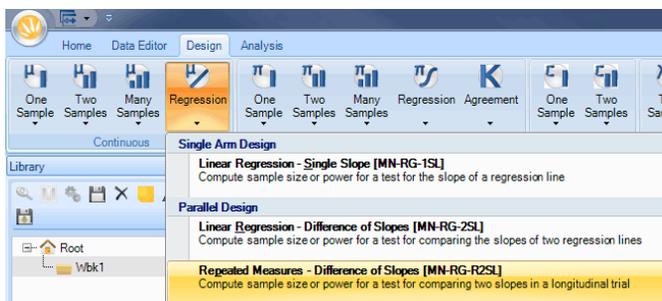
19.3.1 Trial Design

Consider a trial to compare an analgesic to placebo in the treatment of chronic pain using a 10 cm visual analogue scale (VAS). Measurements are taken on each subject at baseline and once a month for six months. Thus $M = 7$ and $S = 6$. It is assumed from past data that $\sigma_w = 4$ and $\sigma_b = 6$. We wish to test the null hypothesis $H_0: \theta_t = \theta_c$

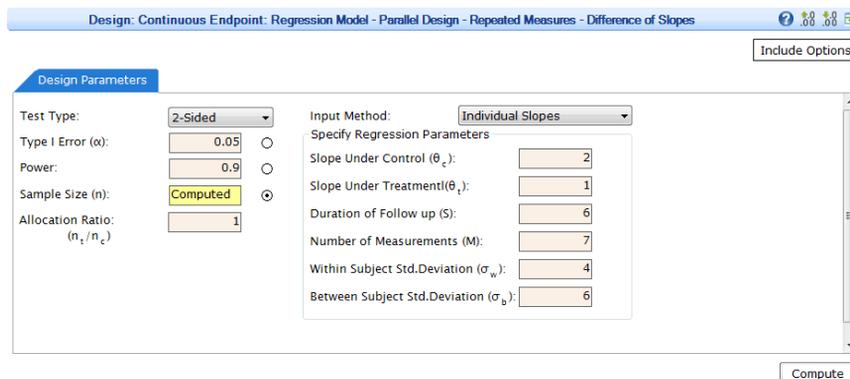
19 Normal Superiority Regression

with a two-sided level-0.05 test having 90% power to detect a 1 cm/month decline in slope, with $\theta_c = 2$ and $\theta_t = 1$ under H_1 .

Start East afresh. Click **Continuous: Regression** on the **Design** tab, and then click **Parallel Design: Repeated Measures - Difference of Slopes**.



This will launch a new input window. Select **2-Sided** for **Test Type**. Enter 0.05 and 0.9 for **Type I Error (α)** and **Power**, respectively. Select **Individual Slopes** for **Input Method**. Enter the values of $\theta_c = 2$, $\theta_t = 1$, **Duration of Follow up (S)** = 6, **Number of Measurements (M)** = 7, $\sigma_w = 4$, and $\sigma_b = 6$.



Click **Compute**. This will calculate the sample size for this design and the output is shown as a row in the **Output Preview** located in the lower pane of this window. The

computed sample size (1538) is highlighted in yellow.

| ▲ | ID | Test Type | Specified α | Power | Sample Size | Slope under Control | Slope under Treatment | nt/nc | Duration of Follow Up | Number of Measurements | Std. Dev. Within Subjects | Std. Dev. Between Subjects |
|---|------|-----------|--------------------|-------|-------------|---------------------|-----------------------|-------|-----------------------|------------------------|---------------------------|----------------------------|
| ☒ | Des1 | 2-Sided | 0.05 | 0.9 | 1538 | 2 | 1 | 1 | 6 | 7 | 4 | 6 |

Des 1 requires 1538 completers in order to attain 90% power. Select this design by clicking anywhere along the row in the **Output Preview** and click . Some of the design details will be displayed in the upper pane, labeled as **Output Summary**.

| Des 1 | |
|----------------------------|------------|
| Mnemonic | MN-RG-R2SL |
| Test Parameters | |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Sample Size | |
| Maximum | 1538 |
| Other Parameters | |
| Slope under Control | 2 |
| Slope under Treatment | 1 |
| Duration of Follow Up | 6 |
| Number of Measurements | 7 |
| Std. Dev. Within Subjects | 4 |
| Std. Dev. Between Subjects | 6 |

Volume 3 *Binomial and Categorical Endpoints*

| | | |
|----|--|-----|
| 20 | <i>Introduction to Volume 3</i> | 344 |
| 21 | <i>Tutorial: Binomial Endpoint</i> | 350 |
| 22 | <i>Binomial Superiority One-Sample</i> | 363 |
| 23 | <i>Binomial Superiority Two-Sample</i> | 394 |
| 24 | <i>Binomial Non-Inferiority Two-Sample</i> | 474 |
| 25 | <i>Binomial Equivalence Two-Sample</i> | 535 |
| 26 | <i>Binomial Superiority n-Sample</i> | 549 |
| 27 | <i>Multiple Comparison Procedures for Discrete Data</i> | 577 |
| 28 | <i>Multiple Endpoints-Gatekeeping Procedures for Discrete Data</i> | 601 |
| 29 | <i>Two-Stage Multi-arm Designs using p-value combination</i> | 621 |
| 30 | <i>Binomial Superiority Regression</i> | 644 |
| 31 | <i>Agreement</i> | 649 |

32 *Dose Escalation* 658

20 *Introduction to Volume 3*

This volume describes the procedures for discrete endpoints (binomial) applicable to one-sample, two-samples, many-samples, regression and agreement situations. All the three type of designs - superiority, non-inferiority and equivalence are discussed in detail.

Chapter 21 introduces you to East on the Architect platform, using an example clinical trial to test difference of proportions.

Chapter 22 deals with the design and interim monitoring of two types of tests involving binomial response rates that can be described as superiority one sample situation. Section 22.1 discusses designs in which an observed binomial response rate is compared to a fixed response rate, possibly derived from historical data. Section 22.2 deals with McNemar's test for comparing matched pairs of binomial responses. Chapter 38 discusses in detail the Simon's Two stage design.

Chapter 23 discusses the superiority two-sample situation where the aim is to compare independent samples from two populations in terms of the proportion of sampling units presenting a given trait. East supports the design and interim monitoring of clinical trials in which this comparison is based on the difference of proportions, the ratio of proportions, or the odds ratio of the two populations, common odds ratio of the two populations. The four cases are discussed in Sections 23.1, 23.2, 23.3 and 23.4, respectively. Section 23.5 discusses the Fisher's exact test for single look design.

Chapter 24 presents an account of designing and monitoring non-inferiority trials in which the non-inferiority margin is expressed as either a difference, a ratio, or an odds ratio of two binomial proportions. The difference is examined in Section 24.1. This is followed by two formulations for the ratio: the Wald formulation in Section 24.2 and the Farrington-Manning formulation in Section 24.3. The odds ratio formulation is presented in Section 24.4.

Chapter 25 narrates the details of the design and interim monitoring in equivalence two-sample situation where the goal is neither establishing superiority nor non-inferiority, but equivalence. Examples of this include showing that an aggressive therapy yields a similar rate of a specified adverse event to the established control, such as the bleeding rates associated with thrombolytic therapy or cardiac outcomes with a new stent.

Chapter 26 details the design and interim monitoring superiority k-sample experimental situations where there are several binomial distributions indexed by an

ordinal variable and where it is required to examine changes in the probabilities of success as the levels of the indexing variable changes. Examples of this include the examination of a dose-related presence of a response or a particular side effect, dose-related tumorigenicity, or presence of fetal malformations relative to levels of maternal exposure to a particular toxin, such as alcohol, tobacco, or environmental factors.

Chapter 27 details the Multiple Comparison Procedures (MCP) for discrete data. It is often the case that multiple objectives are to be addressed in one single trial. These objectives are formulated into a family of hypotheses. Multiple comparison (MC) procedures provide a guard against inflation of type I error while testing these multiple hypotheses. East supports several parametric and p-value based MC procedures. This chapter explains how to design a study using a chosen MC procedure that strongly maintains FWER.

Chapter 30 describes how East may be used to design and monitor two-arm randomized clinical trials with a binomial endpoints, while adjusting for the effects of covariates through the logistic regression model. These methods are limited to binary and categorical covariates only. A more general approach, not limited to categorical covariates, is to base the design on statistical information rather than sample size. This approach is further explained in Chapter 59

Chapter 31 discusses the tests available to check the inter-rater reliability. In some experimental situations, to check inter-rater reliability, independent sets of measurements are taken by more than one rater and the responses are checked for agreement. For a binary response, Cohen's Kappa test for binary ratings can be used to check inter-rater reliability.

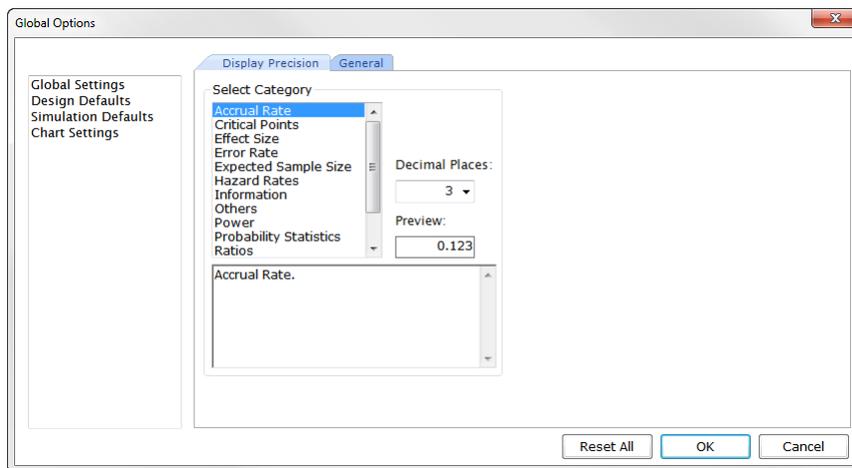
Chapter 32 deals with the design, simulation, and interim monitoring of Phase 1 dose escalation trials. One of the primary goals of Phase I trials in oncology is to find the maximum tolerated dose (MTD). Sections 32.1, 32.2, 32.3 and 32.4 discusses the four commonly used dose escalation methods - 3+3, Continual Reassessment Method (CRM), modified Toxicity Probability Interval (mTPI) and Bayesian Logistic Regression Model (BLRM).

20 Introduction to Volume 3

20.1 Settings



Click the **Global Options** icon in the **Home** menu to adjust default values in East 6.

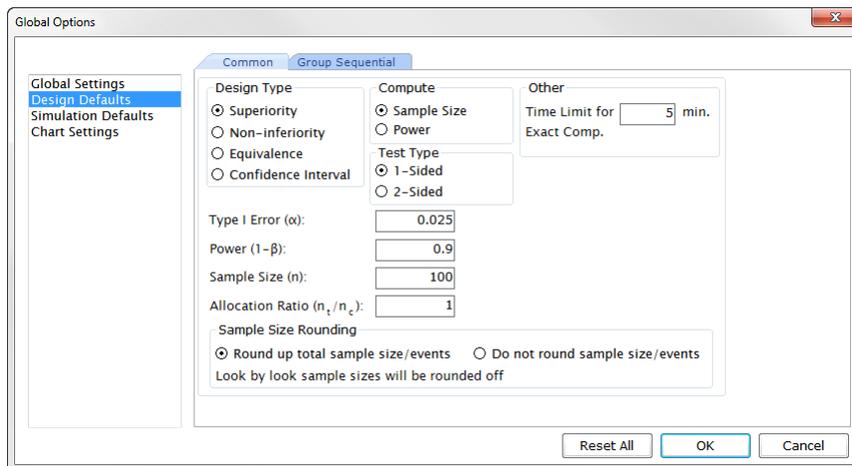


The options provided in the **Display Precision** tab are used to set the decimal places of numerical quantities. The settings indicated here will be applicable to all tests in East 6 under the **Design** and **Analysis** menus.

All these numerical quantities are grouped in different categories depending upon their usage. For example, all the average and expected sample sizes computed at simulation or design stage are grouped together under the category "Expected Sample Size". So to view any of these quantities with greater or lesser precision, select the corresponding category and change the decimal places to any value between 0 to 9.

The **General** tab has the provision of adjusting the paths for storing workbooks, files, and temporary files. These paths will remain throughout the current and future sessions even after East is closed. This is the place where we need to specify the installation directory of the R software in order to use the feature of R Integration in East 6.

The **Design Defaults** is where the user can change the settings for trial design:



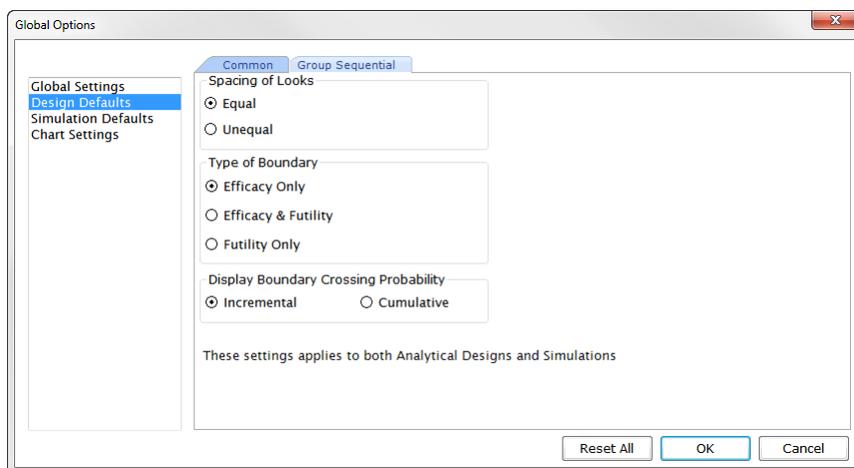
Under the **Common** tab, default values can be set for input design parameters.

You can set up the default choices for the design type, computation type, test type and the default values for type-I error, power, sample size and allocation ratio. When a new design is invoked, the input window will show these default choices.

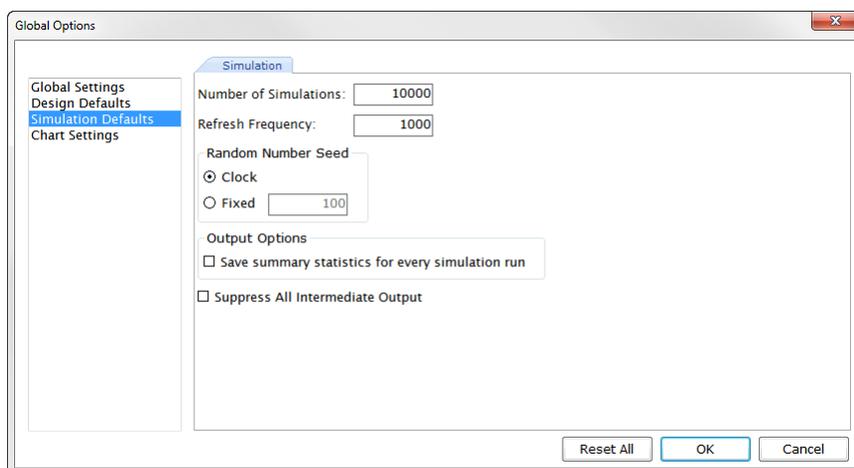
- **Time Limit for Exact Computation**
 This time limit is applicable only to exact designs and charts. Exact methods are computationally intensive and can easily consume several hours of computation time if the likely sample sizes are very large. You can set the maximum time available for any exact test in terms of minutes. If the time limit is reached, the test is terminated and no exact results are provided. Minimum and default value is 5 minutes.
- **Type I Error for MCP**
 If user has selected 2-sided test as default in global settings, then any MCP will use half of the alpha from settings as default since MCP is always a 1-sided test.
- **Sample Size Rounding**
 Notice that by default, East displays the integer sample size (events) by rounding up the actual number computed by the East algorithm. In this case, the look-by-look sample size is rounded off to the nearest integer. One can also see the original floating point sample size by selecting the option "Do not round sample size/events".

20 Introduction to Volume 3

Under the **Group Sequential** tab, defaults are set for boundary information. When a new design is invoked, input fields will contain these specified defaults. We can also set the option to view the Boundary Crossing Probabilities in the detailed output. It can be either Incremental or Cumulative.



Simulation Defaults is where we can change the settings for simulation:

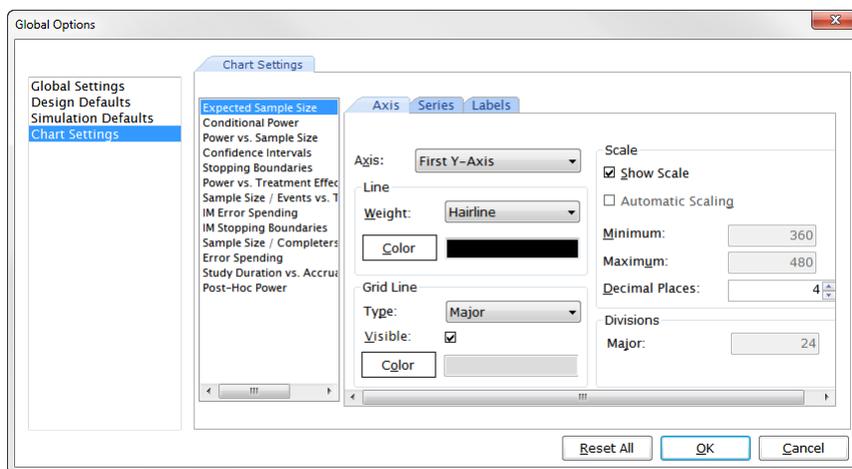


If the checkbox for "Save summary statistics for every simulation" is checked, then East simulations will by default save the per simulation summary data for all the

simulations in the form of a case data.

If the checkbox for "Suppress All Intermediate Output" is checked, the intermediate simulation output window will be always suppressed and you will be directed to the **Output Preview** area.

The **Chart Settings** allows defaults to be set for the following quantities on East6 charts:



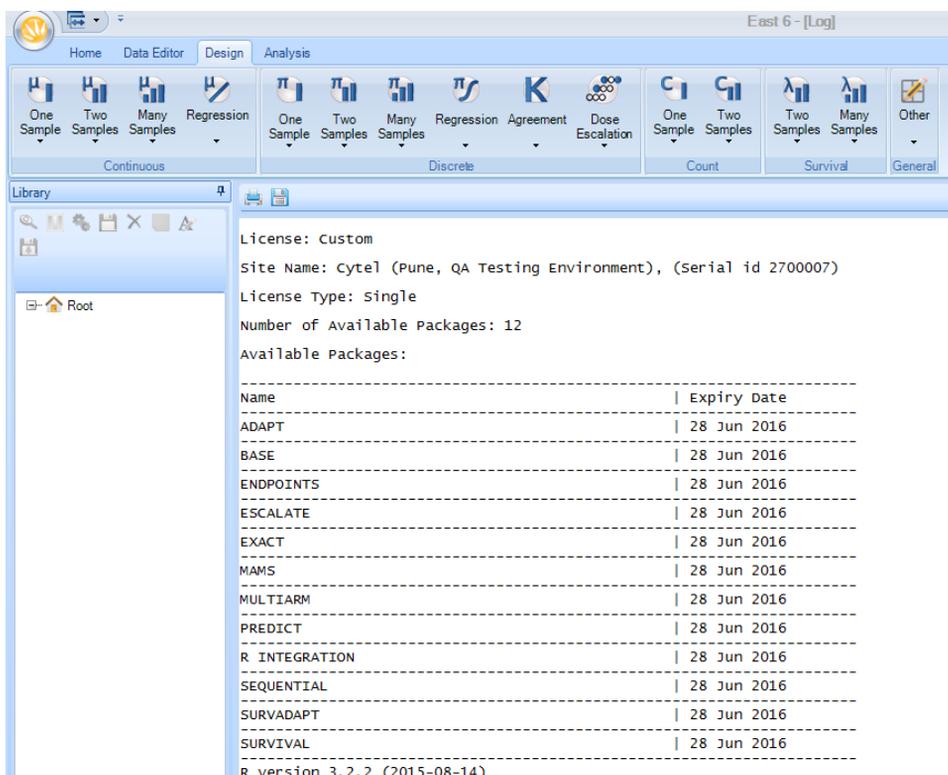
We suggest that you do not alter the defaults until you are quite familiar with the software.

21 *Tutorial: Binomial Endpoint*

This tutorial introduces you to East on the Architect platform, using an example clinical trial to test difference of proportions.

21.1 *Fixed Sample Design*

When you open East, you will see the following screen below.



By default, the Design tab in the ribbon will be active. The items on this tab are grouped under the following categories of endpoints: Continuous, Discrete, Count, Survival, and General. Click **Discrete: Two Samples**, and then **Parallel Design: Difference of Proportions**.

The following input window will appear.

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

Design Type: Superiority Number of Looks: 1

Test Parameters

Test Type: 1-Sided

Type I Error (α): 0.025

Power: 0.9

Sample Size (n): Computed

Allocation Ratio: 1
(n_t/n_c)

Specify Proportion Response
 Prop. under Control (π_c): 0.1

Specify Alternative Hypothesis
 Prop. under Treatment (π_t): 0.5
 Diff. in Prop. ($\delta_1 = \pi_t - \pi_c$): 0.4

Perform Exact Computations

Specify Variance
 Pooled Estimate
 Unpooled Estimate

Use Casagrande-Pike-Smith
 Correction (Ignored if alloc. ratio is not 1)

Assurance (Probability of Success):

By default, the radio button for **Sample Size (n)** is selected, indicating that it is the variable to be computed. The default values shown for **Type I Error** and **Power** are 0.025 and 0.9. Keep the same for this design. Since the default inputs provide all of the necessary input information, you are ready to compute sample size by clicking the **Compute** button. The calculated result will appear in the **Output Preview** pane, as shown below.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | π_c | Prop. Treatment (Alt.) | δ_1 | Variance |
|------|-------------|--------------|-----------|--------------------|-------|-------|-------------|---------|------------------------|------------|-------------------|
| Des1 | Superiority | 1 | 1-Sided | 0.025 | 0.902 | 1 | 45 | 0.1 | 0.5 | 0.4 | Unpooled Estimate |

This single row of output contains relevant details of inputs and the computed result of total sample size (and total completers) of 45. Select this row and save it in the **Library** under a workbook by clicking icon. Select this node in the Library, and click icon to display a summary of the design details in the upper pane

21 Tutorial: Binomial Endpoint

(known as **Output Summary**).

| Wbk1:Des1 | |
|--|-------------------|
| Mnemonic | PN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.902 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.1 |
| Proportion under Treatment (π_t) | 0.5 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0.4 |
| Variance | Unpooled Estimate |
| Sample Size | |
| Maximum | 45 |

The discussion so far gives you a quick feel of the software for computing sample size for a single look design. We will describe further features in an example for a group sequential design in the next section.

21.2 Group Sequential Design for a Binomial Superiority Trial

21.2.1 Study Background

Design objectives and interim results from CAPTURE, a prospective randomized trial of placebo versus Abciximab for patients with refractory unstable angina were presented at a workshop on clinical trial data monitoring committees (Anderson, 2002). The primary endpoint was reduction in death or MI within 30 days of entering the study. The study was designed for 80% power to detect a reduction in the event rate from 15% on the placebo arm to 10% on the Abciximab arm. A two-sided test with a type-1 error of 5% was used. We will illustrate various design, simulation and interim monitoring features of East for studies with binomial endpoints with the help of this example.

Let us modify Des1 to enter above inputs and create a group sequential design for CAPTURE trial. Select the node for Des1 in the **Library** and click the  icon. This will take you back to the input window of Des1. Alternatively, you can also click



the  button on the left hand bottom of East screen to go to the latest

input window.

Select **2-Sided** for **Test Type**, enter **0.05** for **Type I Error**, **0.8** for **Power**, specify the **Prop. under Control** be **0.15**, the **Prop. under Treatment** to be **0.1**. Next, change the **Number of Looks** to be **3**. You will see a new tab, **Boundary Info**, added to the input dialog box.

Design Type: Superiority Number of Looks: 3

Test Parameters

Test Type: 1-Sided

Type I Error (α): 0.05

Power: 0.8

Sample Size (n): Computed

Allocation Ratio: 1 (n_1/n_2)

Specify Proportion Response

Prop. under Control (π_c): 0.15

Specify Alternative Hypothesis

Prop. under Treatment (π_t): 0.1

Diff. in Prop. ($\delta_1 = \pi_t - \pi_c$): -0.05

Specify Variance

Pooled Estimate

Unpooled Estimate

Use Casagrande-Pike-Smith Correction (ignored if alloc. ratio is not 1)

Assurance (Probability of Success):

Click the **Boundary Info** tab, and you will see the following screen. On this tab, you can choose whether to specify stopping boundaries for efficacy, or futility, or both. For this trial, choose efficacy boundaries only, and leave all other default values. We will implement the Lan-Demets (O'Brien-Fleming) spending function, with equally spaced looks.

Design Type: Superiority Number of Looks: 3

Boundary

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.05

Futility

Boundary Family: None

Spacing of Looks Equal Unequal

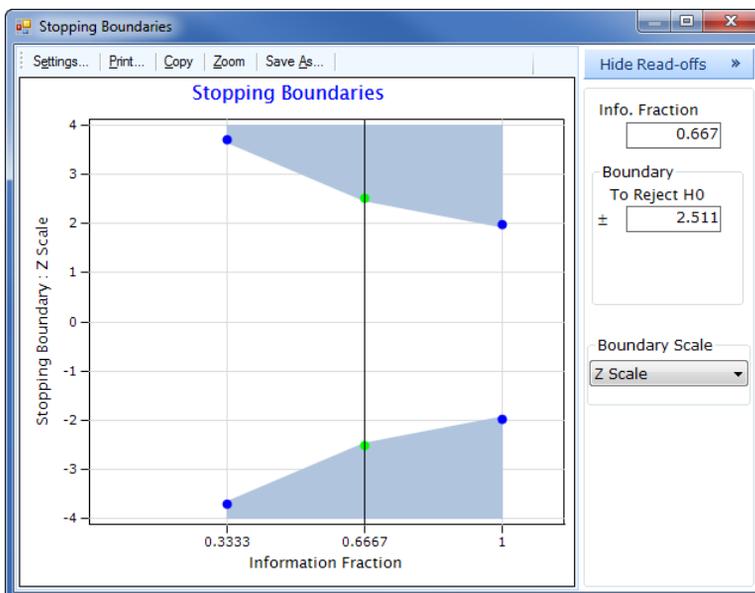
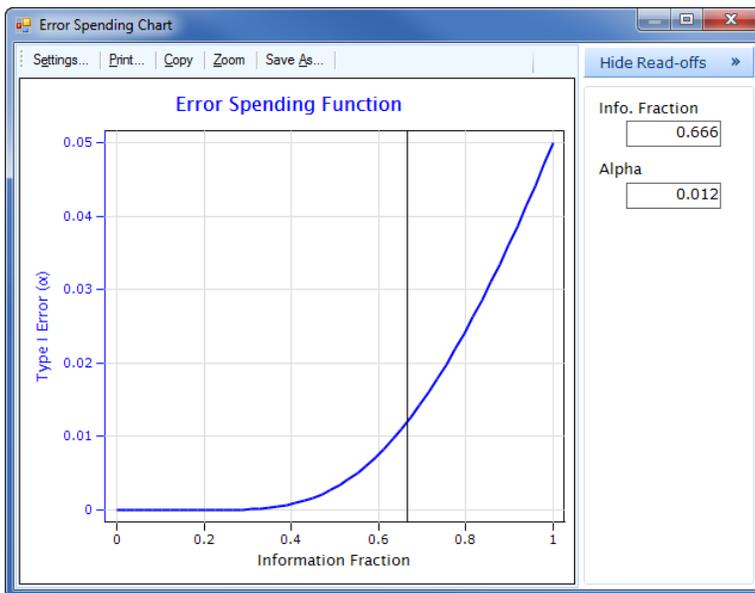
Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.333 | 0.001 | -3.200 |
| 2 | 0.667 | 0.016 | -2.141 |
| 3 | 1.000 | 0.050 | -1.695 |

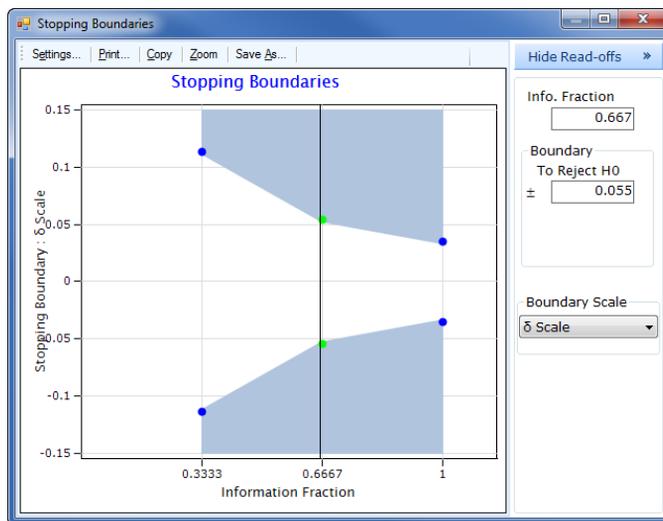
On the **Boundary Info** tab, click on the icons  or , to generate the

21 Tutorial: Binomial Endpoint

following charts.



You can also view these boundaries on different scales like δ scale or p-value scale. Select the desired scale from the dropdown. Let us see the boundaries on δ scale.



Click **Compute**. This will add another row for Des2 in the **Output Preview** area.

The maximum sample size required under this design is 1384. The expected sample sizes under H0 and H1 are 1378 and 1183, respectively. Click  in the **Output Preview** toolbar to save this design to Wbk1 in the **Library**. Double-click on Des2 to generate the following output.

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Test Parameters | |
|-----------------------------------|-------------------|
| Design ID | Des2 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.8 |
| Model Parameters | |
| Prop. under Control (π_c) | 0.15 |
| Prop. under Treatment (π_t) | 0.1 |
| $\delta = \pi_t - \pi_c$ | |
| Under H0 | 0 |
| Under H1 | -0.05 |
| Allocation Ratio (n/n_c) | 1 |
| Variance | Unpooled Estimate |
| Casagrande-Pike-Smith Correction | Not Applied |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|-----------------------------------|-------------|---------------|----------|
| Sample Size (n) | | | |
| Maximum | 692 | 692 | 1384 |
| Expected H1 | 591.548 | 591.13 | 1182.678 |
| Expected H0 | 689.166 | 689.154 | 1378.32 |
| Maximum Information (I): 3181.609 | | | |

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | | Boundary Crossing Probability (Incremental) | | | |
|--------|--------------------------|-----------------|---------------------------|------------|--------|---|----------|----------|-------|
| | | | | Efficacy Z | | Under H0 | | Under H1 | |
| | | | | Upper | Lower | Upper | Lower | Upper | Lower |
| 1 | 0.333 | 461 | 2.059E-4 | 3.712 | -3.712 | 1.029E-4 | 1.029E-4 | 4.661E-8 | 0.019 |
| 2 | 0.667 | 923 | 0.012 | 2.511 | -2.511 | 0.006 | 0.006 | 7.349E-7 | 0.4 |
| 3 | 1 | 1384 | 0.05 | 1.993 | -1.993 | 0.019 | 0.019 | 6.585E-7 | 0.382 |

21 Tutorial: Binomial Endpoint

21.2.2 Creating multiple designs easily

In East, it is easy to create multiple designs by inputting multiple parameter values. In the trial described above, suppose we want to generate designs for all combinations of the following parameter values: **Power** = 0.8, 0.9, and **Difference in Proportions** = -0.04, -0.03, -0.02, -0.01. The number of such combinations is $2 \times 4 = 8$.

East can create all 8 designs by a single specification in the input dialog box. Select Des2 and click  icon. Enter the above values in the **Test Parameters** tab as shown below. The values of **Power** have been entered as a list of comma-separated values, while **Difference in Proportions** has been entered as a colon-separated range of values: -0.04 to -0.01 in steps of 0.01.

Test Type: 2-Sided

Type I Error (α): 0.05

Power: 0.8, 0.9

Sample Size (n): Computed

Allocation Ratio: 1
(n_1 / n_2)

Specify Proportion Response

Prop. under Control (π_c): 0.15

Specify Alternative Hypothesis

Prop. under Treatment (π_t): Computed

Diff. in Prop. ($\delta_1 = \pi_t - \pi_c$): -0.04:-0.01

Specify Variance

Pooled Estimate

Unpooled Estimate

Use Casagrande-Pike-Smith Correction (Ignored if alloc. ratio is not 1)

Now click compute. East computes all 8 designs Des3-Des10, and displays them in the **Output Preview** as shown below. Click  to maximize the **Output Preview**.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt / nc | Sample Size | π_c | Prop. Treatment (Alt.) | δ_1 | Variance | Spacing of Looks | Efficacy Boundary | Expected SS (H0) | Expected SS (H1) |
|-------|-------------|--------------|-----------|--------------------|-------|---------|-------------|---------|------------------------|------------|-------------------|------------------|-------------------|------------------|------------------|
| Des1 | Superiority | 1 | 1-Sided | 0.025 | 0.902 | 1 | 45 | 0.1 | 0.5 | 0.4 | Unpooled Estimate | | | | |
| Des2 | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | 1384 | 0.15 | 0.1 | -0.05 | Unpooled Estimate | Equal | LD (OF) | 1378.32 | 1182.678 |
| Des3 | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | 2240 | 0.15 | 0.11 | -0.04 | Unpooled Estimate | Equal | LD (OF) | 2230.817 | 1914.336 |
| Des4 | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | 4118 | 0.15 | 0.12 | -0.03 | Unpooled Estimate | Equal | LD (OF) | 4101.115 | 3519.342 |
| Des5 | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | 9563 | 0.15 | 0.13 | -0.02 | Unpooled Estimate | Equal | LD (OF) | 9523.784 | 8172.868 |
| Des6 | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | 39413 | 0.15 | 0.14 | -0.01 | Unpooled Estimate | Equal | LD (OF) | 39251.361 | 33683.628 |
| Des7 | Superiority | 3 | 2-Sided | 0.05 | 0.9 | 1 | 2996 | 0.15 | 0.11 | -0.04 | Unpooled Estimate | Equal | LD (OF) | 2983.717 | 2402.552 |
| Des8 | Superiority | 3 | 2-Sided | 0.05 | 0.9 | 1 | 5508 | 0.15 | 0.12 | -0.03 | Unpooled Estimate | Equal | LD (OF) | 5485.41 | 4417.104 |
| Des9 | Superiority | 3 | 2-Sided | 0.05 | 0.9 | 1 | 12791 | 0.15 | 0.13 | -0.02 | Unpooled Estimate | Equal | LD (OF) | 12738.545 | 10257.754 |
| Des10 | Superiority | 3 | 2-Sided | 0.05 | 0.9 | 1 | 52714 | 0.15 | 0.14 | -0.01 | Unpooled Estimate | Equal | LD (OF) | 52497.802 | 42274.776 |

Select the first Des2 to Des4 using the Ctrl key, and click  to display a summary

of the design details in the upper pane, known as the **Output Summary**.

| | Des2 | Des3 | Des4 |
|--|-------------------|-------------------|-------------------|
| Mnemonic | PN-2S-DI | PN-2S-DI | PN-2S-DI |
| Test Parameters | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 3 | 3 | 3 |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.8 | 0.8 | 0.8 |
| Model Parameters | | | |
| Proportion under Control (π_c) | 0.15 | 0.15 | 0.15 |
| Proportion under Treatment (π_t) | 0.1 | 0.11 | 0.12 |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.05 | -0.04 | -0.03 |
| Variance | Unpooled Estimate | Unpooled Estimate | Unpooled Estimate |
| Allocation Ratio (n_t/n_c) | 1 | 1 | 1 |
| Boundary Parameters | | | |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) |
| Spacing of Looks | Equal | Equal | Equal |
| Sample Size | | | |
| Maximum | 1384 | 2240 | 4118 |
| Expected Under H0 | 1378.32 | 2230.817 | 4101.115 |
| Expected Under H1 | 1182.678 | 1914.336 | 3519.342 |

Des2 is already saved in the workbook. We will use this design for simulation and interim monitoring, as described below. Now that you have saved Des2, delete all designs from the **Output Preview** before continuing, by selecting all designs with the Shift key, and clicking  in the toolbar.

21.2.3 Simulation

Right-click Des2 in the **Library**, and select **Simulate**. Alternatively, you can select Des2 and click the  icon.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

Number of Looks: 3

Simulation Parameters | Response Generation Info | Simulation Control Info

Trial Type: Superiority
 Test Type: 2-Sided
 Sample Size (n): 1384

Specify Variance
 Pooled Estimate
 Unpooled Estimate

| Look # | Info. Fraction | Cum. α Spent | | Efficacy Z | |
|--------|----------------|---------------------|-------|------------|--------|
| | | Upper | Lower | Upper | Lower |
| 1 | 0.333 | 0.000 | 0.000 | 3.712 | -3.712 |
| 2 | 0.667 | 0.006 | 0.006 | 2.511 | -2.511 |
| 3 | 1.000 | 0.025 | 0.025 | 1.993 | -1.993 |

Restore Original Design

We will carry out a simulation of Des2 to check whether it preserves the specified power. Click **Simulate**. East will execute by default 10000 simulations with the specified inputs. Close the intermediate window after examining the results. A row labeled as Sim1 will be added in the **Output Preview**.

21 Tutorial: Binomial Endpoint

Click the  icon to save this simulation to the **Library**. A simulation sub-node, Sim1, will be added under Des2 node. Double clicking on this node will display the detailed simulation output in the work area.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Simulation Parameters | |
|-----------------------------------|-------------------|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 2-Sided |
| Sample Size (n) | 1384 |
| Variance | Unpooled Estimate |
| Avg. Power at Termination | 0.805 |
| Response Generation Parameters | |
| Prop. under Control (π_c) | 0.15 |
| Prop. under Treatment (π_t) | 0.1 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 461 |
| 2 | 923 |
| 3 | 1384 |
| Average | 1182.252 |

Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries Efficacy | | Stopping For | | Total Simulations | |
|--------|-----------------|---------------------|--------|----------------|----------------|-------------------|---------|
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| | | 1 | 461 | 3.712 | -3.712 | 0 | 146 |
| 2 | 923 | 2.511 | -2.511 | 0 | 4084 | 4084 | 40.840% |
| 3 | 1384 | 1.993 | -1.993 | 0 | 3816 | 5770 | 57.700% |
| Total | | | | 0 | 8046 | 10000 | |
| % | | | | 0.000% | 80.460% | | |

Simulation Seed and Elapsed Time

Starting Seed: 68247669
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:10

In 80.46% of the simulated trials, the null hypothesis was rejected. This tells us that the design power of 80% is achieved. **Simulations** is a tool which can be used to assess the study design under various scenarios. The next section will explore interim monitoring with this design.

21.2.4 Interim Monitoring

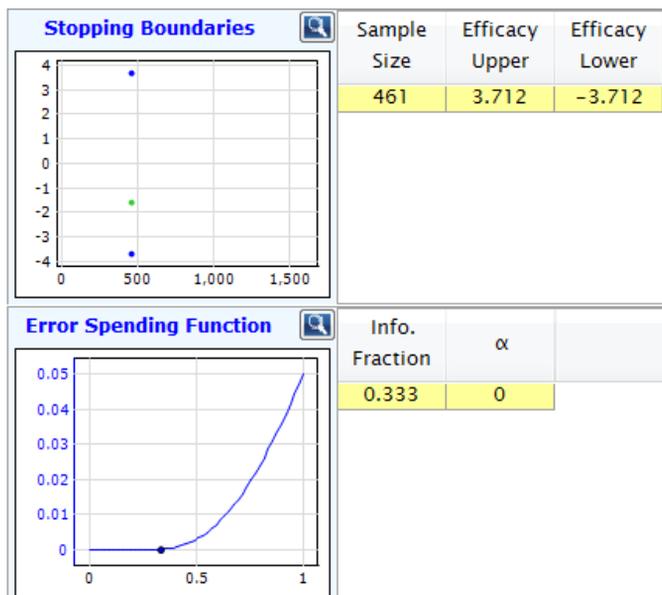
Right-click Des2 in the **Library** and select **Interim Monitoring**. Click the **Enter Interim Data** to open the **Test Statistic Calculator**. Suppose that after 461 subjects, at the first look, you have observed 34 out of 230 responding on Control arm and 23 out of 231 responding on Treatment arm. The calculator computes the difference in proportions as -0.048 and its standard error of 0.031 .

Click **OK** to update the IM Dashboard.

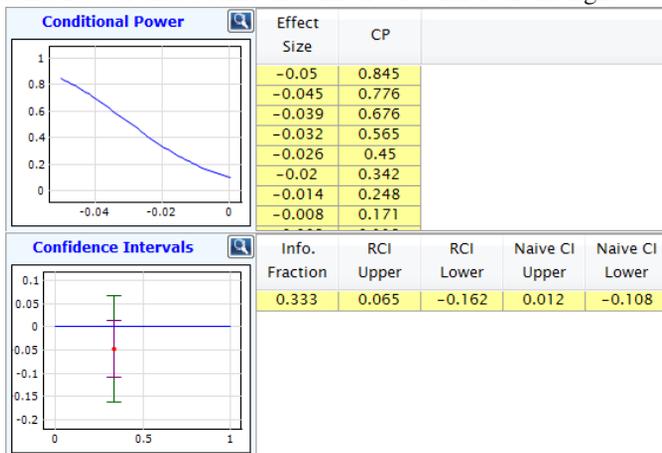
| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | | 95% RCI for δ | | Repeat ... p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|--------|----------------------|-------|--------------------|-------|------------------|
| | | | | | | Upper | Lower | Lower | Upper | | | |
| 1 | 0.333 | 461 | -1.578 | -0.048 | 0.031 | 3.712 | -3.712 | -0.162 | 0.065 | 0.545 | 0.823 | 0.705 |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |

The **Stopping Boundaries** and **Error Spending Function** charts on the left:

21 Tutorial: Binomial Endpoint



The **Conditional Power** and **Confidence Intervals** charts on the right:



Suppose that after 923 subjects, at the second look, you have observed 69 out of 461 responding on Control arm and 23 out of 462 responding on Treatment arm. The

calculator computes the difference in proportions as -0.1 and its standard error of 0.019 .

| | Control | Treatment |
|-------------------------|---------|-----------|
| Cumulative Sample Size: | 461 | 462 |
| Cumulative Response: | 69 | 23 |

Cumulative Sample Size: 923

Estimate of δ : -0.1

$\delta = (\pi_t - \pi_c)$

Standard Error of Estimate of δ : 0.019

Test Statistic: -5.135

Click **Recalc**, and then **OK** to update the IM Dashboard. In this case, a boundary has been crossed, and the following window appears.

Since the value of Test Statistic is \leq the critical point for efficacy, H_0 is rejected.

Although boundary has been crossed, East gives you choice either to stop the study or to continue entering further looks. Please make your decision.

Stop stop the study and bar further looks input

Continue allow the study to continue

Click **Stop** to complete the trial. The IM Dashboard will be updated accordingly, and a

21 Tutorial: Binomial Endpoint

table for **Final Inference** will be displayed as shown below.

| Final Inference | |
|--|--------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0 |
| Adj. Pt. Est. for δ | -0.094 |
| Adj. 95% CI for δ | |
| Upper Confidence Bound | -0.047 |
| Lower Confidence Bound | -0.135 |
| Post-Hoc Power | |

22 Binomial Superiority One-Sample

This chapter deals with the design, simulation, and interim monitoring of two types of tests involving binomial response rates. In Section 22.1, we discuss group sequential designs in which an observed binomial response rate is compared to a fixed response rate, possibly derived from historical data. Section 22.2 deals with McNemar’s test for comparing matched pairs of binomial responses in a group sequential setting.

22.1 Binomial One Sample

22.1.1 Trial Design

22.1.2 Trial Simulation

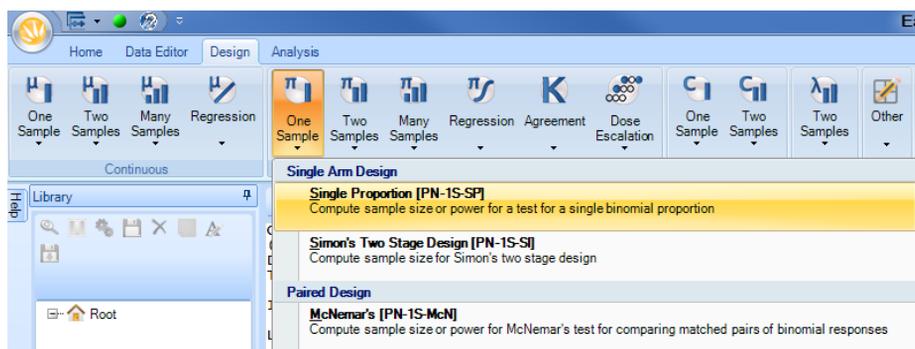
22.1.3 Interim Monitoring

In experimental situations, where the variable of interest has a binomial distribution, it may be of interest to determine whether the response rate π differs from a fixed value π_0 . Specifically we wish to test the null hypothesis $H_0: \pi = \pi_0$ against the two sided alternative hypothesis $H_1: \pi \neq \pi_0$ or against one sided alternatives of the form $H_1: \pi > \pi_0$ or $H_1: \pi < \pi_0$. The sample size, or power, is determined for a specified value of π which is consistent with the alternative hypothesis, denoted π_1 .

22.1.1 Trial Design

Consider the design of a single-arm oncology trial in which we wish to determine if the tumor response rate of a new cytotoxic agent is at least 15%. Thus, it is desired to test the null hypothesis $H_0: \pi = 0.15$ against the one-sided alternative hypothesis $H_1: \pi > 0.15$. We will design this trial with a one sided test that achieves 80% power at $\pi = \pi_1 = 0.25$ with a one-sided level 0.05 test.

Single-Look Design To begin, click **Design** tab, then **Single Sample** under **Discrete** group, and then click **Single Proportion**.



In the ensuing dialog box, choose the test parameters as shown below. We first consider a single-look design, so leave the default value for **Number of Looks** to 1. In the drop down menu, next to **Test Type** select 1-Sided. Enter 0.8 for **Power**. Enter

22 Binomial Superiority One-Sample

0.15 in the box next to **Prop. Response under Null** (π_0) and 0.25 in the box next to **Prop. Response under Alt** (π_1). This dialog box also asks us to specify whether we wish to standardize the test statistic (for performing the hypothesis test of the null hypothesis $H_0: \pi = 0.15$) with the null or the empirical variance. We will discuss the test statistic and the method of standardization in the next subsection. For the present, select the default radio button **Under Null Hypothesis**.

Now click **Compute**. The design is shown as a row in the **Output Preview** located in the lower pane of this window. The sample size required in order to achieve the desired 80% power is 91 subjects.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | Sample Size | π_0 | π_1 | Variance |
|------|-------------|--------------|-----------|--------------------|-------|-------------|---------|---------|-----------------------|
| Des1 | Superiority | 1 | 1-Sided | 0.05 | 0.801 | 91 | 0.15 | 0.25 | Under Null Hypothesis |

You can select this design by clicking anywhere on the row in the **Output Preview**. Click  icon to get the design output summary displayed in the upper pane. In the **Output Preview** toolbar, click  icon to save this design Des1 to workbook Wbk1 in the **Library**. If you hover the cursor over the node Des1 in the Library, a

tooltip will appear that summarizes the input parameters of the design.

The screenshot shows the 'Library' window with a tree view containing 'Root', 'Wbk1', and 'Des1'. A tooltip is displayed over 'Des1' with the following text:

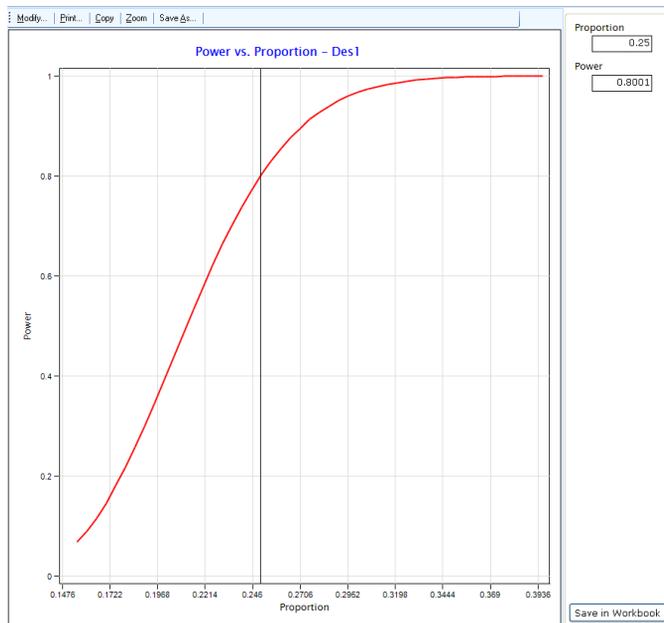
```

Design Type = Superiority
No. of Looks = 1
Test Type = 1-Sided
Specified  $\alpha$  = 0.05
Power = 0.801
Sample Size = 91
Prop. Response under Null ( $\pi_0$ ) = 0.15
Prop. Response under Alt. ( $\pi_1$ ) = 0.25
Variance of Standardized Test Statistic = Under Null Hypothesis
    
```

To the right, a table titled 'Wbk1:Des1' lists parameters:

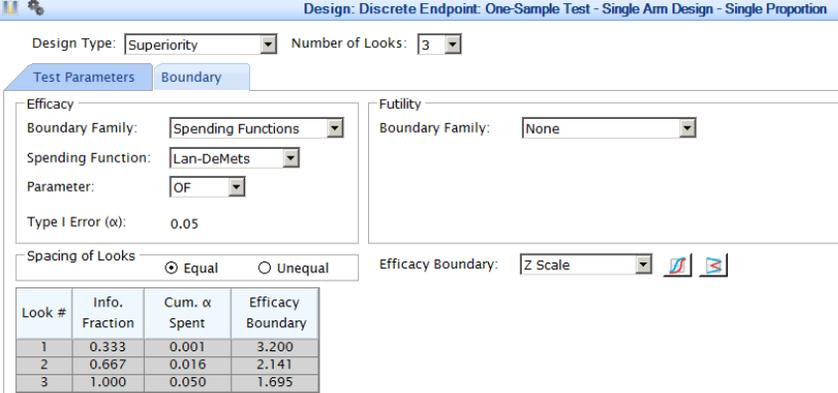
| Wbk1:Des1 | |
|---|-----------------------|
| Mnemonic | PN-1S-SP |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.801 |
| Model Parameters | |
| Prop. Response under Null (π_0) | 0.15 |
| Prop. Response under Alt. (π_1) | 0.25 |
| Variance of Standardized Test Statistic | Under Null Hypothesis |
| Sample Size | |
| Maximum | 91 |

With the design Des1 selected in the Library, click icon on the Library toolbar, and then click **Power vs. Treatment Effect (δ)**. The power curve for this design will be displayed. You can save this chart to the Library by clicking **Save in Workbook**. Alternatively, you can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...**. For now, you may close the chart before continuing.



22 Binomial Superiority One-Sample

Three-Look Design In order to reach an early decision and enter into comparative trials, let us plan to conduct this single-arm study as a group sequential trial with a maximum of 3 looks. Create a new design by selecting Des1 in the **Library**, and clicking the  icon on the **Library** toolbar. Change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. A new tab **Boundary** will appear. Clicking on this tab will reveal the stopping boundary parameters. By default, the **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The left side contains details for the **Efficacy** boundary, and the right side for the **Futility** boundary. By default, there is an efficacy boundary (to reject H0) selected, but no futility boundary (to reject H1). The **Boundary Family** specified is of the **Spending Functions** type. The default **Spending function** is the **Lan-DeMets** (Lan & DeMets, 1983), with **Parameter** OF (O'Brien-Fleming), which generates boundaries that are very similar, though not identical, to the classical stopping boundaries of O'Brien and Fleming (1979). Technical details of these stopping boundaries are available in Appendix F.



Design: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.05

Futility

Boundary Family: None

Spacing of Looks

Equal Unequal

Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|--------------|-------------------|
| 1 | 0.333 | 0.001 | 3.200 |
| 2 | 0.667 | 0.016 | 2.141 |
| 3 | 1.000 | 0.050 | 1.695 |

Return to the test parameters by clicking **Test Parameters** tab. The dialog box requires us to make a selection in the section labeled **Variance of Standardized Test Statistic**. We are being asked to specify to East how we intend to standardize the test statistic when we actually perform the hypothesis tests at the various monitoring time points. There are two options: **Under Null Hypothesis** and **Empirical Estimate**. To understand the difference between these two options, let $\hat{\pi}_j$ denote the estimate of π based on n_j observations, up to and including the j th monitoring time point.

Under Null Hypothesis The test statistic to be used for the interim monitoring is

$$Z_j^{(N)} = \frac{\hat{\pi}_j - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n_j}} \quad (22.1)$$

Empirical The test statistic to be used for the interim monitoring is

$$Z_j^{(E)} = \frac{\hat{\pi}_j - \pi_0}{\sqrt{\hat{\pi}_j(1 - \hat{\pi}_j)/n_j}} \quad (22.2)$$

The choice of variance should not make much of a difference to the type 1 error or power for studies in which the sample size is large. In the present case however, it might matter. We shall therefore examine both the options. First, we select the **Under Null Hypothesis** radio button.

Click **Compute** button to generate output for Design Des2. With Des2 selected in the **Output Preview**, click  icon to save Des2 to the **Library**. In order to see the stopping probabilities, as well as other characteristics, select Des2 in the **Library**, and click  icon. The cumulative boundary stopping probabilities are shown in the **Stopping Boundaries** table. We see that for Des2 the maximum sample size is 91 subjects, with 90 expected under the null hypothesis $H_0: \pi = 0.15$ and 73 expected when the true value is $\pi=0.25$.

Design: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion

| Test Parameters | |
|---------------------------------------|-----------------------|
| Design ID | Des2 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.801 |
| Model Parameters | |
| Prop. Response under Null (π_0) | 0.15 |
| Prop. Response under Alt. (π_1) | 0.25 |
| Variance of Std. Test Stat. | Under Null Hypothesis |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |

Sample Size Information

| Sample Size (n) | Total |
|---------------------------------|--------|
| Maximum | 91 |
| Expected H1 | 72.823 |
| Expected H0 | 90.48 |
| Maximum Information (I):485.333 | |

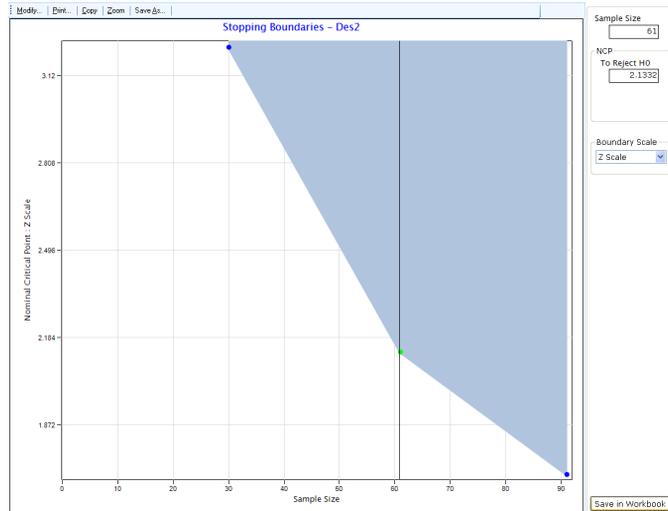
Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries Efficacy Z | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|-----------------------|---|----------|
| | | | | | Under H0 | Under H1 |
| 1 | 0.33 | 30 | 6.412E-4 | 3.22 | 6.412E-4 | 0.082 |
| 2 | 0.67 | 61 | 0.017 | 2.133 | 0.016 | 0.439 |
| 3 | 1 | 91 | 0.05 | 1.696 | 0.033 | 0.28 |

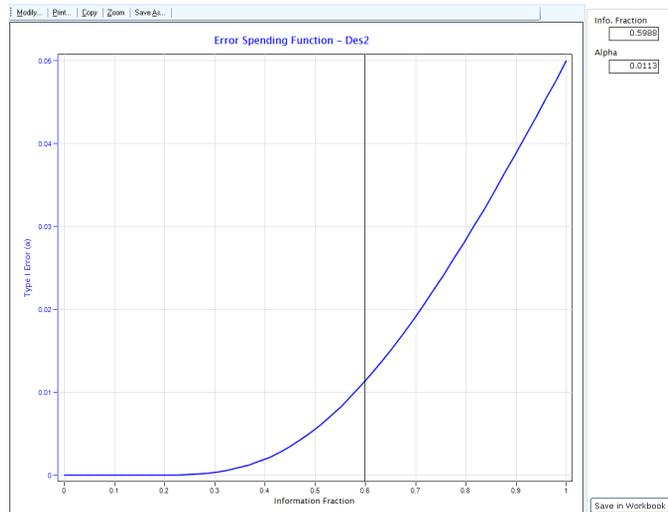
Close the Output window before continuing. The stopping boundary can be displayed by clicking on the  icon on the **Library** toolbar, and then clicking **Stopping**

22 Binomial Superiority One-Sample

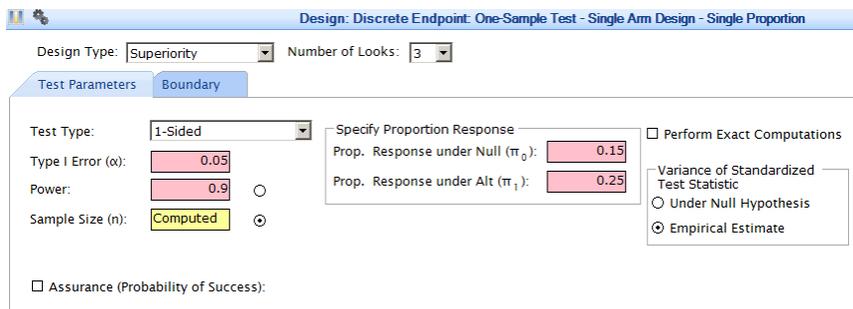
Boundaries. The following chart will appear.



To examine the error spending function, click  icon on the **Library** toolbar, and then click **Error Spending**. The following chart will appear.



To examine the impact of using the empirical variance to standardized test statistic, select Des2 in the **Library**, and click  icon on the **Library** toolbar. In the **Variance of Standardized Test Statistic** box, now select **Empirical Estimate**.



Design: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary

Test Type: 1-Sided

Type I Error (α): 0.05

Power: 0.9

Sample Size (n): Computed

Specify Proportion Response

Prop. Response under Null (π_0): 0.15

Prop. Response under Alt (π_1): 0.25

Perform Exact Computations

Variance of Standardized Test Statistic

Under Null Hypothesis

Empirical Estimate

Assurance (Probability of Success):

Next, click **Compute**. With Des3 selected in the **Output Preview**, click  icon. In the **Library**, select the nodes Des2 and Des3, by holding the Ctrl key, and then click  icon. The upper pane will display the summary details of the two designs side-by-side:

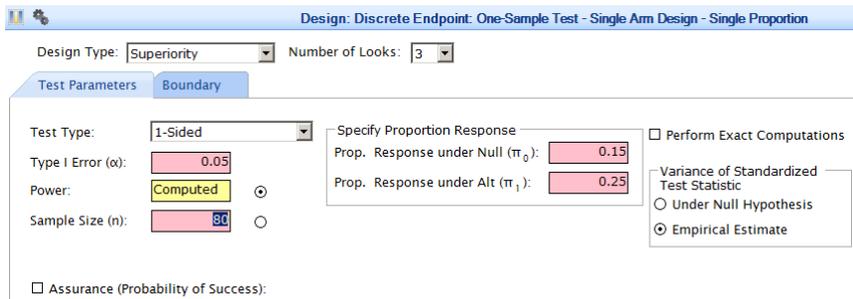
| | Wbk1:Des2 | Wbk1:Des3 |
|---|-----------------------|--------------------|
| Mnemonic | PN-1S-SP | PN-1S-SP |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 3 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.801 | 0.802 |
| Model Parameters | | |
| Prop. Response under Null (π_0) | 0.15 | 0.15 |
| Prop. Response under Alt. (π_1) | 0.25 | 0.25 |
| Variance of Standardized Test Statistic | Under Null Hypothesis | Empirical Estimate |
| Boundary Parameters | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) |
| Sample Size | | |
| Maximum | 91 | 119 |
| Expected Under H0 | 90.48 | 118.326 |
| Expected Under H1 | 72.823 | 98.803 |

The maximum sample size needed for 80% power is 119, and the expected sample size is 99 under the alternative hypothesis H_1 with $\pi_1 = 0.25$, if we intend to standardize the test statistic with the empirical variance. The corresponding maximum and

22 Binomial Superiority One-Sample

expected sample sizes if the null variance is to be used for the standardization are 91 and 73, respectively. Thus, for this configuration of design parameters, it would appear preferable to specify in advance that the test statistic will be standardized by the null variance. Evidently, this is the option with the smaller maximum and expected sample size. These results, however, are based on the large sample theory developed in Appendix B. Since the sample sizes in both Des2 and Des3 are fairly small, it would be advisable to verify that the power and type 1 error of both the plans are preserved by simulating these designs. We show how to simulate these plans in Section 22.1.2.

In some situations, the sample size is subject to external constraints. Then, the power can be computed for a specified maximum sample size. Suppose that in the above situation, using the observed estimates for the computation of the variance, the total sample size is constrained to be at most, 80 subjects. Select Des3 in the **Library** and click  on the **Library** toolbar. Change the selections in the ensuing dialog box so that the trial is now designed to compute power for a maximum sample size of 80 subjects, as shown below.



Design: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary

Test Type: 1-Sided

Type I Error (α): 0.05

Power: Computed

Sample Size (n): 80

Specify Proportion Response

Prop. Response under Null (π_0): 0.15

Prop. Response under Alt (π_1): 0.25

Perform Exact Computations

Variance of Standardized Test Statistic

Under Null Hypothesis

Empirical Estimate

Assurance (Probability of Success):

Click **Compute** button to generate the output for Design Des4. With Des4 selected in the **Output Preview**, click  icon. In the **Library**, select the nodes for Des2, Des3, and Des4 by holding the Ctrl key, and then click  icon. The upper pane

will display the summary details of the three designs side-by-side:

| | Wbk1-Des2 | Wbk1-Des3 | Wbk1-Des4 |
|---|-----------------------|--------------------|--------------------|
| Mnemonic | PN-1S-SP | PN-1S-SP | PN-1S-SP |
| Test Parameters | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 3 | 3 | 3 |
| Test Type | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.801 | 0.802 | 0.655 |
| Model Parameters | | | |
| Prop. Response under Null (π_0) | 0.15 | 0.15 | 0.15 |
| Prop. Response under Alt. (π_1) | 0.25 | 0.25 | 0.25 |
| Variance of Standardized Test Statistic | Under Null Hypothesis | Empirical Estimate | Empirical Estimate |
| Boundary Parameters | | | |
| Spacing of Looks | Equal | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) |
| Sample Size | | | |
| Maximum | 91 | 119 | 80 |
| Expected Under H0 | 90.48 | 118.326 | 79.548 |
| Expected Under H1 | 72.823 | 98.803 | 70.711 |

From this, we can see that Des4 has only 65.5 % power.

22.1.2 Trial Simulation

In Section 22.1.1, we created group sequential designs with two different assumptions for the manner in which the test would be standardized at the interim monitoring stage. Under Des2, we assumed that the null variance, and hence the test statistic (22.1) would be used for the interim monitoring. This plan required a maximum sample size of 91 subjects. Under Des3, we assumed that the empirical variance, and hence the test statistic (22.2) would be used for the interim monitoring. This plan required a maximum sample size of 119 subjects. Since the sample sizes for both plans are fairly small and the calculations involved the use of large sample theory, it would be wise to verify the operating characteristics of these two plans by simulation.

Select Des2 in the **Library**, and click the  icon from **Library** toolbar. Alternatively, right-click on Des2 node and select **Simulate**. A new Simulation

22 Binomial Superiority One-Sample

worksheet will appear.

| Look # | Info. Fraction | Cum. α Spent | Efficacy Z |
|--------|----------------|---------------------|------------|
| 1 | 0.330 | 0.001 | 3.220 |
| 2 | 0.670 | 0.017 | 2.133 |
| 3 | 1.000 | 0.050 | 1.696 |

Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled Sim1. Select Sim1 row in the **Output Preview** and click icon. Note that some of the simulation output details will be displayed in the upper pane. Click icon to save it to the **Library**. Double-click on Sim1 node in the **Library**. The simulation output details will be displayed.

Simulation: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion

| Test Parameters | |
|---------------------------------------|-----------------------|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 91 |
| Prop. Response under Null (π_0) | 0.15 |
| Variance | Under Null Hypothesis |
| Avg. Power at Termination | 0.802 |
| Response Generation Parameters | |
| Proportion Response (π) | 0.25 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 30 |
| 2 | 61 |
| 3 | 91 |
| Average | 70.419 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 30 | 3.22 | 1034 | 1034 | 10.340% |
| 2 | 61 | 2.133 | 4758 | 4758 | 47.580% |
| 3 | 91 | 1.696 | 2230 | 4208 | 42.080% |
| Total | | | 8022 | 10000 | |
| % | | | 80.220% | | |

Simulation Seed and Elapsed Time

Starting Seed: 71540
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:03

Upon running 10,000 simulations with $\pi = 0.25$ we obtain slightly over 80% power as shown above.

Next we run 10,000 simulations under H_0 by setting $\pi = 0.15$ in the choice of simulation parameters. Select Des2 in the **Library**, and click  icon from **Library** toolbar. Under the **Response Generation** tab, change the **Proportion Response** to 0.15. Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled Sim2.

Select Sim2 in the **Output Preview**. Click  icon to save it to the **Library**. Double-click on Sim2 in the **Library**. The simulation output details will be displayed.

Simulation: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion

| Test Parameters | |
|---------------------------------------|-----------------------|
| Simulation ID | Sim2 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 91 |
| Prop. Response under Null (π_0) | 0.15 |
| Variance | Under Null Hypothesis |
| Avg. Power at Termination | 0.067 |
| Response Generation Parameters | |
| Proportion Response (π) | 0.15 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

⊖ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 30 |
| 2 | 61 |
| 3 | 91 |
| Average | 89.797 |

⊖ Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 30 | 3.22 | 34 | 34 | 0.340% |
| 2 | 61 | 2.133 | 332 | 332 | 3.320% |
| 3 | 91 | 1.696 | 303 | 9634 | 96.340% |
| Total | | | 669 | 10000 | |
| % | | | 6.690% | | |

Simulation Seed and Elapsed Time

Starting Seed: 295871
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:03

We observe that 7% of these simulations reject the null hypothesis thereby confirming that these boundaries do indeed preserve the type 1 error (up to Monte Carlo accuracy).

Finally we repeat the same set of simulations for Des3. Select Des3 in the **Library**, and click  icon from **Library** toolbar. Upon running 10,000 simulations with

22 Binomial Superiority One-Sample

$\pi = 0.25$, we obtain 82% power.

Simulation: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion

| Test Parameters | |
|---------------------------------------|--------------------|
| Simulation ID | Sim3 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 119 |
| Prop. Response under Null (π_0) | 0.15 |
| Variance | Empirical Estimate |
| Avg. Power at Termination | 0.822 |
| Response Generation Parameters | |
| Proportion Response (π) | 0.25 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 40 |
| 2 | 79 |
| 3 | 119 |
| Average | 97.33 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | | Total Simulations | |
|--------|-----------------|------------|--------------|-------|-------------------|--|
| | | Efficacy | Efficacy | Count | % | |
| 1 | 40 | 3.189 | 263 | 263 | 2.630% | |
| 2 | 79 | 2.145 | 4898 | 4898 | 48.980% | |
| 3 | 119 | 1.694 | 3062 | 4839 | 48.390% | |
| Total | | | 8223 | 10000 | | |
| % | | | 82.230% | | | |

Simulation Seed and Elapsed Time

Starting Seed: 559158
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:04

However, when we run the simulations under $H_0: \pi = 0.15$, we obtain a type 1 error of about 3% instead of the specified 5% as shown below. While this ensures that the type 1 error is preserved, it also suggests that the use of the empirical variance rather than the null variance to standardize the test statistic might be problematic with small sample sizes.

Simulation: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion

| Simulation Parameters | |
|---------------------------------------|--------------------|
| Simulation ID | Sim4 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 119 |
| Prop. Response under Null (π_0) | 0.15 |
| Variance | Empirical Estimate |
| Avg. Power at Termination | 0.03 |
| Response Generation Parameters | |
| Proportion Response (π) | 0.15 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 40 |
| 2 | 79 |
| 3 | 119 |
| Average | 118.528 |

Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | | Total Simulations | |
|--------|-----------------|------------|--------------|-------|-------------------|--|
| | | Efficacy | Efficacy | Count | % | |
| 1 | 40 | 3.185 | 1 | 1 | 0.010% | |
| 2 | 79 | 2.147 | 116 | 116 | 1.160% | |
| 3 | 119 | 1.694 | 184 | 9883 | 98.830% | |
| Total | | | 301 | 10000 | | |
| % | | | 3.010% | | | |

Simulation Seed and Elapsed Time

Starting Seed: 1362061
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:02

Let us now investigate if the problem disappears with larger studies. Select Des3 in the **Library** and click  on the **Library** toolbar. Change the value of **Prop. Response under Alt (π_1)** from 0.25 to 0.18.

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary

Test Type: 1-Sided

Type I Error (α): 0.05

Power: 0.8

Sample Size (n): Computed

Specify Proportion Response

Prop. Response under Null (π_0): 0.15

Prop. Response under Alt (π_1): 0.18

Perform Exact Computations

Variance of Standardized Test Statistic

Under Null Hypothesis

Empirical Estimate

Assurance (Probability of Success):

Click **Compute** to generate the output for Des5. In the **Output Preview**, we see that Des5 requires a sample size of 1035 subjects. To verify whether the use of the empirical variance will indeed produce the correct type-1 error for this large trial, select Des5 in the **Output Preview** and click  icon. In the **Library**, select Des5 and click  icon from **Library** toolbar . First, run 10,000 trials with $\pi = 0.15$. On the **Response Generation** tab, change **Proportion Response** from 0.18 to 0.15. Next click **Simulate**. Observe that the type-1 error obtained by simulating Des5 is about 4.4%, an improvement over the corresponding type 1 error obtained by simulating Des3.

Simulation: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion

| Test Parameters | |
|---------------------------------------|--------------------|
| Simulation ID | Sim5 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 1035 |
| Prop. Response under Null (π_0) | 0.15 |
| Variance | Empirical Estimate |
| Avg. Power at Termination | 0.044 |
| Response Generation Parameters | |
| Proportion Response (π) | 0.15 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 345 |
| 2 | 690 |
| 3 | 1035 |
| Average | 1030.067 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 345 | 3.2 | 3 | 3 | 0.030% |
| 2 | 690 | 2.141 | 137 | 137 | 1.370% |
| 3 | 1035 | 1.695 | 304 | 9860 | 98.600% |
| Total | | | 444 | | |
| % | | | 4.440% | | |

Simulation Seed and Elapsed Time

Starting Seed: 1550867
Total Number of Simulations: 10000
Elapsed Time: 00:00:03

22 Binomial Superiority One-Sample

Next, verify that a sample size of 1035 suffices for producing 80% power by running 10,000 simulations with $\pi = 0.18$.

Simulation: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion

| Test Parameters | |
|---------------------------------------|--------------------|
| Simulation ID | Sim6 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 1035 |
| Prop. Response under Null (π_0) | 0.15 |
| Variance | Empirical Estimate |
| Avg. Power at Termination | 0.82 |
| Response Generation Parameters | |
| Proportion Response (π) | 0.18 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 345 |
| 2 | 690 |
| 3 | 1035 |
| Average | 857.567 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | | Total Simulations | |
|--------|-----------------|------------|--------------|-------|-------------------|--|
| | | Efficacy | Efficacy | Count | % | |
| 1 | 345 | 3.2 | 250 | 250 | 2.500% | |
| 2 | 690 | 2.141 | 4643 | 4643 | 46.430% | |
| 3 | 1035 | 1.695 | 3302 | 5107 | 51.070% | |
| Total | | | 8195 | 10000 | | |
| % | | | 81.950% | | | |

Simulation Seed and Elapsed Time

Starting Seed: 1852184
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:03

This example has demonstrated the importance of simulating a design to verify that it does indeed possess the operating characteristics that are claimed for it. Since these operating characteristics were derived by large-sample theory, they might not hold for small sample sizes, in which case, the sample size or type-1 error might have to be adjusted appropriately.

22.1.3 Interim Monitoring

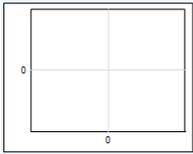
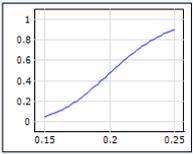
Consider interim monitoring of Des3, the design that has 80% power when the empirical estimate of variance is used to standardize the test statistic. Select Des3 in the **Library**, and click **IM** icon from the Library toolbar. Alternatively, right-click on Des3 and select **Interim Monitoring**. The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs. The bottom section contains four charts, each with a corresponding table to its right. These charts provide graphical and numerical descriptions of the progress of the

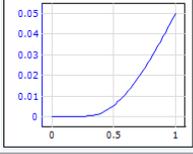
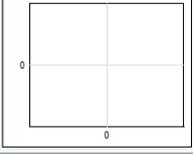
clinical trial and are useful tools for decision making by a data monitoring committee.

Enter Interim Data X [?] [?] [?] [?] - Binonesample.Des3:Interim Monitoring

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | π | Standard Error | Efficacy | 95% RCI for π | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|-------|----------------|----------|-------------------|-------|------------------|----|------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | | | | | | | | | | | |
| 2 | | | | | | | | | | | |
| 3 | | | | | | | | | | | |

Click the "Enter Interim Data" button to enter the Look # 1 data.

| Stopping Boundaries | | Sample Size | Efficacy | Conditional Power | | Prop. | CP |
|---|--|-------------|----------|---|--|-------|-------|
|  | | | |  | | 0.15 | 0.05 |
| | | | | | | 0.16 | 0.098 |
| | | | | | | 0.172 | 0.185 |
| | | | | | | 0.185 | 0.303 |
| | | | | | | 0.197 | 0.439 |
| | | | | | | 0.209 | 0.577 |
| | | | | | | 0.221 | 0.703 |
| | | | | | | 0.234 | 0.805 |
| | | | | | | 0.246 | 0.88 |

| Error Spending Function | | Info. Fraction | α | Confidence Intervals | | Info. Fraction | RCI Upper |
|---|--|----------------|----------|---|--|----------------|-----------|
|  | | | |  | | | |
| | | | | | | | |

Simulation Details | IM Interim Monitoring

At the first interim look, when 40 subjects have enrolled, suppose that the observed response rate is 0.35. Click **Enter Interim Data** icon to invoke the **Test Statistic Calculator**. In the box next to **Cumulative Sample Size** enter 40. Enter 0.35 in the box next to **Estimate of π** . In the box next to **Standard Error of Estimate of π** enter

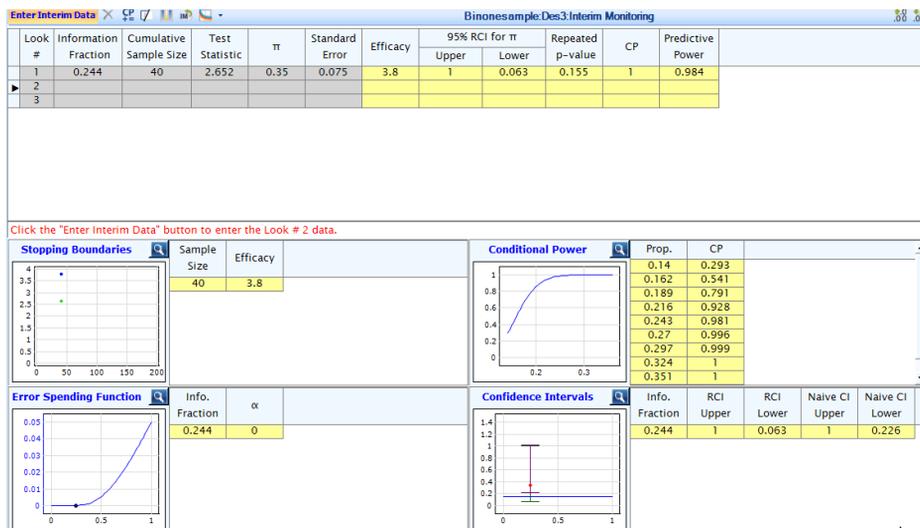
22 Binomial Superiority One-Sample

0.07542. Next click **Recalc**.

The screenshot shows a dialog box titled "Test Statistic Calculator". At the top right, there are two small icons with ".00" below them. The main area is titled "Editing Look #1" and contains a checkbox "Set Current Look as Last" which is unchecked. Below this are three input fields: "Cumulative Sample Size" with the value 40, "Estimate of π " with the value 0.35, and "Standard Error of Estimate of π " with the value 0.07542. These three fields are grouped under the heading "Input for Binomial end point". Below this is an "Output" section with two fields: "Estimate of $\pi - \pi_0$ " with the value 0.2, and "Test Statistic" with the value 2.652. At the bottom of the dialog are three buttons: "Recalc", "OK", and "Cancel".

Observe that upon pressing the **Recalc** button, the test statistic calculator automatically computes the value of the test statistic as 2.652.

Clicking **OK** results in the following output.

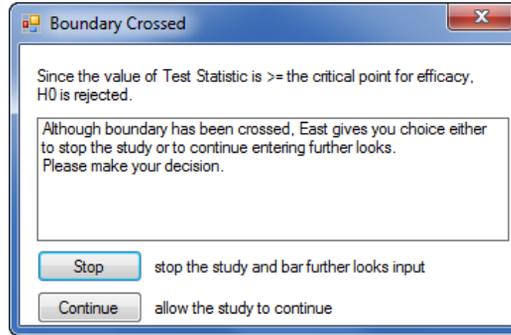


Since our test statistic, 2.652, is smaller than the stopping boundary, 3.185, the trial continues.

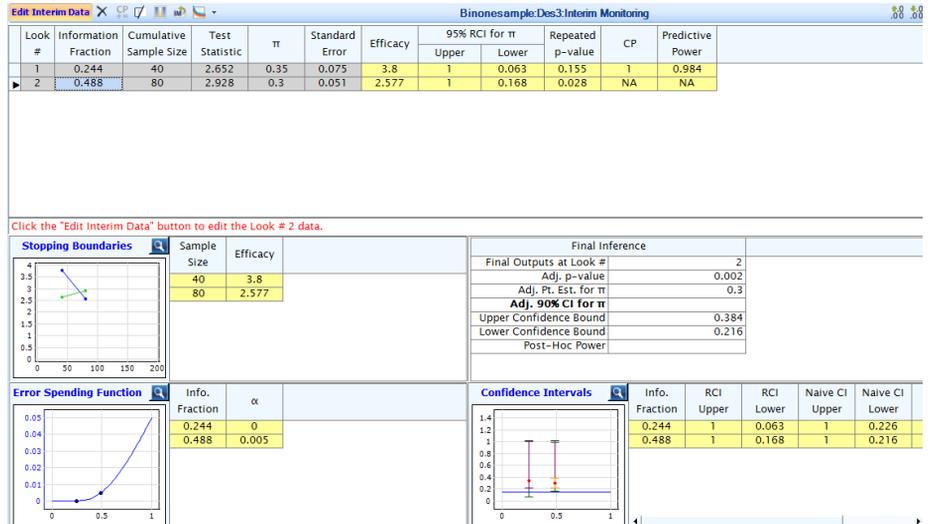
At the second interim monitoring time point, after 80 subjects have enrolled, suppose that the estimate of $\hat{\pi}$ based on all data up to that point is 0.30. Click on the second row in the table in the upper section. Then click **Enter Interim Data** icon. In the box next to **Cumulative Sample Size** enter 80. Enter 0.30 in the box next to **Estimate of π** . In the box next to **Standard Error of Estimate of π** enter 0.05123. Next click **Recalc**. Upon clicking **OK** we observe that the stopping boundary is crossed and the following

22 Binomial Superiority One-Sample

message is displayed.



We can conclude that $\pi > 0.15$ and terminate the trial. Clicking **Stop** yields the following output.



22.2 McNemar's Test

McNemar's Test is used in experimental situations where paired comparisons are observed. In a typical application, two binary response measurements are made on each subject – perhaps from two different treatments, or from two different time points. For example, in a comparative clinical trial, subjects are matched on baseline demographics and disease characteristics and then randomized with one subject in the

pair receiving the experimental treatment and the other subject receiving the control. Another example is the cross over clinical trial in which each subject receives both treatments. By random assignment, some subjects receive the experimental treatment followed by the control while others receive the control followed by the experimental treatment. Let π_c and π_t denote the response probabilities for the control and experimental treatments, respectively. The probability parameters for McNemar's test are displayed in Table 22.1.

Table 22.1: A 2 x 2 Table of Probabilities for McNemar's Test

| Control | Experimental | | Total Probability |
|-------------------|--------------|------------|-------------------|
| | No Response | Response | |
| No Response | π_{00} | π_{01} | $1 - \pi_c$ |
| Response | π_{10} | π_{11} | π_c |
| Total Probability | $1 - \pi_t$ | π_t | 1 |

The null hypothesis

$$H_0: \pi_c = \pi_t$$

is tested against the alternative hypothesis

$$H_1: \pi_c \neq \pi_t$$

for the two sided testing problem or the alternative hypothesis

$$H_1: \pi_c > \pi_t$$

(or $H_1: \pi_c < \pi_t$) for the one-sided testing problem. Since $\pi_t = \pi_c$ if and only if $\pi_{01} = \pi_{10}$, the null hypothesis is also expressed as

$$H_0: \pi_{01} = \pi_{10},$$

and is tested against corresponding one and two sided alternatives. The power of this test depends on two quantities:

1. The difference between the two discordant probabilities (which is also the difference between the response rates of the two treatments)

$$\delta = \pi_{01} - \pi_{10} = \pi_t - \pi_c;$$

22 Binomial Superiority One-Sample

2. The sum of the two discordant probabilities

$$\xi = \pi_{10} + \pi_{01} .$$

East accepts these two parameters as inputs at the design stage.

We next specify the test statistic to be used during the interim monitoring stage. Suppose we intend to execute McNemar's test a maximum of K times in a group sequential setting. Let the cumulative data up to and including the j th interim look consist of $N(j)$ matched pairs arranged in the form of the following 2×2 contingency table of counts:

Table 22.2: 2×2 Contingency Table of Counts of Matched Pairs at Look j

| Control | Experimental | | Total Probability |
|-------------------|--------------|-------------|-------------------|
| | No Response | Response | |
| No Response | $n_{00}(j)$ | $n_{01}(j)$ | $r_0(j)$ |
| Response | $n_{10}(j)$ | $n_{11}(j)$ | $r_1(j)$ |
| Total Probability | $c_0(j)$ | $c_1(j)$ | $N(j)$ |

For $a = 0, 1$ and $b = 0, 1$ define

$$\hat{\pi}_{ab}(j) = \frac{n_{ab}(j)}{N(j)} \tag{22.3}$$

Then the sequentially computed McNemar test statistic at look j is

$$Z_j = \frac{\hat{\delta}_j}{\text{se}(\hat{\delta}_j)} \tag{22.4}$$

where

$$\hat{\delta}_j = \hat{\pi}_{01}(j) - \hat{\pi}_{10}(j) \tag{22.5}$$

and

$$\text{se}(\hat{\delta}_j) = \frac{\sqrt{[n_{01}(j) + n_{10}(j)]}}{N(j)} \tag{22.6}$$

We now show how to use East to design and monitor a clinical trial based on McNemar's test.

22.2.1 Trial Design

Consider a trial in which we wish to determine whether a transdermal delivery system (TDS) can be improved with a new adhesive. Subjects are to wear the old TDS (control) and new TDS (experimental) in the same area of the body for one week each. A response is said to occur if the TDS remains on for the entire one week observation period. From historical data, it is known that control has a response rate of 85% ($\pi_c = 0.85$). It is hoped that the new adhesive will increase this to 95% ($\pi_t = 0.95$). Furthermore, of the 15% of the subjects who did not respond on the control, it is hoped that 87% will respond on the experimental system. That is, $\pi_{01} = 0.87 \times 0.15 = 0.13$. Based on these data, we can fill in all the entries of Table 22.1 as displayed in Table 22.2.

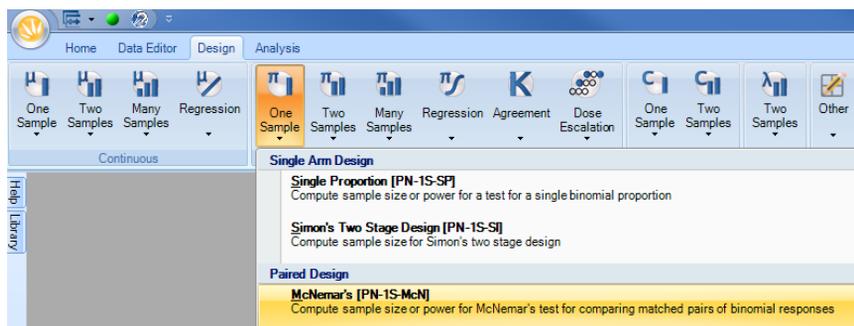
Table 22.3: McNemar Probabilities for the TDS Trial

| Control | Experimental | | Total Probability |
|-------------------|--------------|----------|-------------------|
| | No Response | Response | |
| No Response | 0.02 | 0.13 | 0.15 |
| Response | 0.03 | 0.82 | 0.85 |
| Total Probability | 0.05 | 0.95 | 1 |

Although it is expected that the new adhesive will increase the adherence rate, the comparison is posed as a two-sided testing problem, testing $H_0: \pi_c = \pi_t$ against $H_1: \pi_c \neq \pi_t$ at the 0.05 level. We wish to determine the sample size to have 90% power for the values displayed in Table 22.3. To design this trial, click **Design** tab, then **Single Sample** on the **Discrete** group, and then click **McNemar's Test for**

22 Binomial Superiority One-Sample

Matched Pairs.



Single-Look Design First, consider a study with no interim analyses, and 90% power for two sided test at $\alpha = 0.05$. Choose the design parameters as shown below. We first consider a single-look design, so leave the default value for **Number of Looks** to 1. Enter 0.9 for **Power**. As shown in Table 22.2, we must specify $\delta_1 = \pi_t - \pi_c = 0.1$ and $\xi = \pi_{01} + \pi_{10} = 0.16$.

Design Type: Superiority Number of Looks: 1

Test Parameters

Test Type: 2-Sided
 Type I Error (α): 0.05
 Power: 0.9
 Sample Size (n): Computed

Difference in Probabilities (δ_1): $(\delta_1 = \pi_t - \pi_c)$ 0.1
 Prop. of Discordant Pairs (ξ): $(\xi = \pi_{01} + \pi_{10})$ 0.16

Perform Exact Computations

Probability Allocation: Row = Control, Column = Treatment

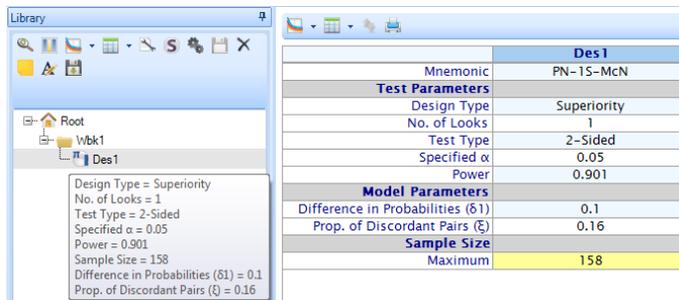
| | | | |
|-------------|-------------|------------|-------------|
| | No Response | Response | Total |
| No Response | π_{00} | π_{01} | $1 - \pi_c$ |
| Response | π_{10} | π_{11} | π_c |
| Total | $1 - \pi_t$ | π_t | 1 |

Click **Compute**. The design Des1 is shown as a row in the Output Preview located in the lower pane of this window. A total of 158 subjects is required to have 90% power.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | Sample Size | δ_1 | ξ |
|------|-------------|--------------|-----------|--------------------|-------|-------------|------------|-------|
| Des1 | Superiority | 1 | 1-Sided | 0.025 | 0.901 | 158 | 0.1 | 0.16 |

You can select this design by clicking anywhere on the row in the **Output Preview**.

Click on  icon to get the output summary displayed in the upper pane. In the **Output Preview** toolbar, click the  icon to save this design Des1 to workbook Wbk1 in the **Library**. If you hover the cursor over Des1 in the Library, a tooltip will appear that summarizes the input parameters of the design.



Five-Look Design Now consider the same design with a maximum of 5 looks, using the default Lan-DeMets (O'Brien-Fleming) spending function. Create a new design by selecting Des1 in the **Library**, and clicking  icon on the **Library** toolbar. Change the **Number of Looks** from 1 to 5, to generate a study with four interim looks and a final analysis. A new tab **Boundary** will appear. Clicking on this tab will reveal the stopping boundary parameters. By default, the **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The left side contains details for the **Efficacy** boundary, and the right side for the **Futility** boundary. By default, there is an efficacy boundary (to reject H0) selected, but no futility boundary (to reject H1). The **Boundary Family** specified is of the Spending Functions type. The default **Spending function** is the Lan-DeMets (Lan & DeMets, 1983), with **Parameter** OF (O'Brien-Fleming), which generates boundaries that are very similar, though not identical, to the classical stopping boundaries of O'Brien and Fleming (1979).

22 Binomial Superiority One-Sample

Technical details of these stopping boundaries are available in Appendix F.

Design Type: Number of Looks:

Test Parameters **Boundary**

Efficacy
 Boundary Family:
 Spending Function:
 Parameter:
 Type I Error (α): 0.05

Futility
 Boundary Family:

Spacing of Looks Equal Unequal Efficacy Boundary:

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.200 | 0.000 | 4.877 | -4.877 |
| 2 | 0.400 | 0.001 | 3.357 | -3.357 |
| 3 | 0.600 | 0.008 | 2.680 | -2.680 |
| 4 | 0.800 | 0.024 | 2.290 | -2.290 |
| 5 | 1.000 | 0.050 | 2.031 | -2.031 |

Click **Compute** to generate output for Des2. With Des2 selected in the **Output Preview**, click the icon to save Des2 to the **Library**. In the **Library**, select the nodes for both Des1 and Des2, by holding the Ctrl key, and then click the icon. The upper pane will display the output summary of the two designs side-by-side:

| | Des1 | Des2 |
|--|-------------|-------------|
| Mnemonic | PN-1S-McN | PN-1S-McN |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 5 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.901 | 0.901 |
| Model Parameters | | |
| Difference in Probabilities (δ_1) | 0.1 | 0.1 |
| Prop. of Discordant Pairs (ξ) | 0.16 | 0.16 |
| Boundary Parameters | | |
| Spacing of Looks | | Equal |
| Efficacy Boundary | | LD (OF) |
| Sample Size | | |
| Maximum | 158 | 162 |
| Expected Under H0 | | 160.935 |
| Expected Under H1 | | 119.965 |

There has been a slight inflation in the maximum sample size, from 158 to 162. However, the expected sample size is 120 subjects if the alternative hypothesis of $\delta_1 = 0.10$ and $\xi = 0.16$ holds. The stopping boundary, spending function, and Power

vs. Sample Size charts can all be displayed by clicking on the appropriate icons from the **Library** toolbar.

22.2.2 Interim Monitoring

Consider interim monitoring of Des2. Select Des2 in the **Library**, and click **IM** icon from the Library toolbar. Alternatively, right-click on Des2 and select **Interim Monitoring**. A new IM worksheet will appear.

Suppose, that the results are to be analyzed after results are available for every 32 subjects. After the first 32 subjects were enrolled, one subject responded on the control arm and did not respond on the treatment arm; four subjects responded on the treatment arm but did not respond on the control arm; 10 subjects did not respond on either treatment; 17 subjects responded on both the arms. This information is sufficient to complete all the entries in Table 22.3 and hence to evaluate the test statistic value.

Click **Enter Interim Data** icon to invoke the **Test Statistic Calculator**. In the box next to **Cumulative Sample Size** enter 32. Enter the values in the table as shown below and click **Recalc**.

Test Statistic Calculator

Editing Look #1
 Set Current Look as Last

Input for Binomial end point

Row = Control, Column = Treatment

| | No Response | Response | Total |
|-------------|-------------|----------|-------|
| No Response | 10 | 4 | 14 |
| Response | 1 | 17 | 18 |
| Total | 11 | 21 | 32 |

Output
 Test Statistic: 1.342

Recalc OK Cancel

22 Binomial Superiority One-Sample

Clicking **OK** results in the following entry in the first look row.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | | 95% RCI for δ | | Repeat... p-value | CP | Predict... |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|--------|----------------------|--------|-------------------|-------|------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.198 | 32 | 1.342 | 0.094 | 0.07 | 4.909 | -4.909 | 0.437 | -0.249 | 0.902 | 0.874 | 0.703 |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | |

As you can see the value of the test statistic, 1.342, is within the stopping boundaries, (4.909,-4.909). Thus, the trial continues.

The second interim analysis was performed after data were available for 64 subjects. A total of two subjects responded on the control arm and failed to respond on the treatment arm; seven subjects responded on the treatment arm and failed to respond on the control arm; 20 subjects responded on neither arm; 35 subjects responded on both the arms.

Click on the second row in the table in the upper section. Then click **Enter Interim Data** icon. Enter the appropriate values in the table as shown below and click **Recalc**.

Test Statistic Calculator
✕

↑.0 ↓.0
.00 .00

Editing Look #2

Set Current Look as Last

Input for Binomial end point

Row = Control, Column = Treatment

| | | | |
|-------------|-------------|----------|-------|
| | No Response | Response | Total |
| No Response | 20 | 7 | 27 |
| Response | 2 | 35 | 37 |
| Total | 22 | 42 | 64 |

Output

Test Statistic: 1.667

Recalc OK Cancel

Then click **OK**. This results in the following screen.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | | 95% RCI for δ | | Repeat... p-value | CP | Predict... Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|--------|----------------------|--------|-------------------|-------|------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.198 | 32 | 1.342 | 0.094 | 0.07 | 4.909 | -4.909 | 0.437 | -0.249 | 0.902 | 0.874 | 0.703 |
| 2 | 0.395 | 64 | 1.667 | 0.078 | 0.047 | 3.38 | -3.38 | 0.237 | -0.08 | 0.434 | 0.803 | 0.707 |
| 3 | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | |

At the third interim analysis, after 96 subjects were enrolled, a total of two subjects responded on the control arm and failed to respond on the treatment arm; 13 subjects responded on the treatment arm and failed to respond on the control arm; 32 subjects did not respond on either arm; 49 subjects responded on both the arms.

Click on the third row in the table in the upper section. Then click **Enter Interim Data** icon. Enter the appropriate values in the table as shown below and click **Recalc**.

Test Statistic Calculator

Editing Look #3

Set Current Look as Last

Input for Binomial end point

Row = Control, Column = Treatment

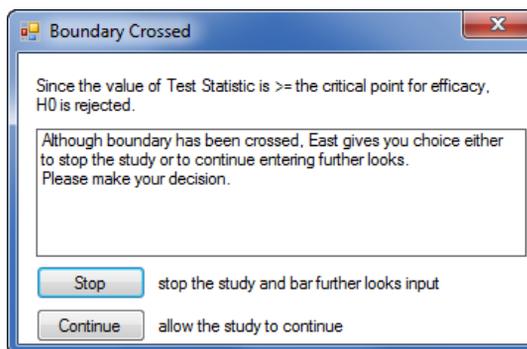
| | No Response | Response | Total |
|-------------|---------------------------------|---------------------------------|---------------------------------|
| No Response | <input type="text" value="32"/> | <input type="text" value="13"/> | <input type="text" value="45"/> |
| Response | <input type="text" value="2"/> | <input type="text" value="49"/> | <input type="text" value="51"/> |
| Total | <input type="text" value="34"/> | <input type="text" value="62"/> | <input type="text" value="96"/> |

Output

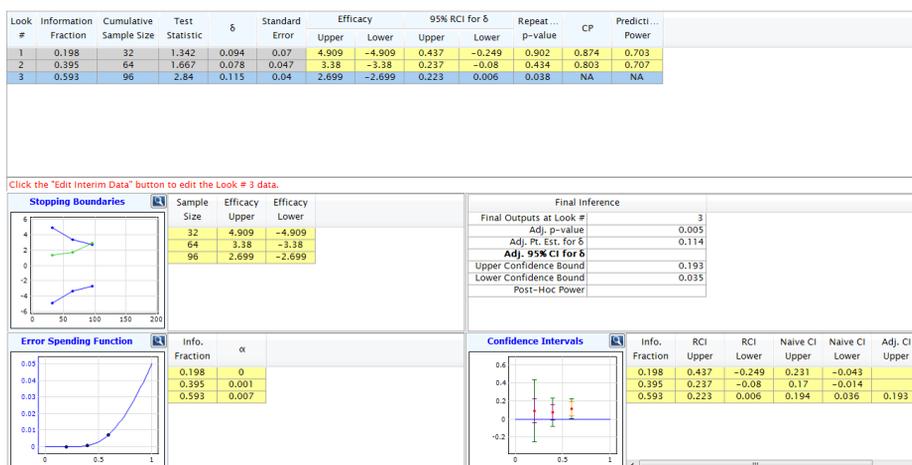
Test Statistic:

22 Binomial Superiority One-Sample

Then click **OK**. This results in the following message box.



Clicking on **Stop** yields the following Interim Monitoring output.



We reject the null hypothesis that $\delta = 0$, based on these data.

22.2.3 Simulation

Des2 can be simulated to examine the properties for different values of the parameters. First, we verify the results under the alternative hypothesis at which the power is to be controlled, namely $\delta_1=0.10$ and $\xi=0.16$.

Select Des2 in the **Library**, and click  icon from **Library** toolbar. Alternatively,

right-click on Des2 and select **Simulate**. A new Simulation worksheet will appear.

Number of Looks: 5

Test Parameters Response Generation Simulation Controls

Trial Type: Superiority

Test Type: 2-Sided

Sample Size (n): 162

| Look # | Info. Fraction | Cum. α Spent | | Efficacy Z | |
|--------|----------------|---------------------|-------|------------|--------|
| | | Upper | Lower | Upper | Lower |
| 1 | 0.198 | 0.000 | 0.000 | 4.909 | -4.909 |
| 2 | 0.401 | 0.000 | 0.000 | 3.351 | -3.351 |
| 3 | 0.599 | 0.004 | 0.004 | 2.684 | -2.684 |
| 4 | 0.802 | 0.012 | 0.012 | 2.285 | -2.285 |
| 5 | 1.000 | 0.025 | 0.025 | 2.032 | -2.032 |

Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled Sim1. Select Sim1 in the **Output Preview**. If you click  icon, you will see some of the simulation output details displayed in the upper pane. Click  icon to save it to the **Library**. Double-click on Sim1 in the **Library**. The simulation output details will be displayed

22 Binomial Superiority One-Sample

as shown below. The results confirm that the power is at about 90%.

Simulation: Discrete Endpoint: One-Sample Test - Paired Design - McNemar's

| Test Parameters | |
|--|-------------|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 5 |
| Test Type | 2-Sided |
| Sample Size (n) | 162 |
| Avg. Power at Termination | 0.905 |
| Response Generation Parameters | |
| Difference in Probabilities (δ_1) | 0.1 |
| Prop. of Discordant Pairs (ξ) | 0.16 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 32 |
| 2 | 65 |
| 3 | 97 |
| 4 | 130 |
| 5 | 162 |
| Average | 125.485 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries Efficacy | | Stopping For | | Total Simulations | |
|--------|-----------------|---------------------|--------|----------------|----------------|-------------------|---------|
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 32 | 4.909 | -4.909 | 0 | 0 | 0 | 0.000% |
| 2 | 65 | 3.351 | -3.351 | 184 | 0 | 184 | 1.840% |
| 3 | 97 | 2.684 | -2.684 | 3528 | 0 | 3528 | 35.280% |
| 4 | 130 | 2.285 | -2.285 | 3687 | 0 | 3687 | 36.870% |
| 5 | 162 | 2.032 | -2.032 | 1653 | 0 | 2601 | 26.010% |
| Total | | | | 9052 | 0 | 10000 | |
| % | | | | 90.520% | 0.000% | | |

To confirm the results under the null hypothesis, set $\delta_1 = 0$ in the **Response Generation** tab in the simulation worksheet and then click **Simulate**. The results, which confirm that the type-1 error rate is approximately 5%, are given below.

| Test Parameters | |
|--|-------------|
| Simulation ID | Sim2 |
| Design Type | Superiority |
| Number of Looks | 5 |
| Test Type | 2-Sided |
| Sample Size (n) | 162 |
| Avg. Power at Termination | 0.043 |
| Response Generation Parameters | |
| Difference in Probabilities (δ_1) | 0 |
| Prop. of Discordant Pairs (ξ) | 0.16 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 32 |
| 2 | 65 |
| 3 | 97 |
| 4 | 130 |
| 5 | 162 |
| Average | 161.176 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries Efficacy | | Stopping For | | Total Simulations | |
|--------|-----------------|---------------------|--------|----------------|----------------|-------------------|---------|
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 32 | 4.909 | -4.909 | 0 | 0 | 0 | 0.000% |
| 2 | 65 | 3.351 | -3.351 | 0 | 0 | 0 | 0.000% |
| 3 | 97 | 2.684 | -2.684 | 19 | 26 | 45 | 0.450% |
| 4 | 130 | 2.285 | -2.285 | 79 | 87 | 166 | 1.660% |
| 5 | 162 | 2.032 | -2.032 | 114 | 105 | 9789 | 97.890% |
| Total | | | | 212 | 218 | 10000 | |
| % | | | | 2.120% | 2.180% | | |

Simulation Seed and Elapsed Time

Starting Seed: 3285368
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:04

While it is often difficult to specify the absolute difference of the discordant probabilities, δ_1 , it is even more difficult to specify the sum of the discordant probabilities, ξ . Simulation can be used to examine the effects of misspecification of ξ . Run the simulations again, now with $\delta_1=0.10$ and $\xi=0.2$. The results are given below.

| Test Parameters | |
|--|-------------|
| Simulation ID | Sim3 |
| Design Type | Superiority |
| Number of Looks | 5 |
| Test Type | 2-Sided |
| Sample Size (n) | 162 |
| Avg. Power at Termination | 0.819 |
| Response Generation Parameters | |
| Difference in Probabilities (δ_1) | 0.1 |
| Prop. of Discordant Pairs (ξ) | 0.2 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

☺ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 32 |
| 2 | 65 |
| 3 | 97 |
| 4 | 130 |
| 5 | 162 |
| Average | 132.415 |

☺ Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|--------|----------------|----------------|-------------------|---------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 32 | 4.909 | -4.909 | 0 | 0 | 0 | 0.000% |
| 2 | 65 | 3.351 | -3.351 | 219 | 0 | 219 | 2.190% |
| 3 | 97 | 2.684 | -2.684 | 2543 | 0 | 2543 | 25.430% |
| 4 | 130 | 2.285 | -2.285 | 3416 | 0 | 3416 | 34.160% |
| 5 | 162 | 2.032 | -2.032 | 2013 | 0 | 3822 | 38.220% |
| Total | | | | 8191 | 0 | 10000 | |
| % | | | | 81.910% | 0.000% | | |

Simulation Seed and Elapsed Time

Starting Seed: 3686640
Total Number of Simulations: 10000
Elapsed Time: 00:00:03

Notice that this provides a power of approximately 81%. Larger values of ξ would further decrease the power. However, values of $\xi > 0.2$ with $\delta_1 = 0.1$ would be inconsistent with the initial assumption of $\pi_c = 0.85$ and $\pi_t = 0.95$. Additional simulations for various values of δ and ξ can provide information regarding the consequences of misspecification of the input parameters.

23 Binomial Superiority Two-Sample

In experiments based on binomial data, the aim is to compare independent samples from two populations in terms of the proportion of sampling units presenting a given trait. In medical research, outcomes such as the proportion of patients responding to a therapy, developing a certain side effect, or requiring specialized care, would satisfy this definition. East supports the design, simulation, and interim monitoring of clinical trials in which this comparison is based on the difference of proportions, the ratio of proportions, or the odds ratio of the two populations. The three cases are discussed in the following sections.

23.1 Difference of Two Binomial Proportions

23.1.1 Trial Design

23.1.2 Interim Monitoring

23.1.3 Pooled versus Unpooled Designs

Let π_c and π_t denote the binomial probabilities for the control and treatment arms, respectively, and let $\delta = \pi_t - \pi_c$. Interest lies in testing the null hypothesis that $\delta = 0$ against one and two-sided alternatives. A special characteristic of binomial designs is the dependence of the variance of a binomial random variable on its mean. Because of this dependence, even if we keep all other test parameters the same, the maximum sample size required to achieve a specified power will be affected by how we intend to standardize the difference of binomial response rates when computing the test statistic at the interim monitoring stage. There are two options for computing the test statistic – use either the **unpooled** or **pooled** estimate of variance for standardizing the observed treatment difference. Suppose, for instance, that at the j th interim look the observed response rate on the treatment arm is $\hat{\pi}_{tj}$, and the observed response rate on the control arm is $\hat{\pi}_{cj}$. Let n_{tj} and n_{cj} be the number of patients on the treatment and control arms, respectively. Then the test statistic based on the unpooled variance is

$$Z_j^{(u)} = \frac{\hat{\pi}_{tj} - \hat{\pi}_{cj}}{\sqrt{\frac{\hat{\pi}_{tj}(1-\hat{\pi}_{tj})}{n_{tj}} + \frac{\hat{\pi}_{cj}(1-\hat{\pi}_{cj})}{n_{cj}}}} \tag{23.1}$$

In contrast, the test statistic based on the pooled variance is

$$Z_j^{(p)} = \frac{\hat{\pi}_{tj} - \hat{\pi}_{cj}}{\sqrt{\hat{\pi}_j(1-\hat{\pi}_j)\left[\frac{1}{n_{tj}} + \frac{1}{n_{cj}}\right]}}, \tag{23.2}$$

where

$$\hat{\pi}_j = \frac{n_{tj}\hat{\pi}_{tj} + n_{cj}\hat{\pi}_{cj}}{n_{tj} + n_{cj}}. \tag{23.3}$$

It can be shown that $[Z_j^{(p)}]^2$ is the familiar Pearson chi-square statistic computed from all the data accumulated by the j th look.

The maximum sample size required to achieve a given power depends on whether, at the interim monitoring stage, we intend to use the unpooled statistic (23.1) or the pooled statistic (23.2) to determine statistical significance. The technical details of the sample size computations for these two options are given in Appendix B, Section B.2.5. The CAPTURE clinical trial is designed in Section 23.1.1 and monitored in Section 23.1.2 under the assumption that the unpooled statistic will be used for interim monitoring. In Section 23.1.3, however, the same trial is re-designed, on the basis of the pooled variance. It is seen that the difference in sample size due to the two design assumptions is almost negligible. This is because the CAPTURE trial utilized balanced randomization. We show further in Section 23.1.3 that if the randomization is unbalanced, the difference in sample size based on the two design assumptions can be substantial.

23.1.1 Trial Design

Design objectives and interim results from CAPTURE, a prospective randomized trial of placebo versus Abciximab for patients with refractory unstable angina were presented at a workshop on clinical trial data monitoring committees (Anderson, 2002). The primary endpoint was reduction in death or MI within 30 days of entering the study. The study was designed for 80% power to detect a reduction in the event rate from 15% on the placebo arm to 10% on the Abciximab arm. A two-sided test with a type-1 error of 5% was used. We will illustrate various design and interim monitoring features of East for studies with binomial endpoints with the help of this example. Thereby this example can serve as a model for designing and monitoring your own binomial studies.

Single Look Design To begin, click **Design** tab, then **Two Samples** on the **Discrete** group, and then click **Difference of Proportions**.

The goal of this study is to test the null hypothesis, H_0 , that the Abciximab and placebo arms both have an event rate of 15%, versus the alternative hypothesis, H_1 , that Abciximab reduces the event rate by 5%, from 15% to 10%. It is desired to have a two sided test with three looks at the data, a type-1 error of $\alpha = 0.05$ and a power of $(1 - \beta) = 0.8$.

Choose the test parameters as shown below. We first consider a single-look design, so leave the default value for **Number of Looks** to 1. Enter 0.8 for the **Power**. To specify the appropriate effect size, enter 0.15 for the **Prop. Under Control** and 0.10 for the **Prop. Under Treatment**. Notice that you have the option to select the manner in which the test statistic will be standardized at the hypothesis testing stage. If you choose **Unpooled Estimate**, the standardization will be according to equation (23.1).

23 Binomial Superiority Two-Sample

If you choose **Pooled Estimate**, the standardization will be according to equation (23.2). For the present, choose the **Unpooled Estimate** option. The other choice in this dialog box is whether or not to use the Casagrande-Pike-Smith (1978) correction for small sample sizes. This is not usually necessary as can be verified by the simulation options in East. The dialog box containing the test parameters will now look as shown below.

Design Type: Superiority Number of Looks: 1

Test Parameters

Test Type: 2-Sided

Type I Error (α): 0.05

Power: 0.8

Sample Size (n): Computed

Allocation Ratio: 1 (n_1/n_2)

Specify Proportion Response
Prop. under Control (π_c): 0.15

Specify Alternative Hypothesis
Prop. under Treatment (π_t): 0.1
Diff. in Prop. ($\delta_1 = \pi_t - \pi_c$): -0.05

Perform Exact Computations
 Specify Variance
 Pooled Estimate
 Unpooled Estimate

Use Casagrande-Pike-Smith Correction (Ignored if alloc. ratio is not 1)

Assurance (Probability of Success):

Next, click **Compute** button. The design is shown as a row in the Output Preview located in the lower pane of this window. The computed sample size (1366 subjects) is highlighted in yellow.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | π_c | Prop. Treatment (Alt.) | δ_1 | Variance |
|------|-------------|--------------|-----------|--------------------|-------|-------|-------------|---------|------------------------|------------|-------------------|
| Des1 | Superiority | 1 | 2-Sided | 0.05 | 0.8 | 1 | 1366 | 0.15 | 0.1 | -0.05 | Unpooled Estimate |

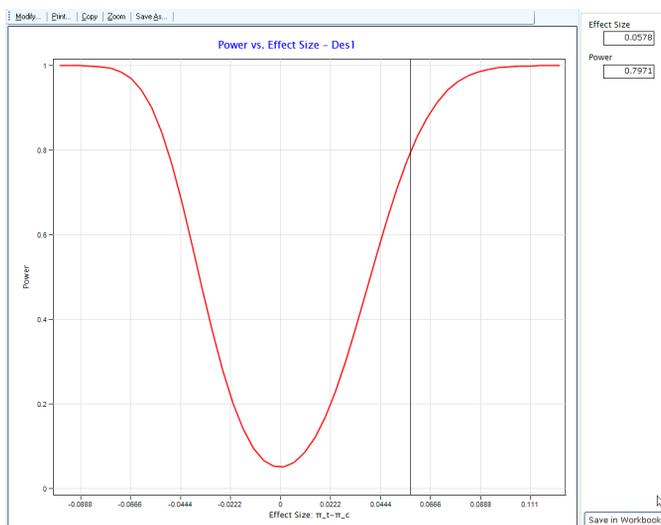
You can select this design Des1 by clicking anywhere on the row in the **Output Preview**. Now you can click  icon to see the output summary displayed in the upper pane. In the **Output Preview** toolbar, click  icon to save this design Des1 to Workbook Wbk1 in the **Library**. If you hover the cursor over Des1 in the Library, a

tooltip will appear that summarizes the input parameters of the design.

The screenshot shows a software window with a 'Library' pane on the left and a parameter table on the right. The 'Library' pane shows a tree structure with 'Root', 'Wbk1', and 'Des1'. A tooltip for 'Des1' lists the following parameters: Design Type = Superiority, No. of Looks = 1, Test Type = 2-Sided, Specified α = 0.05, Power = 0.8, Allocation Ratio (nt/nc) = 1, Sample Size = 1366, Proportion under Control (π_c) = 0.15, Proportion under Treatment (π_t) = 0.1, Diff. in Prop. ($\pi_t - \pi_c$) = -0.05, and Variance = Unpooled Estimate.

| Des 1 | |
|--|-------------------|
| Mnemonic | PN-25-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.8 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.15 |
| Proportion under Treatment (π_t) | 0.1 |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.05 |
| Variance | Unpooled Estimate |
| Sample Size | |
| Maximum | 1366 |

With Des1 selected in the **Library**, click the icon on the **Library** toolbar, and then click **Power vs Treatment Effect** (δ). The resulting power curve for this design is shown. You can save this chart to the **Library** by clicking **Keep**. Alternatively, you can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...** For now, you may close the chart before continuing.



Group Sequential Design Create a new design by selecting Des1 in the **Library**,

23 Binomial Superiority Two-Sample

and clicking  icon on the **Library** toolbar. Change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. A new tab **Boundary** will appear. Clicking on this tab will reveal the stopping boundary parameters. By default, the **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The left side contains details for the **Efficacy** boundary, and the right side for the **Futility** boundary. By default, there is an efficacy boundary (to reject H0) selected, but no futility boundary (to reject H1). The **Boundary Family** specified is of the **Spending Functions** type. The default **Spending function** is the Lan-DeMets (Lan & DeMets, 1983), with **Parameter OF** (O'Brien-Fleming), which generates boundaries that are very similar, though not identical, to the classical stopping boundaries of O'Brien and Fleming (1979). Technical details of these stopping boundaries are available in Appendix F.

Click **Boundary** tab to see the details of cumulative alpha spent, and the boundary values, in the **Look Details** table.

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.05

Futility

Boundary Family: None

Spacing of Looks Equal Unequal

Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.333 | 0.000 | 3.710 | -3.710 |
| 2 | 0.667 | 0.012 | 2.511 | -2.511 |
| 3 | 1.000 | 0.050 | 1.993 | -1.993 |

Click **Compute** to generate output for a new design Des2. The 3-look group sequential design displayed in Des2 requires an upfront commitment of up to a maximum of 1384 patients. That is 18 patients more than the fixed sample design displayed in Des1. Notice, however, that under the alternative hypothesis of a 5% drop in the event rate, the expected sample size is only 1183 patients – a saving of 201 patients relative to the fixed sample design. This is because the test statistic could cross a stopping boundary at one of the interim looks.

With Des2 selected in the **Output Preview**, click  icon to save Des2 to the **Library**. In order to see the stopping probabilities, as well as other characteristics,

select Des2 in the **Library**, and click  icon. The cumulative boundary stopping probabilities are shown in the **Stopping Boundaries** table.

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Test Parameters | |
|-----------------------------------|-------------------|
| Design ID | Des2 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.8 |
| Model Parameters | |
| Prop. under Control (π_0) | 0.15 |
| Prop. under Treatment (π_1) | 0.1 |
| $\delta = \pi_1 - \pi_0$ | |
| Under H0 | 0 |
| Under H1 | -0.05 |
| Allocation Ratio (n/n_0) | 1 |
| Variance | Unpooled Estimate |
| Casagrande-Pike-Smith Correction | Not Applied |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |

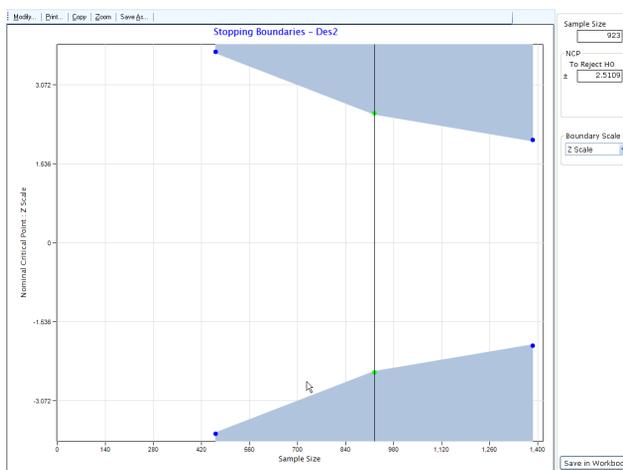
Sample Size Information

| | Control Arm | Treatment Arm | Total |
|-----------------------------------|-------------|---------------|----------|
| Sample Size (n) | | | |
| Maximum | 692 | 692 | 1384 |
| Expected H1 | 591.548 | 591.13 | 1182.678 |
| Expected H0 | 689.166 | 689.154 | 1378.32 |
| Maximum Information (I): 3181.609 | | | |

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | | Boundary Crossing Probability (Incremental) | | | |
|--------|--------------------------|-----------------|---------------------------|------------|--------|---|----------|----------|-------|
| | | | | Efficacy Z | | Under H0 | | Under H1 | |
| | | | | Upper | Lower | Efficacy | | Efficacy | |
| 1 | 0.333 | 461 | 2.058E-4 | 3.712 | -3.712 | 1.029E-4 | 1.029E-4 | 4.661E-8 | 0.019 |
| 2 | 0.667 | 923 | 0.012 | 2.511 | -2.511 | 0.006 | 0.006 | 7.348E-7 | 0.4 |
| 3 | 1 | 1384 | 0.05 | 1.993 | -1.993 | 0.019 | 0.019 | 6.585E-7 | 0.382 |

Close the Output window before continuing. The stopping boundary chart can be brought up by clicking  icon on the **Library** toolbar, and then clicking **Stopping Boundaries**. The following chart will appear.

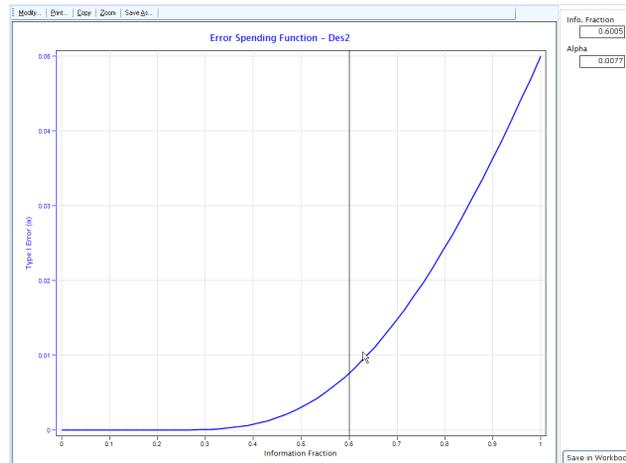


Lan-DeMets Spending Function: O'Brien-Fleming Version

Close this chart, and click  icon in the **Library** toolbar and then **Error**

23 Binomial Superiority Two-Sample

Spending The following chart will appear.



This spending function was proposed by Lan and DeMets (1983), and for two-sided tests has the following functional form :

$$\alpha(t) = 4 - 4\Phi\left(\frac{z_{\alpha/4}}{\sqrt{t}}\right) . \quad (23.4)$$

Notice that hardly any type-1 error is spent in the early stages of the trial but the rate of error spending increases rapidly as the trial progresses. This is reflected in the corresponding stopping boundaries. The upper and lower boundary values are rather wide apart initially (± 3.712 standard deviations) but come closer together with each succeeding interim look until at the last look the standardized test statistic crosses the boundary at ± 1.993 standard deviations. This is not too far off from the corresponding boundary values, ± 1.96 , required to declare statistical significance at the 0.05 level for a fixed sample design. For this reason this spending function is often adopted in preference to other spending functions that spend the type-1 error more aggressively and thereby reduce the expected sample size under H_1 by a greater amount.

Lan-DeMets Spending Function: Pocock Version

A more aggressive spending function, also proposed by Lan and DeMets (1983), is **PK** which refers to Pocock. This spending function captures the spirit of the Pocock (1977) stopping boundary belonging to the Wang and Tsatis (1987) power family, and

has the functional form

$$\alpha(t) = \alpha \log(1 + (e - 1)t) . \tag{23.5}$$

Select Des2 in the **Library**, and click  icon on the **Library** toolbar. On the **Boundary** tab, change the **Parameter** from OF to PK, and click **Compute** to create design Des3. With Des3 selected in the **Output Preview**, click the  icon. In the **Library**, select the nodes for both Des2 and Des3, by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the two designs side-by-side:

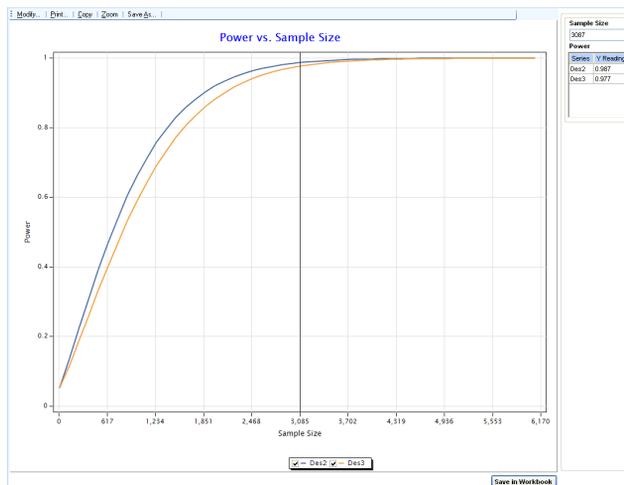
| | Des2 | Des3 |
|--|-------------------|-------------------|
| Mnemonic | PN-2S-DI | PN-2S-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 3 | 3 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.8 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Proportion under Control (π_c) | 0.15 | 0.15 |
| Proportion under Treatment (π_t) | 0.1 | 0.1 |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.05 | -0.05 |
| Variance | Unpooled Estimate | Unpooled Estimate |
| Boundary Parameters | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (PK) |
| Sample Size | | |
| Maximum | 1384 | 1599 |
| Expected Under H0 | 1378.32 | 1566.588 |
| Expected Under H1 | 1182.678 | 1119.294 |

Under Des3, you must make an up-front commitment of up to 1599 patients, considerably more than you would need for a fixed sample design. However, because the type-1 error is spent more aggressively in the early stages, the expected sample size is only 1119 patients.

For now, close this output window, and click  icon on the **Library** toolbar to

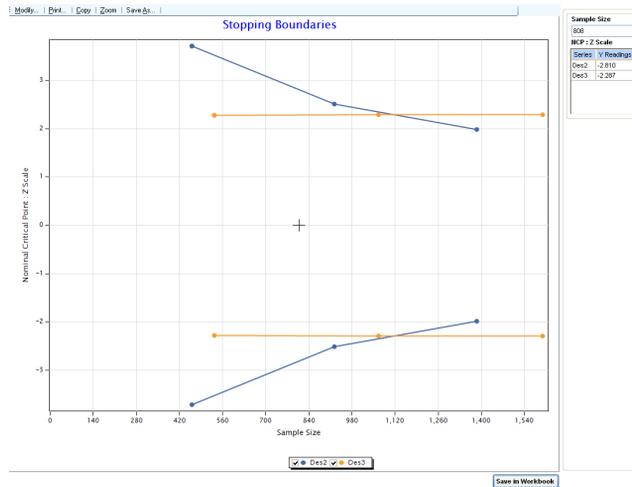
23 Binomial Superiority Two-Sample

compare the two designs according to Power vs. Sample Size.

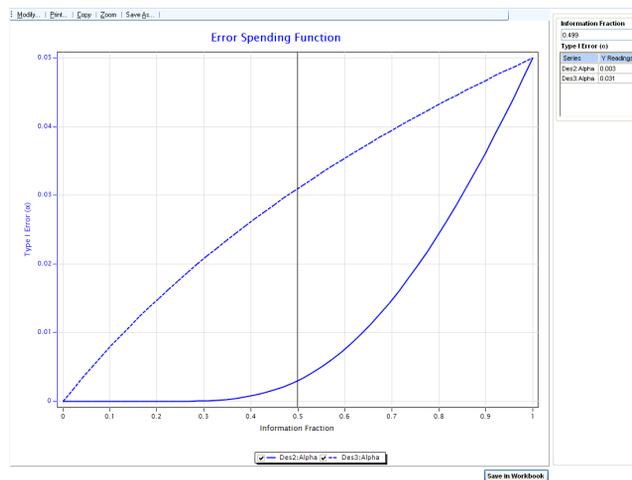


Using the same icon, select Stopping Boundaries. Notice, by moving the cursor from right to left in the stopping boundary charts, that the stopping boundary derived from the PK spending function is approximately flat, requiring ± 2.28 standard deviations at the first look and ± 2.29 standard deviations at the second and ± 2.30 third looks. In contrast, the stopping boundary derived from the OF spending function requires ± 3.71 standard deviations at the first look, ± 2.51 standard deviations at the second look and ± 1.99 standard deviations at the third look. This translates into a smaller expected sample size under H_1 for Des3 than for Des2. This advantage is, however, offset by at least two drawbacks of the stopping boundary derived from the PK spending function; the large up-front commitment of 1599 patients, and the large standardized test statistic of 2.295 (corresponding to a two-sided p value of 0.0217)

required at the last look in order to declare statistical significance.



Using the same icon, select Error Spending to compare the two designs graphically in terms of error spending functions. Des3 (PK) spends the type-1 error probability at a much faster rate than Des2 (OF). Close the chart before continuing.



23 Binomial Superiority Two-Sample

Wang and Tsiatis Power Boundaries

The stopping boundaries generated by the Lan-Demets OF and PK functions closely resemble closely the classical O'Brien-Fleming and Pocock stopping boundaries, respectively. These classical boundaries are a special case of a family of power boundaries proposed by Wang and Tsiatis (1987). For a two-sided α level test, using K equally spaced looks, the power boundaries for the standardized test statistic Z_j at the j -th look are of the form

$$Z_j \geq \frac{C(\Delta, \alpha, K)}{(j/K)^{0.5-\Delta}} \quad (23.6)$$

The normalizing constant $C(\Delta, \alpha, K)$ is evaluated by recursive integration so as to ensure that the K -look group sequential test has type-1 error equal to α .

Select Des3 in the **Library** and click  on the **Library** toolbar. On the **Boundary** tab, change the **Boundary Family** from Spending Functions to Wang-Tsiatis. Leave the default value of Δ as 0 and click **Compute** to create design Des4.

Efficacy

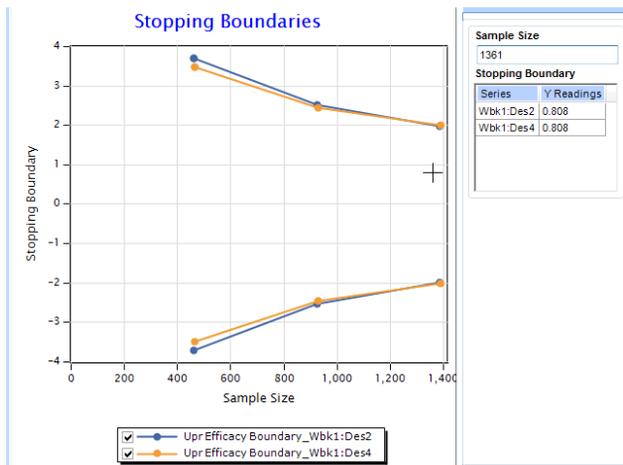
Boundary Family: Wang-Tsiatis

Shape Parameter (Δ):

Type I Error (α): 0.05

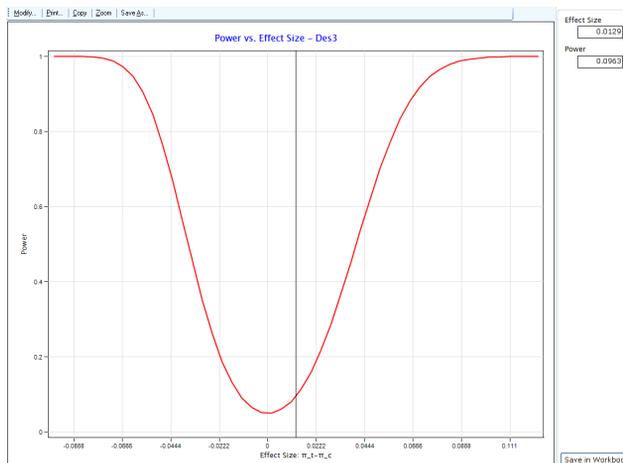
With Des4 selected in the **Output Preview**, click the  icon. In the **Library**, select both Des2 and Des4 by holding the Ctrl key. Click  icon, and under **Select Chart** on the right, select Stopping Boundaries. As expected, the boundary values for Des2 (Lan-Demets, OF) and Des4 (Wang-Tsiatis, $\Delta = 0$) are very similar.

Close the chart before continuing.



The Power Chart and the ASN Chart

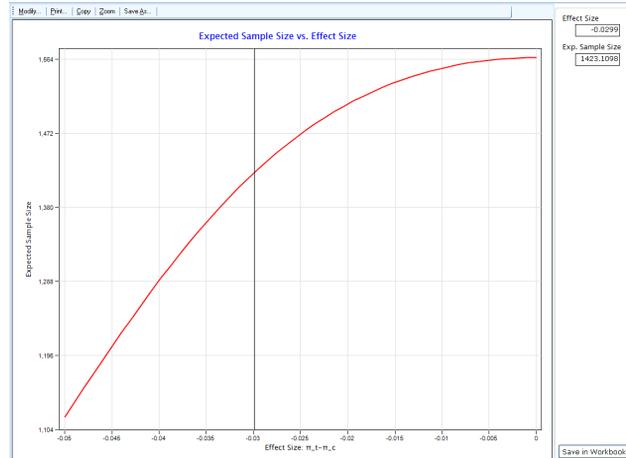
East provides some additional tools for evaluating study designs. Select Des3 in the **Library**, click the  icon, and then click **Power vs. Treatment effect (δ)**. By scrolling from left to right with the vertical line cursor, one can observe the power for various values of the effect size.



Close this chart, and with Des3 selected, click the  icon again. Then click

23 Binomial Superiority Two-Sample

Expected Sample Size. The following chart appears:



This chart displays the Expected Sample Size as a function of the effect size and confirms that for Des3 the average sample size is 1566 under H_0 (effect size, zero) and 1120 under H_1 (effect size, -0.05).

Unequally spaced analysis time points

In the above designs, we have assumed that analyses were equally spaced. This assumption can be relaxed if you know when interim analyses are likely to be performed (e.g. for administrative reasons). In either case, departures from this assumption are allowed during the actual interim monitoring of the study, but sample size requirements will be more accurate if allowance is made for this knowledge.

With Des3 selected in the **Library**, click the  icon. Under **Spacing of Looks** in the **Boundary** tab, click the **Unequal** radio button. The column titled **Info. Fraction** in the **Look Details** table can be edited to modify the relative spacing of the analyses. The information fraction refers to the proportion of the maximum (yet unknown) sample size. By default, this table displays equal spacing, but suppose that the two interim analyses will be performed with 0.25 and 0.5 (instead of 0.333 and 0.667) of the maximum sample size. Enter these new information fraction values and click **Compute** to create design Des5. Select Des5 in the **Output Preview** and click  icon to save it in the **Library** for now.

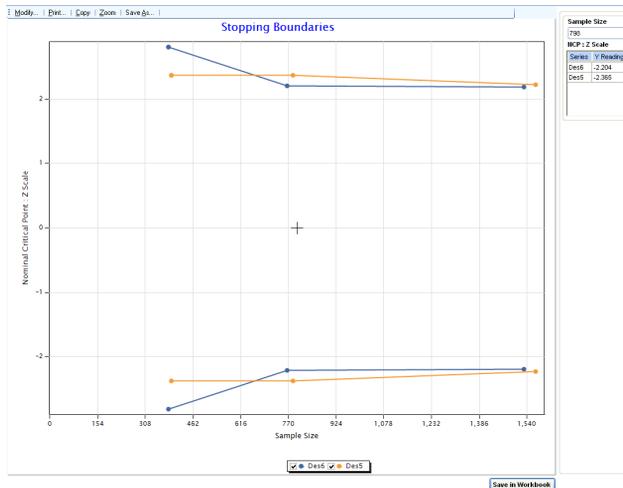
Arbitrary amounts of error probability to be spent at each analysis

Another feature of East is the possibility to specify arbitrary amounts of cumulative error probability to be used at each look. This option can be combined with the option of unequal spacing of the analyses. With Des5 selected in the **Library**, click the  icon on the **Library** toolbar. Under the **Boundary** tab, select **Interpolated** for the **Spending Function**. In the column titled **Cum. α Spent**, enter 0.005 for the first look and 0.03 for the second look, and click **Compute** to create design Des6.

Look Details:

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|-------|
| | | | Upper | Lower |
| 1 | 0.250 | 0.0050 | | |
| 2 | 0.500 | 0.0300 | | |
| 3 | 1.000 | 0.0500 | | |

Select Des6 in the **Output Preview** and click  icon. From the **Library**, select Des5 and Des6 by holding the Ctrl key. Click  icon, and under **Select Chart** on the right, select **Stopping Boundaries**. The following chart will be displayed.



Computing power for a given sample size

When sample size is a given design constraint, East can compute the achieved power, given the other test parameters. Select Des6 in the **Library** and click  icon. On the **Test Parameters** tab, click the radio button for **Power(1 - β)**. You will notice that

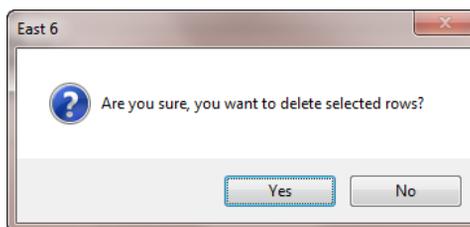
23 Binomial Superiority Two-Sample

the field for power will contain the word **Computed**. You may now enter a value for the sample size: 1250, and click **Compute**.

The following output will appear in **Output Preview** in Des7 row, where, as expected, the achieved power is less than 0.9, namely 0.714.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | tTC | Prop. Treatment (Alt.) | δ_1 | Variance | Spacing of Looks | Efficacy Boundary | Expected SS (H0) | Expected SS (H1) |
|------|-------------|--------------|-----------|--------------------|-------|-------|-------------|------|------------------------|------------|-------------------|------------------|-------------------|------------------|------------------|
| Des1 | Superiority | 1 | 2-Sided | 0.05 | 0.8 | 1 | 1366 | 0.15 | 0.1 | -0.05 | Unpooled Estimate | | | | |
| Des2 | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | 1384 | 0.15 | 0.1 | -0.05 | Unpooled Estimate | Equal | LD (OF) | 1378.32 | 1182.678 |
| Des3 | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | 1599 | 0.15 | 0.1 | -0.05 | Unpooled Estimate | Equal | LD (PK) | 1566.588 | 1119.294 |
| Des4 | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | 1390 | 0.15 | 0.1 | -0.05 | Unpooled Estimate | Equal | WT (0) | 1383.124 | 1169.726 |
| Des5 | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | 1568 | 0.15 | 0.1 | -0.05 | Unpooled Estimate | Unequal | LD (PK) | 1536.687 | 1151.318 |
| Des6 | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | 1529 | 0.15 | 0.1 | -0.05 | Unpooled Estimate | Unequal | Interp. | 1504.165 | 1140.019 |
| Des7 | Superiority | 3 | 2-Sided | 0.05 | 0.714 | 1 | 1250 | 0.15 | 0.1 | -0.05 | Unpooled Estimate | Unequal | Interp. | 1229.69 | 987.482 |

To delete this design, click Des7 in the **Output Preview**, and click icon in the text **Output Preview** toolbar. East will display a warning to make sure that you want to delete the selected row. Click **Yes** to continue.



Stopping Boundaries for Early Rejection of H_0 or H_1 Although both Des2 and Des3 reduce the expected sample size substantially by rejecting H_0 when H_1 is true, they are unable to do so if H_0 is true. It is, however, often desirable to terminate a study early if H_0 is true since that would imply that the new treatment is no different than the standard treatment. East can produce stopping boundaries that result in early termination either under either H_0 or H_1 . Stopping boundaries for early termination if

H_1 is true are known as **efficacy boundaries**. They are obtained by choosing an appropriate α -spending function. These boundaries ensure that the type 1 error does not exceed the pre-specified significance level α . East can also construct stopping boundaries for rejecting H_1 and terminating early if H_0 is true. These stopping boundaries are known as **futility boundaries**. They are obtained by choosing an appropriate β spending function. These boundaries ensure that the type 2 error does not exceed β and thereby ensure that the power of the study is preserved at $1 - \beta$ despite the possibility of early termination for futility. Pampallona and Tsiatis (1994) have extended the error spending function methodology of Lan and DeMets (1983) so as to spend both α , the type-1 error, and β , the type-2 error, and thereby obtain efficacy and futility boundaries simultaneously. East provides you with an entire catalog of published spending functions from which you can take your pick for generating both the H_0 and H_1 boundaries.

For various reasons, investigators usually prefer to be very conservative about early stopping for efficacy but are likely to be more aggressive about cutting their losses and stopping early for futility. Suppose then that you wish to use the conservative Lan-DeMets (OF) spending function for early termination to reject H_0 in favor of H_1 , but use a more aggressive spending function for early termination to reject H_1 in favor of H_0 . Possible choices for spending functions to reject H_1 that are more aggressive than Lan-DeMets(OF) but not as aggressive as Lan-DeMets(PK) are members of the Rho family (Jennison and Turnbull, 2000) and the Gamma family (Hwang, Shih and DeCani, 1990). For illustrative purposes we will use the **Gamam(-1)** spending function from the Gamma family.

Select Des2 in the **Library** and click  icon. For the futility boundary on the **Boundary** tab, select Spending Functions and then select Gamma Family. Set the **Parameter** to -1 . Also, click on the **Binding** option to the right. The screen

23 Binomial Superiority Two-Sample

will look like this:

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.05

Futility

Boundary Family: Spending Functions

Spending Function: Gamma Family

Parameter (γ): -1

Type II Error (β): 0.2

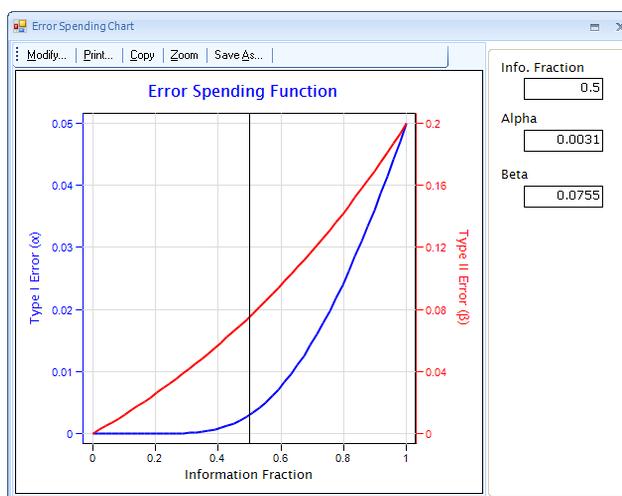
Non-Binding
 Binding

Spacing of Looks: Equal Unequal

Boundary Scale: Z Scale

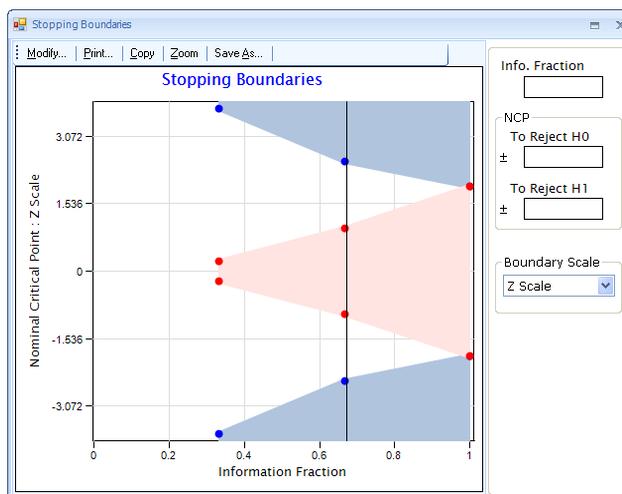
| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | | Cum. β Spent | Futility Boundary | |
|--------|----------------|-------------------------------------|-------------------------------------|---------------------|-------------------|--------|--------------------|-------------------|--------|
| | | | | | Upper | Lower | | Upper | Lower |
| 1 | 0.333 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.000 | 3.710 | -3.710 | 0.046 | 0.232 | -0.232 |
| 2 | 0.667 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.012 | 2.509 | -2.509 | 0.110 | 0.979 | -0.979 |
| 3 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.050 | 1.933 | -1.933 | 0.200 | 1.933 | -1.933 |

On the **Boundary** tab, you may click icon, or icon to view plots of the error spending functions, or stopping boundaries, respectively.



Observe that the β -spending function (upper in red) spends the type-2 error

substantially faster than the α -spending function (lower in blue).



These stopping boundaries are known as **inner-wedge** stopping boundaries. They divide the sample space into three zones corresponding to three possible decisions. If the test statistic enters the lower blue zone, we terminate the trial, reject H_0 , and conclude that the new treatment (Abciximab) is beneficial relative to the placebo. If the test statistic enters the upper blue zone, we terminate the trial, reject H_0 , and conclude that the new treatment is harmful relative to the placebo. If the test statistic enters the center (pink) zone, we terminate the trial, reject H_1 , and conclude that Abciximab offers no benefit relative to placebo. Assuming that the event rate is 0.15 for the placebo arm, this strategy has a 2.5% chance of declaring benefit and a 2.5% chance of declaring harm when the event rate for the Abciximab arm is also 0.15. Furthermore this strategy has a 20% chance of entering the pink zone and declaring no benefit when there actually is a substantial benefit with Abciximab, resulting in a drop in the event rate from 0.15 to 0.1. In other words, Des7 has a two-sided type-1 error of 5% and 80% power.

Click **Compute** and with Des7 selected in the **Output Preview**, click the  icon.

To view the design details, click the  icon. Des7 requires an up-front commitment of 1468 patients, but the expected sample size is 1028 patients under H_0 , and 1164 patients under H_1 . You may wish to save this output (e.g., in HTML format)

by clicking on the  icon, or to print by clicking on the  icon. Close the

23 Binomial Superiority Two-Sample

output window before continuing.

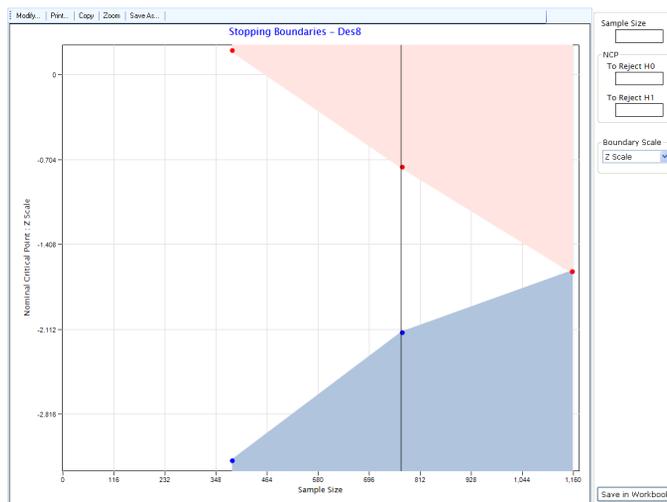
Boundaries with Early Stopping for Benefit or Futility

Next suppose you are interested in designing the clinical trial in such a way that you can reach only two conclusions, not three. You wish to demonstrate either that Abciximab is beneficial relative to placebo or that it offers no benefit relative to placebo, but there is no interest in demonstrating that Abciximab is harmful relative to placebo. To design this two-decision trial select Des7 in the **Library** and click the  icon. Change the entry in the **Test Type** cell from 2-Sided to 1-Sided. Check to ensure other specifications are same as in Des7. Click **Compute** to generate the design.

| | Des 7 | Des 8 |
|--|-------------------|-------------------|
| Mnemonic | PN-2S-DI | PN-2S-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 3 | 3 |
| Test Type | 2-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.8 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Proportion under Control (π_c) | 0.15 | 0.15 |
| Proportion under Treatment (π_t) | 0.1 | 0.1 |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.05 | -0.05 |
| Variance | Unpooled Estimate | Unpooled Estimate |
| Boundary Parameters | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) |
| Futility Boundary | Gm (-1) (B) | Gm (-1) (B) |
| Sample Size | | |
| Maximum | 1468 | 1156 |
| Expected Under H0 | 1028.235 | 680.953 |
| Expected Under H1 | 1163.929 | 891.733 |

The error spending functions are the same but this time the stopping boundaries divide

the sample space into two zones only as shown below.



If the test statistic enters the lower (blue) zone, the null hypothesis is rejected in favor of concluding that Abciximab is beneficial relative to placebo. The probability of this event under H_0 is 0.05. If the test statistic enters the upper (pink) zone the alternative hypothesis is rejected in favor of concluding that Abciximab offers no benefit relative to placebo. The probability of this event under H_1 is 0.2. In other words, Design8 has a one sided type-1 error rate of 5% and 80% power. Since Design8 precludes the possibility of demonstrating that Abciximab is harmful relative to placebo, it requires far fewer patients. It only requires an up-front commitment of 1156 patients and the expected sample size is 681 if H_0 is true and 892 if H_1 is true.

Before continuing to the next section, we will save the current workbook, and open a new workbook. Select the workbook node in the **Library** and Click the  button in the top left hand corner, and click **Save**. Alternatively, select Workbook1 in the **Library** and right-click, then click **Save**. This saves all the work done so far on your directory.

Next, click the  button, click **New**, and then **Workbook**. A new workbook, Wbk2, should appear in the **Library**. Next, close the window to clear all designs from the **Output Preview**.

23 Binomial Superiority Two-Sample

Multiple designs for discrete outcomes East allows the user to easily create multiple designs by specifying a range of values for certain parameters in the design window. In studies with discrete outcomes, East supports the input of multiple key parameters at once to simultaneously create a number of different designs. For example, suppose in a multi-look study the user wants to generate designs for all combinations of the following parameter values in a two sample **Difference of Proportions** test: **Power** = 0.8 and 0.9, and **Alternative Hypothesis - Prop. under Treatment** = 0.4, 0.5 and 0.6. The number of combinations is $2 \times 3 = 6$. East creates all permutations using only a single specification under the **Test Parameters** tab in the design window. As shown below, the values for **Power** are entered as a list of comma separated values, while the **Prop. under Treatment** for the alternative hypothesis are entered as a colon separated range of values, 0.4. to 0.6 in steps of 0.1.

East computes all 6 designs and displays them in the **Output Preview** window:

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | ttc | Prop. Treatment (Alt.) | δ_1 | Variance |
|------|-------------|--------------|-----------|--------------------|-------|-------|------------------|-------------------|-------------|------------------|------------------|-----|------------------------|------------|-------------------|
| Des1 | Superiority | 3 | 2-Sided | 0.05 | 0.805 | 1 | Equal | LD (OF) | 59 | 58.762 | 50.312 | 0.1 | 0.4 | 0.3 | Unpooled Estimate |
| Des2 | Superiority | 3 | 2-Sided | 0.05 | 0.802 | 1 | Equal | LD (OF) | 34 | 33.857 | 29.026 | 0.1 | 0.5 | 0.4 | Unpooled Estimate |
| Des3 | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | Equal | LD (OF) | 21 | 20.914 | 17.945 | 0.1 | 0.6 | 0.5 | Unpooled Estimate |
| Des4 | Superiority | 3 | 2-Sided | 0.05 | 0.9 | 1 | Equal | LD (OF) | 78 | 77.68 | 62.547 | 0.1 | 0.4 | 0.3 | Unpooled Estimate |
| Des5 | Superiority | 3 | 2-Sided | 0.05 | 0.905 | 1 | Equal | LD (OF) | 46 | 45.807 | 36.789 | 0.1 | 0.5 | 0.4 | Unpooled Estimate |
| Des6 | Superiority | 3 | 2-Sided | 0.05 | 0.909 | 1 | Equal | LD (OF) | 29 | 28.885 | 23.016 | 0.1 | 0.6 | 0.5 | Unpooled Estimate |

East provides the capability to analyze multiple designs in ways that make comparisons between the designs visually simple and efficient. To illustrate this, a selection of a few of the above designs can be viewed simultaneously in both the **Output Summary** section as well as in the various tables and plots. The following is a subsection of the designs computed from the above example with differing values for number of looks, power and proportion under treatment. Designs are displayed side by side, allowing details to be easily compared. Save these designs in the newly created workbook.

In addition East allows multiple designs to be viewed simultaneously either graphically or in tabular format: **Stopping Boundaries (table)**

Boundary Scales :

Wbk1:Des1

| Look # | Info. Fraction | Sample Size | Cum. α Spent | Boundaries | | Incremental Boundary Crossing Probabilities | | | |
|--------|----------------|-------------|---------------------|-------------------|--------|---|-------|--------------------------|-------|
| | | | | | | Under H0: $\delta = 0$ | | Under H1: $\delta = 0.3$ | |
| | | | | Efficacy Boundary | | Efficacy | | Efficacy | |
| | | | | Upper | Lower | Upper | Lower | Upper | Lower |
| 1 | 0.339 | 20 | 0 | 3.676 | -3.676 | 0 | 0 | 0.021 | 0 |
| 2 | 0.661 | 39 | 0.012 | 2.524 | -2.524 | 0.006 | 0.006 | 0.393 | 0 |
| 3 | 1 | 59 | 0.05 | 1.992 | -1.992 | 0.019 | 0.019 | 0.391 | 0 |

Wbk1:Des2

| Look # | Info. Fraction | Sample Size | Cum. α Spent | Boundaries | | Incremental Boundary Crossing Probabilities | | | |
|--------|----------------|-------------|---------------------|-------------------|--------|---|-------|--------------------------|-------|
| | | | | | | Under H0: $\delta = 0$ | | Under H1: $\delta = 0.4$ | |
| | | | | Efficacy Boundary | | Efficacy | | Efficacy | |
| | | | | Upper | Lower | Upper | Lower | Upper | Lower |
| 1 | 0.324 | 11 | 0 | 3.771 | -3.771 | 0 | 0 | 0.015 | 0 |
| 2 | 0.676 | 23 | 0.013 | 2.489 | -2.489 | 0.006 | 0.006 | 0.42 | 0 |
| 3 | 1 | 34 | 0.05 | 1.995 | -1.995 | 0.019 | 0.019 | 0.367 | 0 |

23 Binomial Superiority Two-Sample

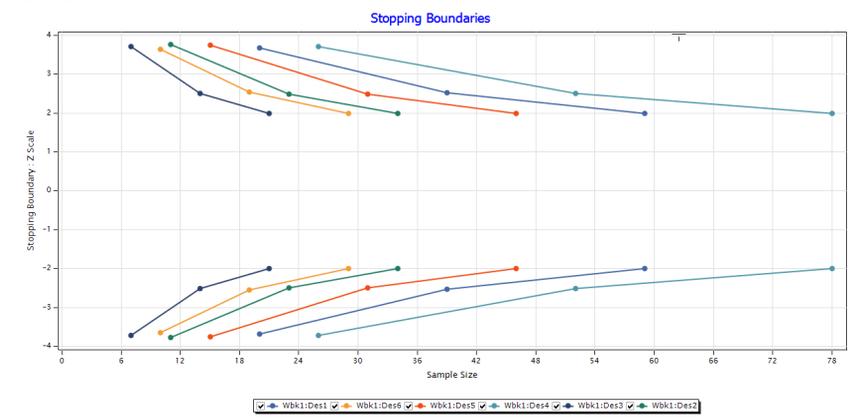
Error Spending (table)

Range for Information Fraction

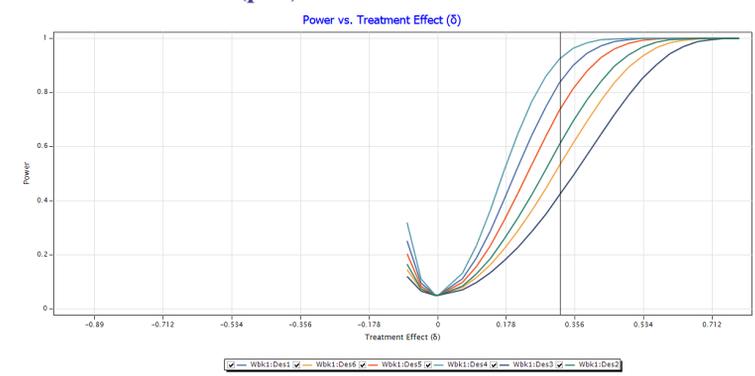
From To Step Size

| Information Fraction | Alpha_Wbk1:Des1 | Alpha_Wbk1:Des6 | Alpha_Wbk1:Des5 | Alpha_Wbk1:Des4 | Alpha_Wbk1:Des3 | Alpha_Wbk1:Des2 |
|----------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 0.5 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 |
| 0.55 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 |
| 0.6 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 |
| 0.65 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 | 0.011 |
| 0.7 | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 | 0.015 |
| 0.75 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 | 0.019 |
| 0.8 | 0.024 | 0.024 | 0.024 | 0.024 | 0.024 | 0.024 |
| 0.85 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| 0.9 | 0.036 | 0.036 | 0.036 | 0.036 | 0.036 | 0.036 |

Stopping Boundaries (plot)



Power vs. Treatment Effect (plot)



This capability allows the user to explore a greater space of possibilities when determining the best choice of study design.

Select individual looks

With Des8 selected in Wbk1, click  icon. In the **Spacing of Looks** table of the **Boundary** tab, notice that there are ticked checkboxes under the columns **Stop for Efficacy** and **Stop for Futility**. East gives you the flexibility to remove one of the stopping boundaries at certain looks, subject to the following constraints: (1) both boundaries must be included at the final two looks, (2) at least one boundary, either efficacy or futility, must be present at each look, (3) once a boundary has been selected all subsequent looks must include this boundary as well and (4) efficacy boundary for the penultimate look cannot be absent.

Untick the checkbox in the first look under the **Stop for Futility** column.

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.05

Futility

Boundary Family: Spending Functions

Spending Function: Gamma Family

Parameter (γ): -1

Type II Error (β): 0.2

Non-Binding
 Binding

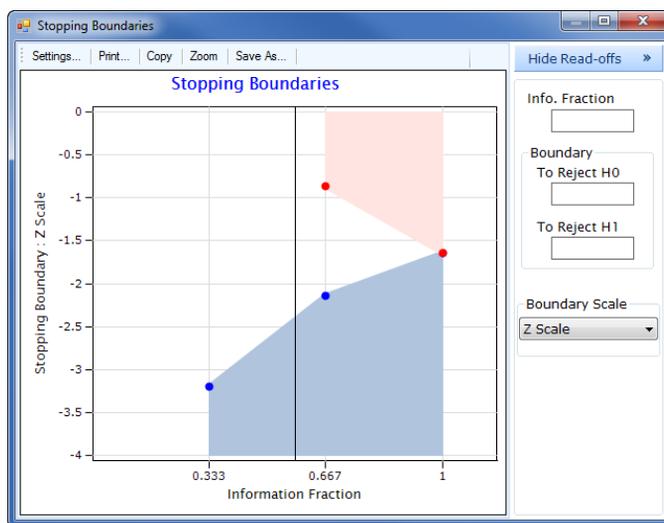
Spacing of Looks: Equal Unequal

Boundary Scale: Z Scale Recalc

| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Cum. β Spent | Futility Boundary |
|--------|----------------|-------------------------------------|-------------------------------------|---------------------|-------------------|--------------------|-------------------|
| 1 | 0.333 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | | | | |
| 2 | 0.667 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | | | | |
| 3 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.050 | | 0.200 | |

23 Binomial Superiority Two-Sample

Click **Recalc**, and click  icon to view the new boundaries. Notice that the futility boundary does not begin until the second look.

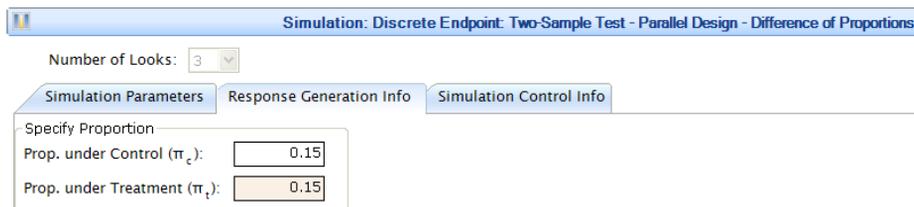


Simulation Tool

Let us verify the operating characteristics of Des8 from Wkbk1 through Simulations.

Select Des8 in the **Library**, and click  icon from **Library** toolbar. Alternatively, right-click on Des8 and select **Simulate**. A new Simulation worksheet will appear.

Let us first verify, by running 10,000 simulated clinical trials that the type-1 error is indeed 5%. That is, we must verify that if the event rate for both the placebo and treatment (Abciximab) arms is 0.15, only about 500 of these simulations will reject H_0 . Click on the **Response Generation** tab, and change the entry in the cell labeled **Prop. Under Treatment** from 0.1 to 0.15.



Next, click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled Sim1.

Select Sim1 in the **Output Preview**. Click  icon to save it to the **Library**. Double-click on Sim1 in the **Library**. The simulation output details will be displayed. In the **Details** output, notice that 487 of the 10,000 simulations rejected H_0 . (This number might vary, depending on the starting seed used for the simulations.) This confirms that the type-1 error is preserved (up to Monte Carlo accuracy) by these stopping boundaries.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Test Parameters | |
|-----------------------------------|-------------------|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 1098 |
| Variance | Unpooled Estimate |
| Avg. Power at Termination | 0.049 |
| Response Generation Parameters | |
| Prop. under Control (π_0) | 0.15 |
| Prop. under Treatment (π_1) | 0.15 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

☰ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 366 |
| 2 | 732 |
| 3 | 1098 |
| Average | 1091.778 |

☰ Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 366 | -3.2 | 5 | 5 | 0.050% |
| 2 | 732 | -2.141 | 160 | 160 | 1.600% |
| 3 | 1098 | -1.695 | 322 | 9835 | 98.350% |
| Total | | | 487 | 10000 | |
| % | | | 4.870% | | |

Simulation Seed and Elapsed Time

Starting Seed: 1323107
Total Number of Simulations: 10000
Elapsed Time: 00:00:04

Next, run 10,000 simulations under the alternative hypothesis H_1 that the event rate for placebo is 0.15 but the event rate for Abciximab is 0.1. Right-click Sim1 in the Library and click Edit Simulation. In the **Response Generation** tab, enter 0.10 for **Prop. Under Treatment**. Leave all other values as they are, and click **Simulate** to create output Sim2. Select Sim2 in the **Output Preview** and save it to Workbook Wbk1. In the **Overall Simulation Result** table, notice that the lower efficacy stopping boundary was crossed in 7996 out of 10000 simulated trials, which is consistent with 80% power (up to Monte Carlo accuracy) for the original design. Moreover, 393 of these simulations were able to reject the null hypothesis at the very first look. Feel free to

23 Binomial Superiority Two-Sample

experiment further with other simulation options before continuing.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Simulation Parameters | |
|-----------------------------------|-------------------|
| Simulation ID | Sim2 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 1156 |
| Variance | Unpooled Estimate |
| Avg. Power at Termination | 0.8 |
| Response Generation Parameters | |
| Prop. under Control (π_0) | 0.15 |
| Prop. under Treatment (π_1) | 0.1 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 385 |
| 2 | 771 |
| 3 | 1156 |
| Average | 896.43 |

Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|----------|--------------|----------|-------------------|---------|
| | | Efficacy | Futility | Efficacy | Futility | Count | % |
| 1 | 385 | -3.202 | 0.198 | 393 | 404 | 797 | 7.970% |
| 2 | 771 | -2.139 | -0.768 | 4450 | 696 | 5146 | 51.460% |
| 3 | 1156 | -1.633 | -1.633 | 3153 | 904 | 4057 | 40.570% |
| Total | | | | 7996 | 2004 | 10000 | |
| % | | | | 79.960% | 20.040% | | |

Simulation Seed and Elapsed Time

Starting Seed: 4677429
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:02

23.1.2 Interim Monitoring

The spending functions discussed above were for illustrative purposes only. They were not used in the actual CAPTURE trial. Instead, the investigators created their own spending function which is closely approximated by the Gamma spending function of Hwang, Shih and DeCani (1990) with parameter -4.5 . The investigators then used this spending function to generate two-sided boundaries for early stopping only to reject H_0 . Moreover since it was felt that the trial would enroll patients rapidly, the study was designed for three unequally spaced looks; one interim analysis after 25% enrollment, a second interim analysis after 50% enrollment, and a final analysis after all the patients had enrolled.

To design this trial, select Des2 in the **Library** and click  icon. In the **Boundary** tab, in the **Efficacy** box, set **Spending Function** to Gamma Family and change the **Parameter** (γ) to -4.5 . In the **Futility** Box, make sure **Boundary Family** is set to None. Click the radio button for **Unequal** in the **Spacing of Looks** box. In the **Looks Details** table change the **Info. Fraction** to 0.25 and 0.50 for Looks 1 and 2,

respectively.

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary

Efficacy
 Boundary Family: Spending Functions
 Spending Function: Gamma Family
 Parameter (γ): -4.5
 Type I Error (α): 0.05

Futility
 Boundary Family: None

Spacing of Looks
 Equal Unequal

Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|--------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.25 | 0.001 | 3.246 | -3.246 |
| 2 | 0.5 | 0.005 | 2.883 | -2.883 |
| 3 | 1 | 0.050 | 1.977 | -1.977 |

Click **Comptue**. In the **Output Preview** toolbar, click  icon to save this design to Wbk1 in the **Library**. Select Des9 in the **Library**, and click the **IM** icon from the **Library** toolbar. Alternatively, right-click on Des9 and select **Interim Monitoring** dashboard. The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs. The bottom section contains four charts, each with a corresponding table to its right. These charts provide graphical and numerical descriptions of the progress of the clinical trial and are useful tools for decision making by a data monitoring committee.

Click on the **Enter Interim Data** icon to invoke the **Test Statistic Calculator**. The first interim look was taken after accruing a total of 350 patients, 175 per treatment arm. There were 30 events on the placebo arm and 14 on the Abciximab arm. Based on these data, the event rate for placebo is $30/175 = 0.17143$ and the event rate for Abciximab is $14/175 = 0.08$. Hence the estimate of

$\delta = 0.08 - 0.17143 = -0.09143$. The unpooled estimate of the SE of $\hat{\delta}$ is

$$\sqrt{\frac{(14/175)(161/175)}{175} + \frac{(30/175)(145/175)}{175}} = 0.035103. \quad (23.7)$$

So the value of test statistic is

$$\frac{\hat{\delta}}{SE} = \frac{-0.09143}{0.035103} = -2.60457 \quad (23.8)$$

We will use the test statistic calculator and specify the values of $\hat{\delta}$ and SE in the same.

23 Binomial Superiority Two-Sample

The test statistic calculator will then compute the test statistic value and post it into the interim monitoring sheet. This process will ensure that the RCI and final adjusted estimates will be computed using the estimates of δ and SE obtained from the observed data.

Click on the **Estimate of δ and Std. Error of δ** radio button. Type in $(14/175) - (30/175)$ for **Estimate of δ** . The **Estimate of δ** is computed as -0.091429 . We can then enter the expression given by (23.7) for the **Std. Error of Estimate of δ** . Click on **Recalc** to get the **Test Statistic** value, then **OK** to continue.

The top panel of the interim monitoring worksheet displays upper and lower stopping boundaries and upper and lower 95% repeated confidence intervals. The lower stopping boundary for rejecting H_0 is -3.239 . Since the current value of the test statistic is -2.605 , the trial continues. The repeated confidence interval is $(-0.205, 0.022)$. We thus conclude, with 95% confidence, that Abciximab arm is unlikely to increase the event rate by any more than 2.2% relative to placebo and might actually reduce the event rate by as much as 20.5%.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | | 95% RCI for δ | | Repeat... p-value | CP | Predicti... Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|--------|----------------------|--------|-------------------|----|-------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.254 | 350 | -2.605 | -0.091 | 0.035 | 3.239 | -3.239 | 0.022 | -0.205 | 0.383 | 1 | 0.97 |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |

Now click on the second row in the table in the upper section. Then click the **Enter Interim Data** icon. A second interim look was taken after accruing a total of 700 patients, 353 on placebo and 347 on Abciximab. By this time point there were a total of 55 events on the placebo arm and 37 events on the Abciximab arm.

Based on these data, the event rate for placebo is $55/353 = 0.15581$ and the event rate for Abciximab is $37/347 = 0.10663$. Hence the estimate of $\delta = 0.10663 - 0.15581 = -0.04918$. The unpooled estimate of the SE of $\hat{\delta}$ is

$$\sqrt{\frac{(37/347)(310/347)}{347} + \frac{(55/353)(298/353)}{353}} = 0.02544. \quad (23.9)$$

So the value of test statistic is

$$\frac{\hat{\delta}}{SE} = \frac{-0.04918}{0.02544} = -1.9332 \quad (23.10)$$

We will now enter the above values of $\hat{\delta}$ and SE in the test statistic calculator for posting the test statistic value into the interim monitoring sheet. Enter the appropriate values for **Cumulative SS** and **Cumulative Response**. Click the **Recalc** button. The calculator updates the fields - total sample size, δ and SE.

Test Statistic Calculator

Editing look #2

Set Current Look as Last

Sample Size and Responses

| | Control | Treatment |
|-------------------------|---------|-----------|
| Cumulative Sample Size: | 353 | 347 |
| Cumulative Response: | 55 | 37 |

Cumulative Sample Size: 700

Input for Binomial end point

Estimate of δ : -0.049

$\delta = (\pi_1 - \pi_2)$

Standard Error of Estimate of δ : 0.025

Output

Test Statistic: -1.933

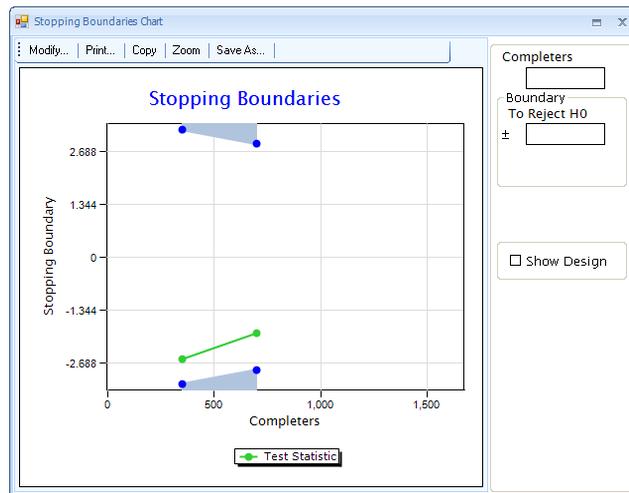
Recalc OK Cancel

23 Binomial Superiority Two-Sample

The updated sheet is displayed below.

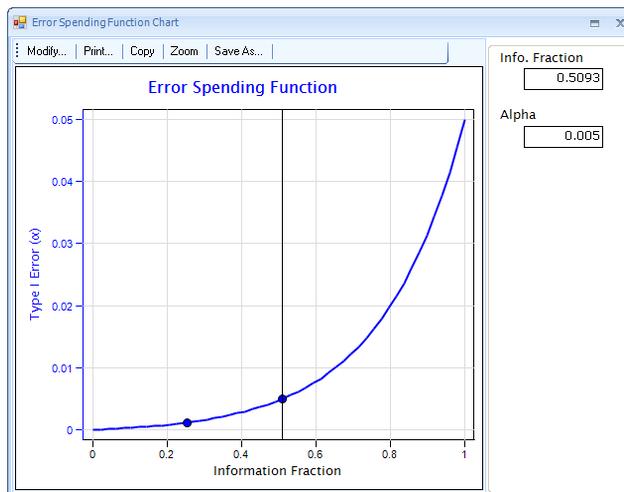
| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | | 95% RCI for δ | | Repeat... p-value | CP | Predicti... Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|--------|----------------------|--------|-------------------|-------|-------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.254 | 350 | -2.605 | -0.091 | 0.035 | 3.239 | -3.239 | 0.022 | -0.205 | 0.383 | 1 | 0.97 |
| 2 | 0.508 | 700 | -1.933 | -0.049 | 0.025 | 2.868 | -2.868 | 0.024 | -0.122 | 0.607 | 0.852 | 0.772 |
| 3 | | | | | | | | | | | | |

At this interim look, the stopping boundary for early rejection of H_0 is ± 2.868 and the 95% repeated confidence interval is still unable to exclude a difference of zero for the two event rates. Thus the study continues. The **Stopping Boundaries** chart of the dashboard displays the path traced out by the test statistic in relation to the upper and lower stopping boundaries at the first two interim looks. To expand this chart to full size, click the  icon located at the top right of the chart.



This full-sized chart displays stopping boundaries that have been recomputed on the basis of the error spent at each look, as shown on the **Error Spending** chart located at the bottom left of the dashboard. To display this full-sized chart, close the current chart

and click the  icon on the **Error Spending** chart.



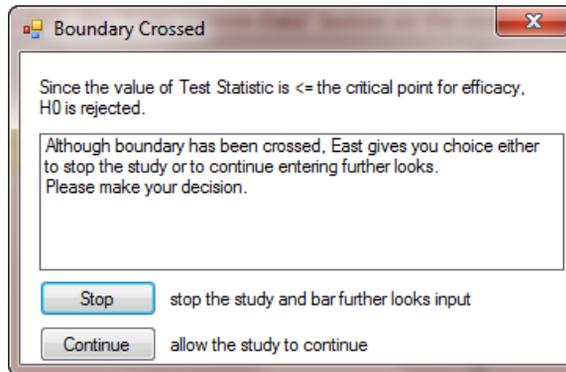
By moving the vertical cursor from left to right on this chart we observe that 0.0012 of the total error was spent by the first interim look and 0.005 of it was spent by the second interim look. Close this chart before continuing.

Although this study was designed for two interim looks and one final look, the data monitoring committee decided to take a third unplanned look after accruing 1050 patients, 532 on placebo and 518 on Abciximab. The error spending function methodology permits this flexibility. Both the timing and number of interim looks may be modified from what was proposed at the design stage. East will recompute the new stopping boundaries on the basis of the error actually spent at each look rather than the error that was proposed to be spent. There were 84 events on the placebo arm and 55 events on the Abciximab arm.

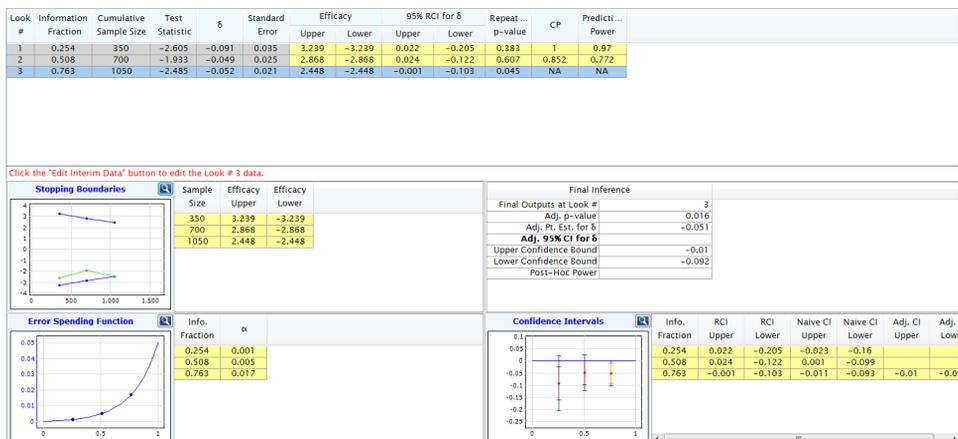
Hence the estimate of $\delta = 0.1062 - 0.1579 = -0.05171$. The unpooled estimate of the SE of δ is 0.02081. So the value of test statistic is -2.4849 . Click the third row of the table in the top portion and then click the **Enter Interim Data** icon. Upon entering this summary information, through the test statistic calculator, into the interim

23 Binomial Superiority Two-Sample

monitoring sheet we observe that the stopping boundary is crossed.



Press the **Stop** button and observe the results in the interim monitoring worksheet.



The 95% repeated confidence interval is $(-0.103, -0.011)$ and it excludes 0 thus confirming that the null hypothesis should be rejected. Once the study is terminated, East computes a final p-value, confidence interval and median unbiased point estimate, all adjusted for the multiple looks, using a stage wise ordering of the sample space as proposed by Tsiatis, Rosner and Mehta (1984). The adjusted p-value is 0.016. The adjusted confidence interval for the difference in event rates is $(-0.092, -0.010)$ and the median unbiased estimate of the difference in event rates is -0.051 . In general, the adjusted confidence interval produced at the end of the study is narrower than the final

repeated confidence interval although both intervals provide valid coverage of the unknown effect size.

23.1.3 Pooled versus Unpooled Designs

The manner in which the data will be analyzed at the interim monitoring stage should be reflected in the study design. We stated at the beginning of this chapter that the test statistic used to track the progress of a binomial endpoint study could be computed by using either the unpooled variance or the pooled variance to standardize the difference of binomial proportions. The design of the CAPTURE trial in Section 23.1.1 and its interim monitoring in Section 23.1.2 were both performed on the basis of the unpooled statistic. In this section we examine how the design would change if we intended to use the pooled statistic for the interim monitoring. It is seen that the change in sample size is negligible if the randomization is balanced. If, however, an unbalanced randomization rule is adopted, there can be substantial sample size differences between the unpooled and pooled designs.

Consider once more the design of the CAPTURE trial with a maximum of $K = 3$ looks, stopping boundaries generated by the **Gamma(-4.5)** Gamma family spending function, and 80% power to detect a drop in the event rate from 0.15 on the placebo arm to 0.1 on the Abciximab arm using a two sided level 0.05 test. We now consider the design of this trial on the basis of the pooled statistic.

Select Des9 in the **Library** and click  icon. Then under the **Test Parameters** tab, in the **Specify Variance** box, select the radio button for **Pooled Estimate**.

| | | | |
|----------------------------|----------|--|---|
| Test Type: | 2-Sided | Specify Proportion Response | <input type="checkbox"/> Perform Exact Computations |
| Type I Error (α): | 0.05 | Prop. under Control (π_c): | 0.15 |
| Power: | 0.8 | <input type="radio"/> Specify Variance | |
| Sample Size (n): | Computed | Specify Alternative Hypothesis | <input checked="" type="radio"/> Pooled Estimate |
| Allocation Ratio: | 1 | Prop. under Treatment (π_t): | 0.1 |
| (n_t/n_c) | | Diff. in Prop. ($\delta = \pi_t - \pi_c$): | -0.05 |
| | | <input type="checkbox"/> Use Casagrande-Pike-Smith Correction (Ignored if alloc. ratio is not 1) | |

Click the **Compute** button to create Des10. Save Des10 to Wbk1. In the **Library**

23 Binomial Superiority Two-Sample

select Des9 and Des10 by holding the Ctrl key, and then click on the  icon.

| | Des9 | Des10 |
|--|-------------------|-----------------|
| Mnemonic | PN-2S-DI | PN-2S-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 3 | 3 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.8 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Proportion under Control (π_c) | 0.15 | 0.15 |
| Proportion under Treatment (π_t) | 0.1 | 0.1 |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.05 | -0.05 |
| Variance | Unpooled Estimate | Pooled Estimate |
| Boundary Parameters | | |
| Spacing of Looks | Unequal | Unequal |
| Efficacy Boundary | Gm (-4.5) | Gm (-4.5) |
| Sample Size | | |
| Maximum | 1377 | 1383 |
| Expected Under H0 | 1373.311 | 1379.295 |
| Expected Under H1 | 1233.561 | 1239.862 |

It is instructive to compare Des9 with Des10. It is important to remember that Des9 utilized the unpooled design while Des10 utilized the pooled design.

When we compare Des9 and Des10 side by side we discover that there is not much difference in terms of either the maximum or expected sample sizes. This is usually the case for balanced designs. If, however, we were to change the value of the **Allocation Ratio** parameter from 1 to 0.333 (which corresponds to assigning 25% of the patients to treatment and 75% to control), then we would find a substantial difference in the sample sizes of the two plans. In the picture below, Des11 utilizes the unpooled design

while Des12 utilizes the pooled design.

| | Des11 | Des12 |
|--|-------------------|-----------------|
| Mnemonic | PN-2S-DI | PN-2S-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 3 | 3 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.8 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (nt / nc) | 0.333 | 0.333 |
| Proportion under Control (π_c) | 0.15 | 0.15 |
| Proportion under Treatment (π_t) | 0.1 | 0.1 |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.05 | -0.05 |
| Variance | Unpooled Estimate | Pooled Estimate |
| Boundary Parameters | | |
| Spacing of Looks | Unequal | Unequal |
| Efficacy Boundary | Gm (-4.5) | Gm (-4.5) |
| Sample Size | | |
| Maximum | 1679 | 1908 |
| Expected Under H0 | 1674.503 | 1902.894 |
| Expected Under H1 | 1504.039 | 1749.991 |

Notice that because of the unbalanced randomization the unpooled design is able to achieve 80% power with 229 fewer patients than the pooled design. Specifically, if we decide to monitor the study with the test statistic (23.2) we need to commit a maximum of 1908 patients (Des12), whereas if we decide to monitor the study with the test statistic (23.1) we need to commit a maximum of only 1679 patients (Des11). We can verify, by simulation that both Des11 and Des12 produce 80% power under the alternative hypothesis.

After saving Des11 and Des12 in Workbook1, select Des11 in the **Library** and click the  icon. Next, click the **Simulate** button. The results are displayed below and demonstrate that the null hypothesis was rejected 7710 times in 10,000 trials (77.10%),

23 Binomial Superiority Two-Sample

very close to the desired 80% power.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Test Parameters | |
|-----------------------------------|-------------------|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 2-Sided |
| Sample Size (n) | 1679 |
| Variance | Unpooled Estimate |
| Avg. Power at Termination | 0.771 |
| Response Generation Parameters | |
| Prop. under Control (π_c) | 0.15 |
| Prop. under Treatment (π_t) | 0.1 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 420 |
| 2 | 840 |
| 3 | 1679 |
| Average | 1488.859 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|--------|----------------|----------------|-------------------|---------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 420 | 3.246 | -3.246 | 0 | 464 | 464 | 4.640% |
| 2 | 840 | 2.882 | -2.882 | 0 | 1570 | 1570 | 15.700% |
| 3 | 1679 | 1.977 | -1.977 | 0 | 5676 | 7966 | 79.660% |
| Total | | | | 0 | 7710 | 10000 | |
| % | | | | 0.000% | 77.100% | | |

Simulation Seed and Elapsed Time

Starting Seed: 2524728
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:04

Next, repeat the procedure for Design12. Observe that once again, the desired power was almost achieved. This time the null hypothesis was rejected 7916 times in 10,000

trials (79.77%), just slightly under the desired 80% power.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Test Parameters | |
|-----------------------------------|-----------------|
| Simulation ID | Sim3 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 2-Sided |
| Sample Size (n) | 1908 |
| Variance | Pooled Estimate |
| Avg. Power at Termination | 0.792 |
| Response Generation Parameters | |
| Prop. under Control (π_0) | 0.15 |
| Prop. under Treatment (π_1) | 0.1 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

⊖ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 477 |
| 2 | 954 |
| 3 | 1908 |
| Average | 1774.535 |

⊖ Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|--------|----------------|----------------|-------------------|---------|
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 477 | 3.246 | -3.246 | 0 | 98 | 98 | 0.980% |
| 2 | 954 | 2.883 | -2.883 | 0 | 1252 | 1252 | 12.520% |
| 3 | 1908 | 1.977 | -1.977 | 0 | 6566 | 8650 | 86.500% |
| Total | | | | 0 | 7916 | 10000 | |
| % | | | | 0.000% | 79.160% | | |

Simulation Seed and Elapsed Time

Starting Seed: 3076919
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:03

The power advantage of the unpooled design over the pooled design gets reversed if the proportion of patients randomized to the treatment arm is 75% instead of 25%. Edit

23 Binomial Superiority Two-Sample

Des11 and Des12, and change the **Allocation Ratio** parameter to 3.

| | Wbk1:Des13 | Wbk1:Des14 |
|--|-------------------|-----------------|
| Mnemonic | PN-2S-DI | PN-2S-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 3 | 3 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.8 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 3 | 3 |
| Proportion under Control (π_c) | 0.15 | 0.15 |
| Proportion under Treatment (π_t) | 0.1 | 0.1 |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.05 | -0.05 |
| Variance | Unpooled Estimate | Pooled Estimate |
| Boundary Parameters | | |
| Spacing of Looks | Unequal | Unequal |
| Efficacy Boundary | Gm (-4.5) | Gm (-4.5) |
| Sample Size | | |
| Maximum | 1995 | 1770 |
| Expected Under H0 | 1989.657 | 1765.263 |
| Expected Under H1 | 1787.135 | 1547.235 |

Now the pooled design (Des14) requires a maximum of 1770 patients whereas the unpooled des (Des13) requires a maximum of 1995 patients. This shows that when planning a binomial study with unbalanced randomization, it is important to try both the pooled and unpooled designs and choose the one that produces the same power with fewer patients. The correct choice will depend on the response rates of the control and treatment arms as well as on the value of the fraction assigned to the treatment arm.

23.2 Ratio of Proportions

23.2.1 Trial Design

23.2.2 Trial Simulation

23.2.3 Interim Monitoring

Let π_c and π_t denote the binomial probabilities for the control and treatment arms, respectively, and let $\rho = \pi_t/\pi_c$. We want to test the null hypothesis that $\rho = 1$ against one or two-sided alternatives. It is mathematically more convenient to express this hypothesis testing problem in terms of the difference of the (natural) logarithms. Thus we define $\delta = \ln(\pi_t) - \ln(\pi_c)$. On this metric, we are interested in testing $H_0: \delta = 0$ against one or two-sided alternative hypotheses. Let $\hat{\pi}_{ij}$ denote the estimate of π_i based on n_{ij} observations from Treatment i , up to and including the j th look, $j = 1, \dots, K$, $i = t, c$, where a maximum of K looks are to be taken. Then the estimate of δ at the j -th look is

$$\hat{\delta}_j = \ln(\hat{\pi}_{tj}) - \ln(\hat{\pi}_{cj}) \quad (23.11)$$

with estimated standard error

$$\hat{s}e_j = \left\{ \frac{(1 - \hat{\pi}_{tj})}{n_{tj}\hat{\pi}_{tj}} + \frac{(1 - \hat{\pi}_{cj})}{n_{cj}\hat{\pi}_{cj}} \right\}^{1/2} \quad (23.12)$$

if we use an unpooled estimate for the variance of $\hat{\delta}$ and estimated standard error

$$\hat{s}e_j = \left\{ \frac{(1 - \hat{\pi}_j)}{\hat{\pi}_j} (n_{tj}^{-1} + n_{cj}^{-1}) \right\}^{1/2}, \quad (23.13)$$

where

$$\hat{\pi}_j = \frac{n_{tj}\hat{\pi}_{tj} + n_{cj}\hat{\pi}_{cj}}{n_{tj} + n_{cj}},$$

if we use a pooled estimate for the variance of $\hat{\delta}$.

In general, for any twice-differentiable function $h(\cdot)$, with derivative $h'(\cdot)$, $h(\hat{\pi}_{ij})$ is approximately normal with mean $h(\pi_i)$ and variance $[h'(\pi_i)]^2 \pi_i(1 - \pi_i)/n_{ij}$ for large values of n_{ij} . Using this asymptotic approximation, the test statistic at the j th look is

$$Z_j^{(u)} = \frac{\ln(\hat{\pi}_{tj}) - \ln(\hat{\pi}_{cj})}{\left\{ \frac{(1 - \hat{\pi}_{tj})}{n_{tj}\hat{\pi}_{tj}} + \frac{(1 - \hat{\pi}_{cj})}{n_{cj}\hat{\pi}_{cj}} \right\}^{1/2}}, \quad (23.14)$$

i.e. the ratio of (23.11) and (23.12), if we use an unpooled estimate for the variance of $\ln(\hat{\pi}_{tj}) - \ln(\hat{\pi}_{cj})$ and

$$Z_j^{(p)} = \frac{\ln(\hat{\pi}_{tj}) - \ln(\hat{\pi}_{cj})}{\left\{ \frac{(1 - \hat{\pi}_j)}{\hat{\pi}_j} (n_{tj}^{-1} + n_{cj}^{-1}) \right\}^{1/2}}, \quad (23.15)$$

i.e. the ratio of (23.11) and (23.13), if we use a pooled estimate for the variance of $\ln(\hat{\pi}_{tj}) - \ln(\hat{\pi}_{cj})$.

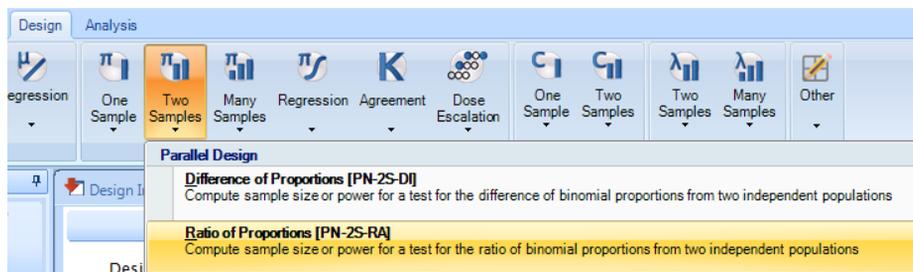
23.2.1 Trial Design

Design objectives and interim results were presented from PRISM, a prospective randomized trial of Heparin alone (control arm), Tirofiban alone (monotherapy arm), and Heparin plus Tirofiban (combination therapy arm), at a DIA Workshop on Flexible Trial Design (Snappin, 2003). The composite endpoint was refractory ischemia, myocardial infarct or death within seven days of randomization. The investigators were interested in comparing the two Tirofiban arms to the control arm with each test being conducted at the 0.025 level of significance (two sided). It was assumed that the control arm has a 30% event rate. Thus, $\pi_t = \pi_c = 0.3$ under H_0 . The investigators wished to determine the sample size to have power of 80% if there was a 25% decline

23 Binomial Superiority Two-Sample

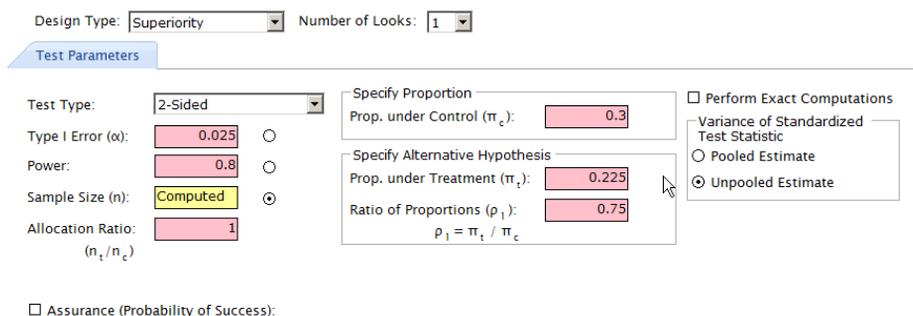
in the event rate, i.e. $\pi_t/\pi_c = 0.75$. It is important to note that the power of the test depends on π_c and π_t , not just the ratio, so different values of the pair (π_c, π_t) with the same ratio will have different solutions.

We will now design a two-arm study that compares the control arm, Heparin, to the combination therapy arm, Heparin plus Tirofiban. First click **Design** tab, then **Two Samples** on the **Discrete** group, and then click **Ratio of Proportions**.



We want to determine the sample size required to have power of 80% when $\pi_c=0.3$ and $\rho = \pi_t/\pi_c=0.75$, using a two-sided test with a type 1 error rate of 0.025.

Single-Look Design- Unpooled Estimate of Variance First consider a study with only one look and equal sample sizes in the two groups. Select the input parameters as displayed below.



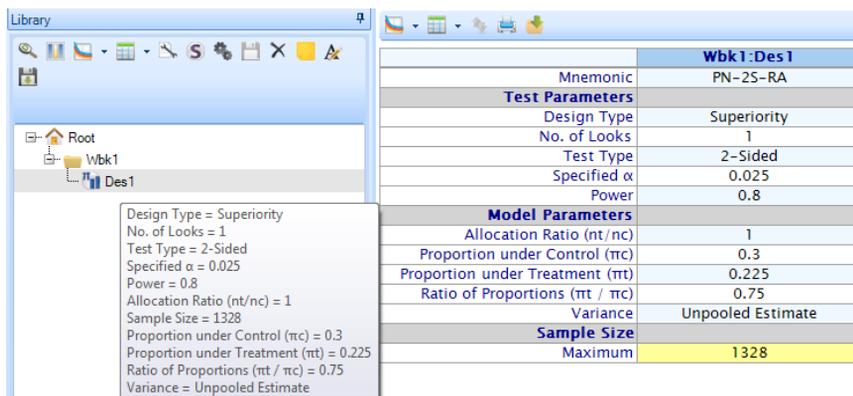
We will use the test statistic (23.14) with the unpooled estimate of the variance. Click the **Compute** button. The design Des1 is shown as a row in the **Output Preview**, located in the lower pane of this window. This single-look design requires a combined

total of 1328 subjects from both treatments in order to attain 80% power.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | π_c | Prop. Treatment (Alt.) | ρ_1 | Variance |
|------|-------------|--------------|-----------|--------------------|-------|-------|-------------|---------|------------------------|----------|-------------------|
| Des1 | Superiority | 1 | 2-Sided | 0.025 | 0.8 | 1 | 1328 | 0.3 | 0.225 | 0.75 | Unpooled Estimate |

You can select this design by clicking anywhere on the row in the **Output Preview**. If you click  , some of the design details will be displayed in the upper pane.

In the **Output Preview** toolbar, click  icon to save this design to Workbook wbk1 in the **Library**.



The screenshot shows the software interface. On the left is the 'Library' pane with a tree view containing 'Root', 'Wbk1', and 'Des1'. A tooltip is displayed over 'Des1' with the following text:

- Design Type = Superiority
- No. of Looks = 1
- Test Type = 2-Sided
- Specified α = 0.025
- Power = 0.8
- Allocation Ratio (nt/nc) = 1
- Sample Size = 1328
- Proportion under Control (π_c) = 0.3
- Proportion under Treatment (π_t) = 0.225
- Ratio of Proportions (π_t / π_c) = 0.75
- Variance = Unpooled Estimate

On the right is the 'Output Preview' pane showing a table for 'Wbk1:Des1' with the following parameters:

| Wbk1:Des1 | |
|--|-------------------|
| Mnemonic | PN-2S-RA |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 2-Sided |
| Specified α | 0.025 |
| Power | 0.8 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.3 |
| Proportion under Treatment (π_t) | 0.225 |
| Ratio of Proportions (π_t / π_c) | 0.75 |
| Variance | Unpooled Estimate |
| Sample Size | |
| Maximum | 1328 |

Three-Look Design - Unpooled Estimate of Variance For the above study, suppose we wish to take up to two equally spaced interim looks and one final look at the accruing data, using the Lan-DeMets (O'Brien-Fleming) stopping boundary. Create a new design by selecting Des1 in the **Library**, and clicking the  icon. In the input, change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. A new tab for **Boundary** will appear. Click this tab to reveal the stopping boundary parameters. By default, the Lan-DeMets (O'Brien-Fleming)

23 Binomial Superiority Two-Sample

stopping boundary and equal spacing of looks are selected.

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary

Efficacy
 Boundary Family: Spending Functions
 Spending Function: Lan-DeMets
 Parameter: OF
 Type I Error (α): 0.025

Futility
 Boundary Family: None

Spacing of Looks Equal Unequal Efficacy Boundary: Z Scale

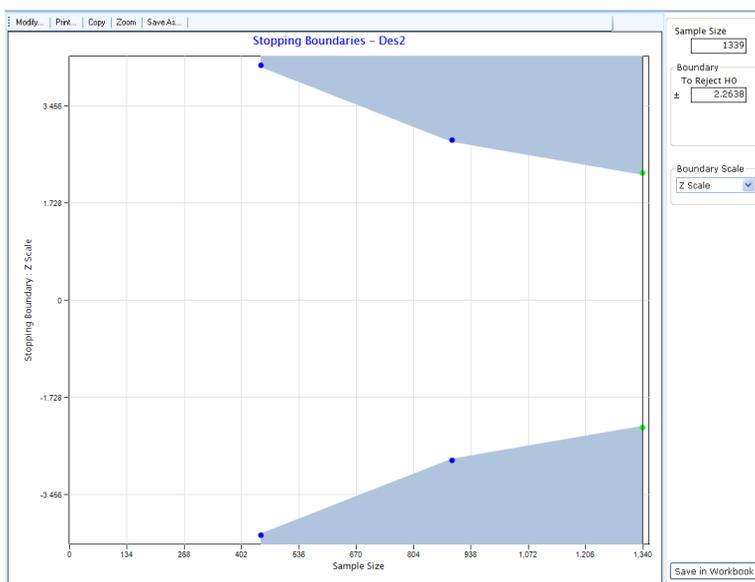
| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.333 | 0.000 | 4.171 | -4.171 |
| 2 | 0.667 | 0.004 | 2.846 | -2.846 |
| 3 | 1.000 | 0.025 | 2.264 | -2.264 |

Click **Compute** to create design Des2. The results of Des2 are shown in the **Output Preview** window. With Des2 selected in the **Output Preview**, click  icon. In the **Library**, select the nodes for both Des1 and Des2, by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the two designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 |
|--|-------------------|-------------------|
| Mnemonic | PN-2S-RA | PN-2S-RA |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 3 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.8 | 0.8 |
| Model Parameters | | |
| Proportion under Control (π_c) | 0.3 | 0.3 |
| Proportion under Treatment (π_t) | 0.225 | 0.225 |
| Ratio of Proportions (π_t / π_c) | 0.75 | 0.75 |
| Variance | Unpooled Estimate | Unpooled Estimate |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Boundary Parameters | | |
| Efficacy Boundary | | LD (OF) |
| Spacing of Looks | | Equal |
| Sample Size | | |
| Maximum | 1328 | 1339 |
| Expected Under H0 | | 1337.002 |
| Expected Under H1 | | 1167.474 |

Although, the maximum sample size has increased from 1328 to 1339, using three planned looks may result in a smaller sample size than that required for the single-look design, with an expected sample size of 1168 subjects under the alternative hypothesis ($\pi_c = 0.3, \rho = 0.75$), and still ensures that the power is 80%.

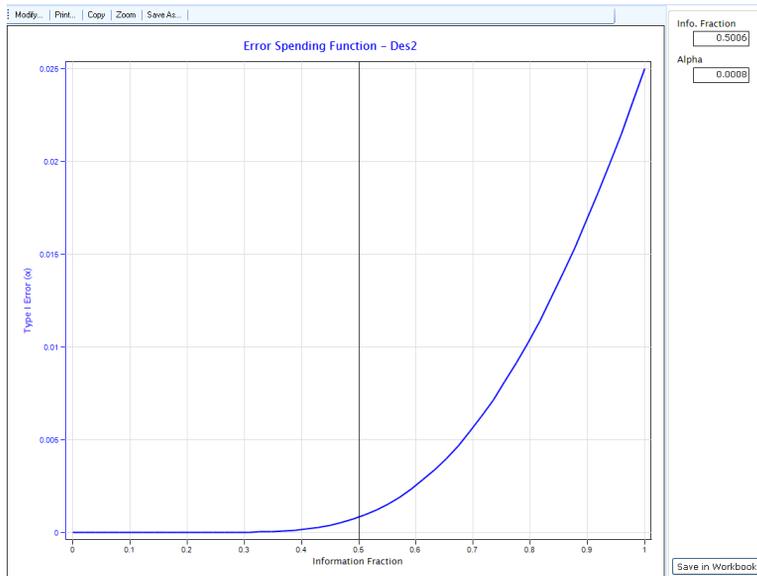
Additional information can also be obtained from Des2. The Lan-DeMets spending function corresponding to the O'Brien-Fleming boundary can be viewed by selecting Des2 in the **Library**, clicking on the  icon, and selecting **Stopping Boundaries**. The following chart will appear:



The alpha-spending function can be viewed by selecting Des2 in the **Library**, clicking

23 Binomial Superiority Two-Sample

on the  icon, and selecting **Error Spending**.



In order to see the stopping probabilities, as well as other characteristics, select Des2 in the **Library**, and click the  icon. The cumulative boundary stopping probabilities are shown in the **Stopping Boundaries** table.

 **Stopping Boundaries: Look by Look**

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | | Boundary Crossing Probability (Incremental) | | | |
|--------|--------------------------|-----------------|---------------------------|------------|--------|---|----------|----------|-------|
| | | | | Efficacy Z | | Under H0 | | Under H1 | |
| | | | | Upper | Lower | Efficacy | | Efficacy | |
| 1 | 0.333 | 446 | 3.013E-5 | 4.173 | -4.173 | 1.506E-5 | 1.506E-5 | 1.263E-9 | 0.009 |
| 2 | 0.667 | 893 | 0.004 | 2.845 | -2.845 | 0.002 | 0.002 | 3.838E-8 | 0.368 |
| 3 | 1 | 1339 | 0.025 | 2.264 | -2.264 | 0.01 | 0.01 | 3.834E-8 | 0.424 |

Close this window before continuing.

Three-Look Design - Pooled Estimate of Variance We now consider this design using the statistic (23.15) with the pooled estimate of the variance. Create a new design by selecting Des2 in the **Library**, and clicking the  icon. Under the **Test Parameters** tab, select the radio button for **Pooled Estimate** in the **Variance of**

Standardized Test Statistic box. Leave everything else unchanged. Click the **Compute** button to generate the output for Des3. Save Des3 by selecting it in the **Output Preview** and clicking the  icon. In the **Library**, select the nodes for Des1, Des2, and Des3, by holding the Ctrl Key, and then click the  icon. The upper pane will display the details of the three designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 | Wbk1:Des3 |
|--|-------------------|-------------------|-----------------|
| Mnemonic | PN-25-RA | PN-25-RA | PN-25-RA |
| Test Parameters | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 1 | 3 | 3 |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.025 | 0.025 | 0.025 |
| Power | 0.8 | 0.8 | 0.8 |
| Model Parameters | | | |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Proportion under Control (π_c) | 0.3 | 0.3 | 0.3 |
| Proportion under Treatment (π_t) | 0.225 | 0.225 | 0.225 |
| Ratio of Proportions (π_t / π_c) | 0.75 | 0.75 | 0.75 |
| Variance | Unpooled Estimate | Unpooled Estimate | Pooled Estimate |
| Boundary Parameters | | | |
| Spacing of Looks | | Equal | Equal |
| Efficacy Boundary | | LD (OF) | LD (OF) |
| Sample Size | | | |
| Maximum | 1328 | 1339 | 1373 |
| Expected Under H0 | | 1337.002 | 1370.956 |
| Expected Under H1 | | 1167.474 | 1200.52 |

For this problem, the test statistic (23.14) with the unpooled estimate of the variance requires a smaller sample size than the test statistic (23.15) with the pooled estimate of the variance. Close this window before continuing.

23.2.2 Trial Simulation

Suppose we want to see the impact of π_t on the behavior of the test statistic (23.14) with the unpooled estimate of the variance. First we consider $\pi_t = 0.225$ as specified by the alternative hypothesis. With Des2 selected in the **Library**, click the  icon. Click on the **Simulate** button. The results of the simulation will appear under Sim1 in the **Output Preview**. Select Sim1 in the **Output Preview** and click  icon to save it to Wbk1. Double-click on Sim1 in the **Library** to display the results of the simulation. Although the actual values may differ, we see that the power is

23 Binomial Superiority Two-Sample

approximately 80% and the probability of stopping early is about 0.37.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Ratio of Proportions

| Test Parameters | |
|--------------------------------------|-------------------|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 2-Sided |
| Sample Size (n) | 1339 |
| Variance | Unpooled Estimate |
| Avg. Power at Termination | 0.805 |
| Response Generation Parameters | |
| Prop. under Control (π_c) | 0.3 |
| Prop. under Treatment (π_{t1}) | 0.225 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 446 |
| 2 | 893 |
| 3 | 1339 |
| Average | 1174.154 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|--------|----------------|----------------|-------------------|---------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 446 | 4.173 | -4.173 | 0 | 42 | 36 | 0.420% |
| 2 | 893 | 2.845 | -2.845 | 0 | 3612 | 36 | 120% |
| 3 | 1339 | 2.264 | -2.264 | 0 | 4400 | 6346 | 63.460% |
| Total | | | | 0 | 8054 | 10000 | |
| % | | | | 0.000% | 80.540% | | |

Simulation Seed and Elapsed Time

Starting Seed: 897592
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:04

Now we consider $\pi_t = 0.25$, which will provide us with the impact if we were too optimistic about the treatment effect. Select Sim1 in the **Library** and click the  icon. Under the **Response Generation** tab, enter the value of 0.25 next to **Prop. Under Treatment (π_{t1})**. Click **Simulate** button. Although the actual values may

differ, we see that the power is approximately 41%.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Ratio of Proportions

| Test Parameters | |
|-----------------------------------|-------------------|
| Simulation ID | Sim2 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 2-Sided |
| Sample Size (n) | 1339 |
| Variance | Unpooled Estimate |
| Avg. Power at Termination | 0.407 |
| Response Generation Parameters | |
| Prop. under Control (π_0) | 0.3 |
| Prop. under Treatment (π_1) | 0.25 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

⊖ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 446 |
| 2 | 893 |
| 3 | 1339 |
| Average | 1290.965 |

⊖ Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|--------|----------------|----------------|-------------------|---------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 446 | 4.173 | -4.173 | 0 | 5 | 5 | 0.050% |
| 2 | 893 | 2.845 | -2.845 | 0 | 1067 | 1067 | 10.670% |
| 3 | 1339 | 2.264 | -2.264 | 0 | 2997 | 8928 | 89.280% |
| Total | | | | 0 | 4069 | 10000 | |
| % | | | | 0.000% | 40.690% | | |

Simulation Seed and Elapsed Time

Starting Seed: 1224804
Total Number of Simulations: 10000
Elapsed Time: 00:00:04

23.2.3 Interim Monitoring

Consider interim monitoring of Des2. Select Des2 in the **Library**, and click the **IM** icon from the **Library** toolbar. The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs. The bottom section contains four charts, each with a corresponding table to its right.

Suppose that the results are to be analyzed after results are available for every 450 subjects. Click on **Enter Interim Data** icon in the upper left to invoke the **Test Statistics Calculator**. Select the radio-button to enter $\hat{\delta}$ and its standard error. Enter 450 in the box next to **Cumulative Sample Size**. Suppose that after the data were available for first 450 subjects, 230 subjects were randomized to the control arm (c) and 220 subjects were randomized to the treatment arm (t). Of the 230 subjects in the control arm, there were 65 events; of the 220 subjects in the treatment arm, there were 45 events. In the box next to **Estimate of δ** enter: $\ln((45/220)/(65/230))$ and then hit Enter. EAST will compute the estimate of δ . Enter 0.169451 in the box next to **Std.**

23 Binomial Superiority Two-Sample

Error of δ . Next click **Recalc**. You should now see the following:

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Sample Size and Responses

| | Control | Treatment |
|-------------------------|---------|-----------|
| Cumulative Sample Size: | N/A | N/A |
| Cumulative Response: | N/A | N/A |

Cumulative Sample Size: 450

Input for Binomial end point

Estimate of δ : -0.323

$\delta = \ln(\pi_t / \pi_c)$

Standard Error of Estimate of δ : 0.169

Output

Test Statistic: -1.911

Recalc OK Cancel

Next, click **OK**. The following table will appear in the top section of the IM Dashboard.

| Look # | Information Fraction | Cumulative Completers | Test Statistic | Est. of p | Est. of δ | Std. Error of Est. of δ | Efficacy | | 97.5% RCI for p | | 97.5% RCI for δ | | Repeat ... p-value | CP | Predictive Power |
|--------|----------------------|-----------------------|----------------|-----------|------------------|--------------------------------|----------|--------|-----------------|-------|------------------------|--------|--------------------|-------|------------------|
| | | | | | | | Upper | Lower | Upper | Lower | Upper | Lower | | | |
| 1 | 0.336 | 450 | -1.908 | 0.724 | -0.323 | 0.169 | 4.153 | -4.153 | 1.463 | 0.358 | 0.38 | -1.027 | 0.407 | 0.899 | 0.771 |
| 2 | | | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | | | |

Note - Click on  icon to hide or unhide the columns of your interest. RCI for δ . Keeping all the four boxes checked can display RCI on both the scales.

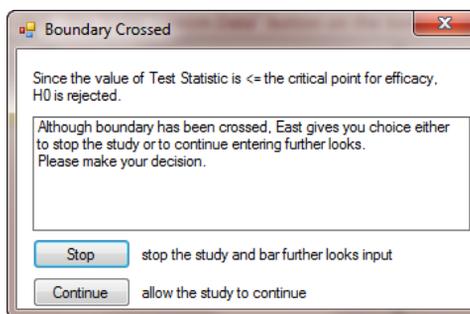
The boundary was not crossed as the value of the test statistic Test Statistic is -1.911, which is within the boundaries (-4.153, 4.153), so the trial continues. After data were available for an additional 450 subjects, the second analysis is performed. Suppose that among the 900 subjects, 448 were randomized to control (*c*) and 452 were randomized to (*t*). Of the 448 subjects in the control arm, there were 132 events; of the 452 subjects in the treatment arm, there were 90 events.

Click on the second row in the table in the upper section. Then click

Enter Interim Data

icon. Enter 900 box next to **Sample Size (Overall)**. Then in the box next to **Estimate of δ** enter: $\ln((90/452)/(132/448))$. Next hit **Enter**, then enter 0.119341 in the box next to **Std. Error of δ** . Click **Recalc** then **OK**.

The value of the test statistic is -3.284, which is less than -2.833, the value of the lower boundary, so the following dialog box appears.



Click on **Stop** to stop any further analyses. The **Final Inference** Table shows that the adjusted point estimate of $\ln(\rho)$ is -0.392 ($p = 0.001$) and the final adjusted 97.5% confidence interval for $\ln(\rho)$ is (-0.659, -0.124).

| Final Inference | |
|--|--------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.001 |
| Adj. Pt. Est. for p | 0.676 |
| Adj. 97.5% CI for p | |
| Upper Confidence Bound | 0.883 |
| Lower Confidence Bound | 0.517 |
| Adj. Pt. Est. for δ | -0.392 |
| Adj. 97.5% CI for δ | |
| Upper Confidence Bound | -0.124 |
| Lower Confidence Bound | -0.659 |

23 Binomial Superiority Two-Sample

23.3 Odds Ratio of Proportions

23.3.1 Trial Design

23.3.2 Trial Simulation

23.3.3 Interim Monitoring

Let π_t and π_c denote the two binomial probabilities associated with the treatment and the control, respectively. Furthermore, let the odds ratio be

$$\psi = \frac{\pi_t/(1 - \pi_t)}{\pi_c/(1 - \pi_c)} = \frac{\pi_t(1 - \pi_c)}{\pi_c(1 - \pi_t)}. \quad (23.16)$$

We are interested in testing $H_0: \psi = 1$ against the two-sided alternative $H_1: \psi \neq 1$ or against a one-sided alternative $H_1: \psi < 1$ or $H_1: \psi > 1$. It is convenient to express this hypothesis testing problem in terms of the (natural) logarithm of ψ . Let $\hat{\pi}_{tj}$ and $\hat{\pi}_{cj}$ denote the estimates of π_t and π_c based on n_{tj} and n_{cj} observations from the treatment and the control, respectively, up to and including the j th look, $j = 1, \dots, K$, where a maximum of K looks are to be made.

The difference between treatments at the j th look is assessed using

$$\hat{\delta}_j = \ln(\hat{\pi}_{tj}/(1 - \hat{\pi}_{tj})) - \ln(\hat{\pi}_{cj}/(1 - \hat{\pi}_{cj})). \quad (23.17)$$

Using the asymptotic approximation presented in section 23.2, the estimate of the standard error of $\hat{\delta}_j$ at the j th look is

$$\hat{s}e_j = \{1/n_{tj}\hat{\pi}_{tj}(1 - \hat{\pi}_{tj}) + 1/n_{cj}\hat{\pi}_{cj}(1 - \hat{\pi}_{cj})\}^{1/2}, \quad (23.18)$$

and the test statistic at the j -th look is the ratio of $\hat{\delta}_j$, given by (23.17), and the estimate of the standard error of $\hat{\delta}_j$, given by (23.18), namely,

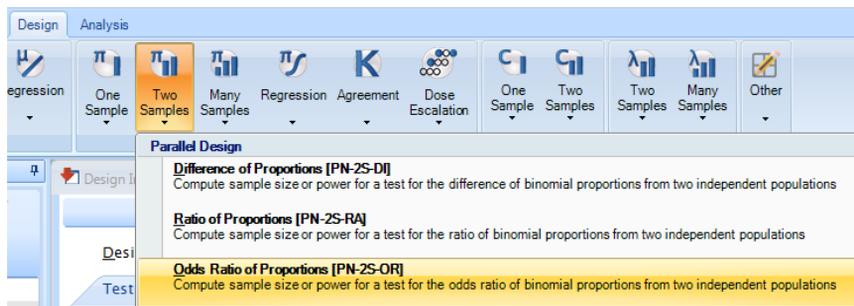
$$Z_j = \frac{\ln(\hat{\pi}_{tj}/(1 - \hat{\pi}_{tj})) - \ln(\hat{\pi}_{cj}/(1 - \hat{\pi}_{cj}))}{\{1/n_{tj}\hat{\pi}_{tj}(1 - \hat{\pi}_{tj}) + 1/n_{cj}\hat{\pi}_{cj}(1 - \hat{\pi}_{cj})\}^{1/2}}. \quad (23.19)$$

23.3.1 Trial Design

Suppose that the response rate for the control treatment is 10% and we hope that the experimental treatment can triple the odds ratio; that is, we desire to increase the response rate to 25%. Although we hope to increase the odds ratio, we solve this problem using a two-sided testing formulation. The null hypothesis $H_0: \psi = 1$ is tested against the two-sided alternative $H_1: \psi \neq 1$. The power of the test is computed at specified values of π_c and ψ . Note that the power of the test depends on π_c and ψ , or equivalently π_c and π_t , not just the odds ratio. Thus, different values of π_c with the same odds ratio will have different solutions.

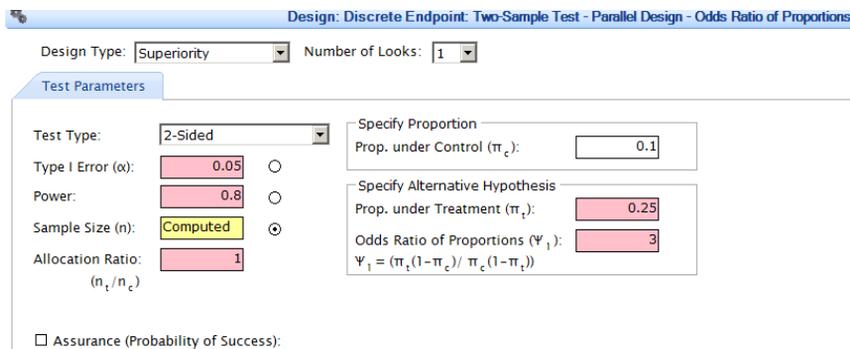
First, click **Design** tab, then click **Two Samples** in the **Discrete** group, and then click

Odds Ratio of Proportions.



Suppose we want to determine the sample size required to have power of 80% when $\pi_c = 0.1$ and $\psi_1 = 3$ using a two-sided test with a type-1 error rate of 0.05.

Single-Look Design First consider a study with only one look and equal sample sizes in the two groups. Enter the appropriate design parameters so that the dialog box appears as shown. Then click **Compute**.



The design Des1 is shown as a row in the **Output Preview**, located in the lower pane of this window. This single-look design requires a combined total of 214 subjects from

23 Binomial Superiority Two-Sample

both treatments in order to attain 80% power.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | π_c | Prop. Treatment (Alt.) | ψ_1 |
|------|-------------|--------------|-----------|--------------------|-------|-------|-------------|---------|------------------------|----------|
| Des1 | Superiority | 1 | 2-Sided | 0.05 | 0.8 | 1 | 214 | 0.1 | 0.25 | 3 |

You can select this design by clicking anywhere on the row in the **Output Preview**. If you click  icon, some of the design details will be displayed in the upper pane. In the **Output Preview** toolbar, click the  icon to save this design to Wbk1 in the **Library**.

| Wbk1:Des1 | |
|--|-------------|
| Mnemonic | PN-2S-OR |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.8 |
| Model Parameters | |
| Proportion under Control (π_c) | 0.1 |
| Proportion under Treatment (π_t) | 0.25 |
| Odds Ratio of Proportions (ψ_1) | 3 |
| Allocation Ratio (nt/nc) | 1 |
| Sample Size | |
| Maximum | 214 |

Three-Look Design For the above study, suppose we wish to take up to two equally spaced interim looks and one final look at the accruing data, using the Lan- DeMets (O’Brien-Fleming) stopping boundary. Create a new design by selecting Des1 in the **Library**, and clicking  icon. In the input, change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. A new tab for **Boundary** will appear. Click this tab to reveal the stopping boundary parameters. By default, the Lan-DeMets (O’Brien-Fleming) stopping boundary and equal spacing of

looks are selected.

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary

Efficacy
 Boundary Family: Spending Functions
 Spending Function: Lan-DeMets
 Parameter: OF
 Type I Error (α): 0.05

Futility
 Boundary Family: None

Spacing of Looks Equal Unequal Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|--------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.333 | 0.000 | 3.710 | -3.710 |
| 2 | 0.667 | 0.012 | 2.511 | -2.511 |
| 3 | 1.000 | 0.050 | 1.993 | -1.993 |

Click **Compute** button to design Des2. The results of Des2 are shown in the **Output Preview** window. With Des2 selected in the **Output Preview**, click the  icon. In the **Library**, select the rows for both Des1 and Des2, by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the two designs

23 Binomial Superiority Two-Sample

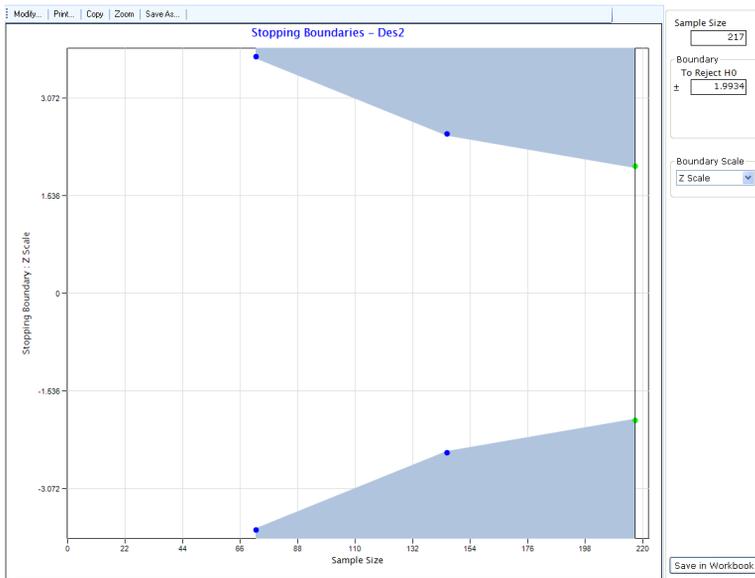
side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 |
|--|-------------|-------------|
| Mnemonic | PN-2S-OR | PN-2S-OR |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 3 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.8 | 0.801 |
| Model Parameters | | |
| Proportion under Control (π_c) | 0.1 | 0.1 |
| Proportion under Treatment (π_t) | 0.25 | 0.25 |
| Odds Ratio of Proportions (ψ) | 3 | 3 |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Boundary Parameters | | |
| Efficacy Boundary | | LD (OF) |
| Spacing of Looks | | Equal |
| Sample Size | | |
| Maximum | 214 | 217 |
| Expected Under H0 | | 216.106 |
| Expected Under H1 | | 185.392 |

Using three planned looks may result in a smaller sample size than that required for the single-look design, with an expected sample size of 186 subjects under the alternative hypothesis ($\pi_c = 0.1, \psi = 3$), and still ensures the power is 80%.

Additional information can also be obtained from Des2. The Lan-DeMets spending function corresponding to the O'Brien-Fleming boundary can be viewed by selecting Des2 in the **Library**, clicking on the  icon, and selecting **Stopping Boundaries**.

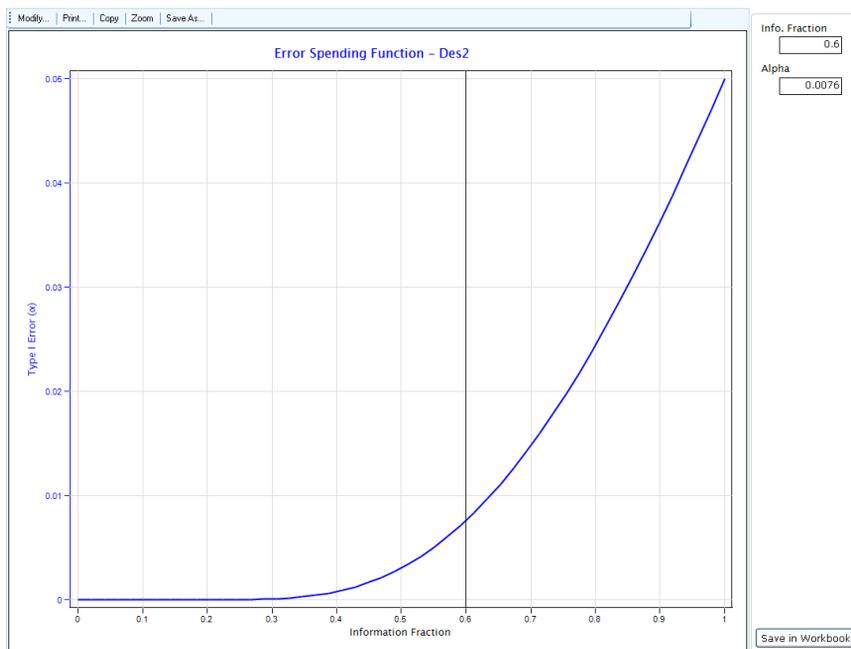
The following chart will appear:



The alpha-spending function can be viewed by selecting Des2 in the **Library**, clicking

23 Binomial Superiority Two-Sample

on the  icon, and selecting **Error Spending**.



In order to see the stopping probabilities, as well as other characteristics, select Des2 in the **Library**, and click the  icon. The cumulative boundary stopping probabilities are shown in the **Stopping Boundaries** table. East displays the stopping boundary, the type-1 error spent and the boundary crossing probabilities under $H_0: \pi_c = 0.1, \psi = 1$ and the alternative hypothesis $H_1: \pi_c = 0.1, \psi = 3$.

☯ **Stopping Boundaries: Look by Look**

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | | Boundary Crossing Probability (Incremental) | | | |
|--------|--------------------------|-----------------|--------------------|------------|--------|---|----------|----------|----------|
| | | | | | | Under H0 | | Under H1 | |
| | | | | Efficacy Z | | Efficacy | | Efficacy | |
| | | | | Upper | Lower | Upper | Lower | Upper | Lower |
| 1 | 0.332 | 72 | 1.995E-4 | 3.72 | -3.72 | 9.975E-5 | 9.975E-5 | 0.018 | 4.517E-8 |
| 2 | 0.668 | 145 | 0.012 | 2.508 | -2.508 | 0.006 | 0.006 | 0.403 | 7.326E-7 |
| 3 | 1 | 217 | 0.05 | 1.993 | -1.993 | 0.019 | 0.019 | 0.38 | 6.513E-7 |

Close this window before continuing.

23.3.2 Trial Simulation

Suppose we want to see the impact of π_t on the behavior of the test statistic (23.19). First we consider $\pi_t = 0.25$ as specified by the alternative hypothesis. With Des2 selected in the **Library**, click  icon. Next, click **Simulate** button. The results of the simulation will appear under Sim1 in the **Output Preview**. Highlight Sim1 in the **Output Preview** and click  icon to save it to workbook Wbk1. Double-click on Sim1 in the **Library** to display the results of the simulation. Although your results may differ slightly, we see that the power is approximately 83% and the probability of stopping early is about 0.39.

☰ Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|--------|----------------|----------------|-------------------|---------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 72 | 3.72 | -3.72 | 0 | 0 | 0 | 0.000% |
| 2 | 145 | 2.508 | -2.508 | 3878 | 0 | 3878 | 38.780% |
| 3 | 217 | 1.993 | -1.993 | 4432 | 0 | 6122 | 61.220% |
| Total | | | | 8310 | 0 | 10000 | |
| % | | | | 83.100% | 0.000% | | |

Now we consider $\pi_t = 0.225$, which will provide us with the impact if we were too optimistic about the treatment effect. Select Sim1 in the **Library** and click  icon. Under the **Response Generation** tab, enter the value of 0.225 next to **Prop. Under Treatment** (π_t). Click **Simulate**. Although, the actual values may differ, we see that the power is approximately 68% and the probability of stopping early is about 0.26.

☰ Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|--------|----------------|----------------|-------------------|---------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 72 | 3.72 | -3.72 | 0 | 0 | 0 | 0.000% |
| 2 | 145 | 2.508 | -2.508 | 2602 | 0 | 2602 | 26.020% |
| 3 | 217 | 1.993 | -1.993 | 4260 | 0 | 7398 | 73.980% |
| Total | | | | 6862 | 0 | 10000 | |
| % | | | | 68.620% | 0.000% | | |

23.3.3 Interim Monitoring

Consider interim monitoring of Des2. Select Des2 in the **Library**, and click 

23 Binomial Superiority Two-Sample

icon from the **Library** toolbar. The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs. The bottom section contains four charts, each with a corresponding table to its right.

Suppose that the results are to be analyzed after results are available for every 70 subjects. Click on **Enter Interim Data** icon in the upper left to invoke the **Test Statistics Calculator**. Select the second radio button on the calculator to enter values of $\hat{\delta}$ and its standard error. Before that enter 70 in the box next to **Cumulative Sample Size**. Suppose, after the data were available for first 70 subjects, 35 subjects were randomized to the control arm (c), of whom 5 experienced a response, and 35 subjects were randomized to the treatment arm (t), of whom 9 subjects experienced a response. In the box next to **Estimate of δ** enter 0.730888 and in the box next to **Std. Error of δ** enter 0.618794. Next click **Recalc**. You should now see the following:

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Sample Size and Responses

| | Control | Treatment |
|-------------------------|---------|-----------|
| Cumulative Sample Size: | N/A | N/A |
| Cumulative Response: | N/A | N/A |

Cumulative Sample Size: 70

Input for Binomial end point

Estimate of δ : 0.730888

$\delta = \ln(\pi_t(1 - \pi_c)/(\pi_c(1 - \pi_t)))$

Standard Error of Estimate of δ : 0.618794

Output

Test Statistic: 1.181

Recalc OK Cancel

Click **OK** and the following entry will appear in the top section of the IM Dashboard.

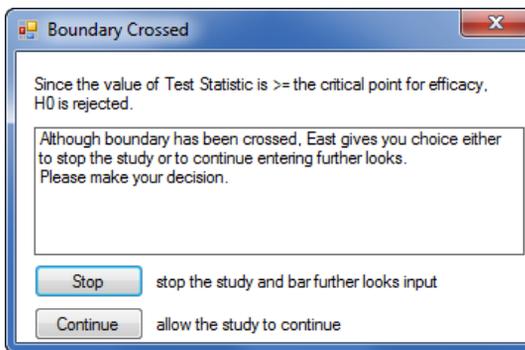
| Look # | Information Fraction | Cumulative Completers | Test Statistic | Est. of ψ | Est. of δ | Std. Error of Est. of δ | Efficacy | | 95% RCI for ψ | | 95% RCI for δ | | Repeat... p-value | CP | Predicti... Power |
|--------|----------------------|-----------------------|----------------|----------------|------------------|--------------------------------|----------|--------|--------------------|-------|----------------------|--------|-------------------|-------|-------------------|
| | | | | | | | Upper | Lower | Upper | Lower | Upper | Lower | | | |
| 1 | 0.323 | 70 | 1.181 | 2.077 | 0.731 | 0.619 | 3.777 | -3.777 | 21.502 | 0.201 | 3.068 | -1.606 | 0.751 | 0.548 | 0.53 |
| 2 | | | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | | | |

Note - Click on  icon to hide or unhide the columns of your interest.

The boundary was not crossed, as value of the test statistic (1.181) is within the boundaries (-3.777, 3.777), so the trial continues. After data were available for an additional 70 subjects, the second analysis was performed. Suppose that among the 140 subjects, 71 were randomized to *c* and 69 were randomized to *t*.

Click on the second row in the table in the upper section. Then click **Enter Interim Data** icon. Enter 140 in the box next to **Cumulative Sample Size**. Then in the box next to **Estimate of δ** enter: 1.067841 and in the box next to **Std. Error of δ** enter: 0.414083. Next, click on **Recalc** then **OK**.

The test statistic 2.579 exceeds the upper boundary (2.56), so the following screen appears.



Click **Stop** to halt any further analyses. The **Final Inference** Table shows that the adjusted point estimate of $\ln(\psi)$ is 1.068 ($p = 0.01$) and the adjusted 95% confidence

23 Binomial Superiority Two-Sample

interval for $\ln(\psi)$ is (0.256, 1.879).

| Final Inference | |
|--|-------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.01 |
| Adj. Pt. Est. for Ψ | 2.908 |
| Adj. 95% CI for Ψ | |
| Upper Confidence Bound | 6.548 |
| Lower Confidence Bound | 1.291 |
| Adj. Pt. Est. for δ | 1.068 |
| Adj. 95% CI for δ | |
| Upper Confidence Bound | 1.879 |
| Lower Confidence Bound | 0.256 |

23.4 Common Odds Ratio of Stratified Tables

23.4.1 Trial Design

23.4.2 Interim Monitoring

Some experiments are performed with several disjoint groups (strata) within each treatment group. For example, multicenter clinical trials are conducted using several investigator sites. Other situations include descriptive subsets, such as baseline and demographic characteristics. Let π_{tg} and π_{cg} denote the two binomial probabilities in Group g , $g = 1, \dots, G$, for the treatment and control, respectively. It is assumed that the odds ratio

$$\psi = \frac{\pi_{tg}/(1-\pi_{tg})}{\pi_{cg}/(1-\pi_{cg})} = \frac{\pi_{tg}(1-\pi_{cg})}{\pi_{cg}(1-\pi_{tg})} \quad (23.20)$$

is the same for each group (stratum). The Cochran-Mantel-Haensel test is used for testing $H_0: \psi = 1$ against the two-sided alternative $H_1: \psi \neq 1$ or against a one-sided alternative $H_1: \psi > 1$ or $H_1: \psi < 1$.

Let $\hat{\pi}_{tjg}$ and $\hat{\pi}_{c jg}$ denote the estimates of π_t and π_c based on n_{tjg} and $n_{c jg}$ observations in Group g from the treatment (t) and the control (c), respectively, up to and including the j th look, $j = 1, \dots, K$, where a maximum of K looks are to be taken.

Then the estimate of $\delta = \ln(\psi)$ from the g -th group at the j -th look is

$$\hat{\delta}_{jg} = \ln\left(\frac{\hat{\pi}_{tjg}}{1-\hat{\pi}_{tjg}}\right) - \ln\left(\frac{\hat{\pi}_{c jg}}{1-\hat{\pi}_{c jg}}\right).$$

Then the estimate of $\delta = \ln(\psi)$ at the j -th look is the average of $\hat{\delta}_{jg}$, $g = 1, \dots, G$;

namely,

$$\hat{\delta}_j = \frac{\sum_{g=1}^G \hat{\delta}_{jg}}{G}$$

or, equivalently,

$$\hat{\delta}_j = \frac{\sum_{g=1}^G (\ln(\frac{\hat{\pi}_{tjg}}{1-\hat{\pi}_{tjg}}) \ln(\frac{\hat{\pi}_{cjg}}{1-\hat{\pi}_{cjg}}))}{G}. \tag{23.21}$$

The estimate of the standard error of $\hat{\delta}_{jg}$ at the j th look is

$$\hat{se}_{jg} = \left\{ \frac{1}{n_{tjg} \hat{\pi}_{tjg} (1 - \hat{\pi}_{tjg})} + \frac{1}{n_{cjg} \hat{\pi}_{cjg} (1 - \hat{\pi}_{cjg})} \right\}^{1/2}.$$

The estimated variance of $\hat{\delta}$ at the j -th look is the average of the variances of $\hat{\delta}_{jg}, g = 1, \dots, G$. Thus,

$$\hat{se}_j = \left\{ \frac{\sum_{g=1}^G \hat{se}_{jg}^2}{G} \right\}^{1/2}. \tag{23.22}$$

The test statistic used at the j -th look is

$$Z_j = \frac{\hat{\delta}_j}{\hat{se}_j}. \tag{23.23}$$

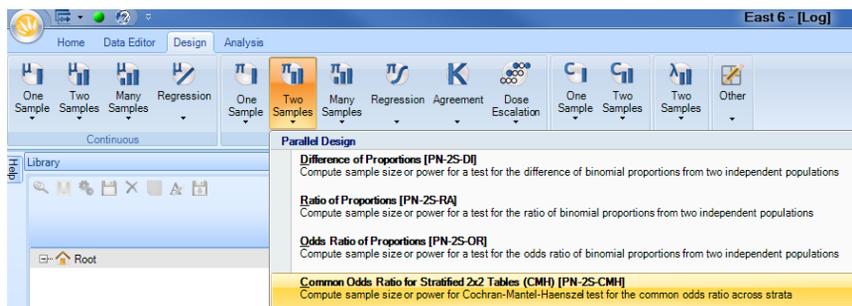
23.4.1 Trial Design

First consider a simple example with two strata, such as males and females, with an equal number of subjects in each stratum and the same response rate of 60% for the control in each stratum. We hope that the experimental treatment can triple the odds ratio. Although we hope to increase the odds ratio, we solve this problem using a two-sided testing formulation. The null hypothesis $H_0: \psi = 1$ is tested against the two-sided alternative $H_1: \psi \neq 1$. The power of the test is computed at specified values of $\pi_{cg}, g = 1, \dots, G$, and ψ .

To begin, click **Design** tab, then click **Two Samples** in the **Discrete** group, and then

23 Binomial Superiority Two-Sample

click **Common Odds Ratio for Stratified 2 x 2 Tables**.



Suppose that we want to determine the sample size required to have power of 80% when $\pi_{c1} = \pi_{c2} = 0.6$ and $\psi = 3$ using a two-sided test with a type-1 error rate of 0.05.

Single-Look Design - Equal Response Rates First consider a study with only one look and equal sample sizes in the two groups. Enter the appropriate test parameters so that the dialog box appears as shown. Then click **Compute**.

Design Type: Superiority Number of Looks: 1

Test Parameters

Test Type: 2-Sided Common Odds Ratio (Ψ): 3 Number of Strata: 2
 $\Psi = \pi_t (1 - \pi_c) / \pi_c (1 - \pi_t)$ (G)

Type I Error (α): 0.05 Power: 0.8

Sample Size (n): Computed Allocation Ratio: 1
 (n_t / n_c)

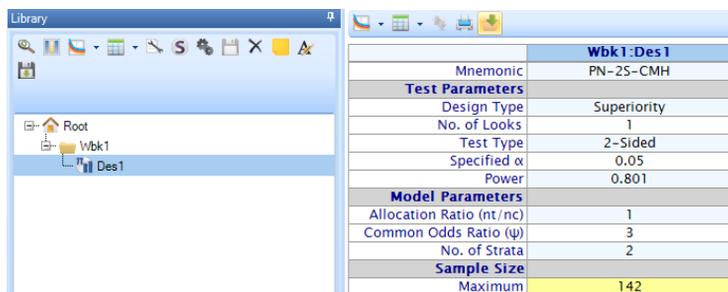
Stratum Specific Input: Equal Unequal

| Stratum # | Stratum Fraction | π_c | π_t |
|-----------|------------------|---------|---------|
| 1 | 0.500 | 0.6 | 0.818 |
| 2 | 0.500 | 0.6 | 0.818 |

The design is shown as a row in the **Output Preview**, located in the lower pane of this window. This single-look design requires a combined total of 142 subjects from both treatments in order to attain 80% power.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | ψ | No. of Strata |
|-------|-------------|--------------|-----------|--------------------|-------|-------|-------------|--------|---------------|
| Des 1 | Superiority | 1 | 2-Sided | 0.05 | 0.801 | 1 | 142 | 3 | 2 |

You can select this design by clicking anywhere on the row in the **Output Preview**. If you click  icon, some of the design details will be displayed in the upper pane. In the **Output Preview** toolbar, click  icon to save this design to workbook Wbk1 in the **Library**.



| Wbk1:Des1 | |
|------------------------------|-------------|
| Mnemonic | PN-2S-CMH |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.801 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Common Odds Ratio (ψ) | 3 |
| No. of Strata | 2 |
| Sample Size | |
| Maximum | 142 |

Single-Look Design - Unequal Response Rates Now, we consider a more realistic clinical trial. Suppose that males and females respond differently, so that the response rate for males is $\pi_{c1} = 0.6$ and the response rate for females is $\pi_{c2} = 0.3$. First, we consider a study without any interim analyses.

Create a new design by selecting Des1 in the **Library**, and clicking the  icon. Change π_{c2} in the **Stratum Specific Input** table to 0.3 as shown below.

Design Type: Superiority Number of Looks: 1

Test Parameters

Test Type: 2-Sided Common Odds Ratio (Ψ): 3 Number of Strata: 2
 $\Psi = \pi_t (1 - \pi_c) / \pi_c (1 - \pi_t)$ (G)

Type I Error (α): 0.05

Power: 0.8

Sample Size (n): Computed

Allocation Ratio: 1
 (n_t/n_c)

Stratum Specific Input Equal Unequal

| Stratum # | Stratum Fraction | $\pi_{c,c}$ | $\pi_{c,t}$ |
|-----------|------------------|-------------|-------------|
| 1 | 0.500 | 0.6 | 0.818 |
| 2 | 0.500 | 0.3 | 0.563 |

Click **Compute** to create design Des2. The results of Des2 are shown in the **Output Preview** window. With Des2 selected in the **Output Preview**, click  icon. In the **Library**, select the rows for both Des1 and Des2, by holding the Ctrl key, and then

23 Binomial Superiority Two-Sample

click the  icon. The upper pane will display the details of the two designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 |
|------------------------------|-------------|-------------|
| Mnemonic | PN-2S-CMH | PN-2S-CMH |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 1 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.801 | 0.801 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Common Odds Ratio (ψ) | 3 | 3 |
| No. of Strata | 2 | 2 |
| Sample Size | | |
| Maximum | 142 | 127 |

This single-look design requires a combined total of 127 subjects from both treatments in order to attain 80% power.

Three-Look Design - Unequal Response Rates For the above study, suppose we wish to take up to two equally spaced interim looks and one final look at the accruing data, using the Lan-DeMets (O'Brien-Fleming) stopping boundary. Create a new design by selecting Des2 in the **Library**, and clicking the  icon. In the input, change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. A new tab for **Boundary** will appear. Click this tab to reveal the stopping boundary parameters. By default, the Lan-DeMets (O'Brien-Fleming) stopping boundary and equal spacing of looks are selected.

Design Type: Superiority Number of Looks:

Test Parameters **Boundary**

Efficacy

Boundary Family:

Spending Function:

Parameter:

Type I Error (α): 0.05

Futility

Boundary Family:

Spacing of Looks Equal Unequal Efficacy Boundary:  

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.333 | 0.000 | 3.710 | -3.710 |
| 2 | 0.667 | 0.012 | 2.511 | -2.511 |
| 3 | 1.000 | 0.050 | 1.993 | -1.993 |

Click the **Compute** button to generate output for Des3. The results of Des3 are shown in the **Output Preview** window. With Des3 selected in the **Output Preview**, click  icon. In the **Library**, select the nodes for Des1, Des2, and Des3 by holding the Ctrl Key, and then click the  icon. The upper pane will display the details of the three designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 | Wbk1:Des3 |
|------------------------------|-------------|-------------|-------------|
| Mnemonic | PN-2S-CMH | PN-2S-CMH | PN-2S-CMH |
| Test Parameters | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 1 | 1 | 3 |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.801 | 0.801 | 0.802 |
| Model Parameters | | | |
| Common Odds Ratio (ψ) | 3 | 3 | 3 |
| No. of Strata | 2 | 2 | 2 |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Boundary Parameters | | | |
| Spacing of Looks | | | Equal |
| Efficacy Boundary | | | LD (OF) |
| Sample Size | | | |
| Maximum | 142 | 127 | 129 |
| Expected Under H0 | | | 128.471 |
| Expected Under H1 | | | 110.157 |

Using three planned looks requires an up-front commitment of 129 subjects, a slight increase over the single-look design, which required 127 subjects. However, the three look design may result in a smaller sample size than that required for the single look design, with an expected sample size of 111 subjects under the alternative hypothesis ($\pi_{c1} = 0.6, \pi_{c2} = 0.3, \psi = 3$), and still ensures that the power is 80%.

By selecting only Des3 in the **Library** and clicking  icon, East displays the stopping boundary, the type-1 error spent and the boundary crossing probabilities under $H_0: \pi_{c1} = 0.6, \pi_{c2} = 0.3, \psi = 1$ and the alternative hypothesis

23 Binomial Superiority Two-Sample

$$H_1: \pi_{c1} = 0.6, \pi_{c2} = 0.3, \psi = 3.$$

☰ Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | | Boundary Crossing Probability (Incremental) | | | |
|--------|--------------------------|-----------------|---------------------------|------------|--------|---|----------|----------|----------|
| | | | | Efficacy Z | | Under H0 | | Under H1 | |
| | | | | Upper | Lower | Upper | Lower | Upper | Lower |
| 1 | 0.333 | 43 | 2.07E-4 | 3.71 | -3.71 | 1.035E-4 | 1.035E-4 | 0.019 | 4.605E-8 |
| 2 | 0.667 | 86 | 0.012 | 2.511 | -2.511 | 0.006 | 0.006 | 0.401 | 7.186E-7 |
| 3 | 1 | 129 | 0.05 | 1.993 | -1.993 | 0.019 | 0.019 | 0.382 | 6.416E-7 |

Close this window before continuing.

Three-Look Design - Unequal Response Rates - Unequal Strata Sizes Some disorders have different prevalence rates across various strata. Consider the above example, but with the expectation that 30% of the subjects will be males and 70% of the subjects will be females. Create a new design by selecting Des3 in the **Library**, and clicking the  icon. Under the **Test Parameters** tab in the **Stratum Specific Input** box select the radio button **Unequal**. You can now edit the **Stratum Fraction** column for Stratum 1. Change this value from 0.5 to 0.3 as shown below.

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary

Test Type: 2-Sided Common Odds Ratio (Ψ): 3 Number of Strata: 2

Type I Error (α): 0.05 $\Psi = \pi_t (1 - \pi_c) / \pi_c (1 - \pi_t)$ (G)

Power: 0.8 Equal Unequal

Sample Size (n): Computed

Allocation Ratio: 1 (n_t/n_c)

| Stratum # | Stratum Fraction | π_c | π_t |
|-----------|------------------|---------|---------|
| 1 | 0.3 | 0.6 | 0.818 |
| 2 | 0.7 | 0.3 | 0.563 |

Click the **Compute** button to generate output for Des4. The results of Des4 are shown in the **Output Preview** window. With Des4 selected in the **Output Preview**, click the  icon. In the **Library**, select the rows for Des1, Des2, Des3, and Des4 by holding the Ctrl key, and then click  icon. The upper pane will display the details of the

four designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 | Wbk1:Des3 | Wbk1:Des4 |
|------------------------------|-------------|-------------|-------------|-------------|
| Mnemonic | PN-2S-CMH | PN-2S-CMH | PN-2S-CMH | PN-2S-CMH |
| Test Parameters | | | | |
| Design Type | Superiority | Superiority | Superiority | Superiority |
| No. of Looks | 1 | 1 | 3 | 3 |
| Test Type | 2-Sided | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 | 0.05 |
| Power | 0.801 | 0.801 | 0.802 | 0.802 |
| Model Parameters | | | | |
| Common Odds Ratio (ψ) | 3 | 3 | 3 | 3 |
| No. of Strata | 2 | 2 | 2 | 2 |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 | 1 |
| Boundary Parameters | | | | |
| Spacing of Looks | | | Equal | Equal |
| Efficacy Boundary | | | LD (OF) | LD (OF) |
| Sample Size | | | | |
| Maximum | 142 | 127 | 129 | 124 |
| Expected Under H0 | | | 128.471 | 123.487 |
| Expected Under H1 | | | 110.157 | 105.856 |

Note that, for this example, unequal sample sizes for the two strata result in a smaller total sample size than that required for equal sample sizes for the two strata.

23.4.2 Interim Monitoring

Consider interim monitoring of Des4. Select Des4 in the **Library**, and click the **IM** icon from the **Library** toolbar. The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs. The bottom section contains four charts, each with a corresponding table to its right.

Suppose that the results are to be analyzed after results are available for every 40 subjects. Click on the **Enter Interim Data** icon in the upper left to invoke the **Test Statistics Calculator**. Enter 40 in the box next to **Cumulative Sample Size**. Suppose that $\hat{\delta}_1 = 0.58$ and $\hat{s}e_1 = 0.23$. Enter these values and click on **Recalc**. You should

23 Binomial Superiority Two-Sample

now see the following:

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Cumulative Sample Size:

Input for Binomial end point

Estimate of δ :
 $\delta = (\sum \ln(\psi_g))/G$

Standard Error of Estimate of δ :

Output

Test Statistic:

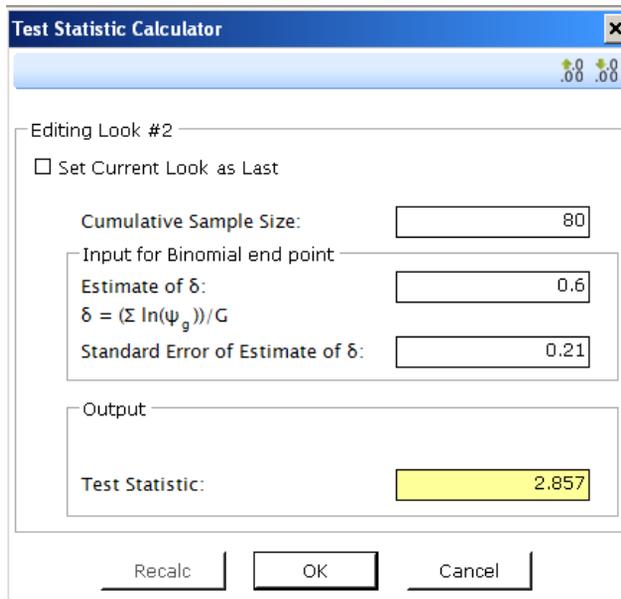
Recalc OK Cancel

Click **OK** and the following table will appear in the top section of the IM Dashboard.

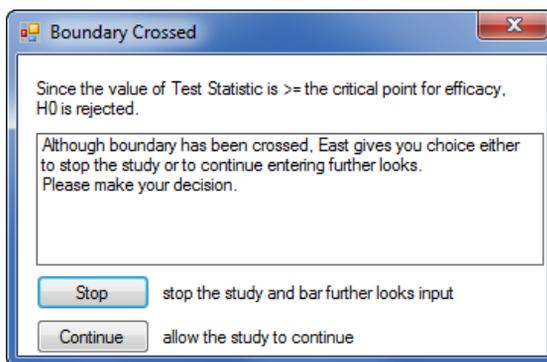
| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | | 95% RCI for δ | | Repeat ... p-value | CP | Predicti... Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|--------|----------------------|--------|--------------------|-------|-------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.323 | 40 | 2.522 | 0.58 | 0.23 | 3.777 | -3.777 | 1.449 | -0.289 | 0.235 | 0.999 | 0.959 |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |

The boundary was not crossed, as value of the test statistic (2.522) is within the boundaries (-3.777, 3.777), so the trial continues. Click on the second row in the table in the upper section. Then click the **Enter Interim Data** icon. Enter 80 in the box next to **Cumulative Sample Size**. Suppose that $\hat{\delta}_2 = 0.60$ and $\hat{s}e_2 = 0.21$. Enter these

values and click **Recalc**. You should now see the following:



Click the **OK** button. The test statistic 2.857 exceeds the upper boundary (2.56), so the following dialog box appears.



Click **Stop** to stop any further analyses. The **Final Inference** Table shows the adjusted point estimate of $\ln(\psi)$ is 0.600 ($p = 0.004$) and the adjusted 95% confidence interval

23 Binomial Superiority Two-Sample

for $\ln(\psi)$ is (0.188, 1.011).

| Final Inference | |
|--|-------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.004 |
| Adj. Pt. Est. for δ | 0.6 |
| Adj. 95% CI for δ | |
| Upper Confidence Bound | 1.011 |
| Lower Confidence Bound | 0.188 |
| Post-Hoc Power | |

23.5 Fisher's Exact Test (Single Look)

23.5.1 Trial Design

In some experimental situations, the normal approximation to the binomial distribution may not be appropriate, such as the probabilities of interest are large or small. This may lead to incorrect p-values, and thus the incorrect conclusion. For this reason, Fisher's exact test may be used. Let π_t and π_c denote the two response probabilities for the treatment and the control, respectively. Interest lies in testing $H_0: \pi_t = \pi_c$ against the two-sided alternative $H_1: \pi_t \neq \pi_c$. Results are presented here only for the situation where there is a single analysis; that is, no interim analysis, for the two-sided test with equal sample sizes for the two treatments.

Let $\hat{\pi}_t$ and $\hat{\pi}_c$ denote the estimates of π_t and π_c , respectively, based on $n_t = n_c = 0.5N$ observations from the treatment (t) and the control (c). The parameter of interest is $\delta = \pi_t - \pi_c$, which is estimated by $\hat{\delta} = \hat{\pi}_t - \hat{\pi}_c$. The estimate of the standard error used in the proposed test statistic uses of the pooled estimate of the common value of π_t and π_c under H_0 , given by

$$\hat{s}e = \frac{2\{\hat{\pi}(1 - \hat{\pi})\}^{1/2}}{N}, \tag{23.24}$$

where $\hat{\pi} = 0.5(\hat{\pi}_t + \hat{\pi}_c)$.

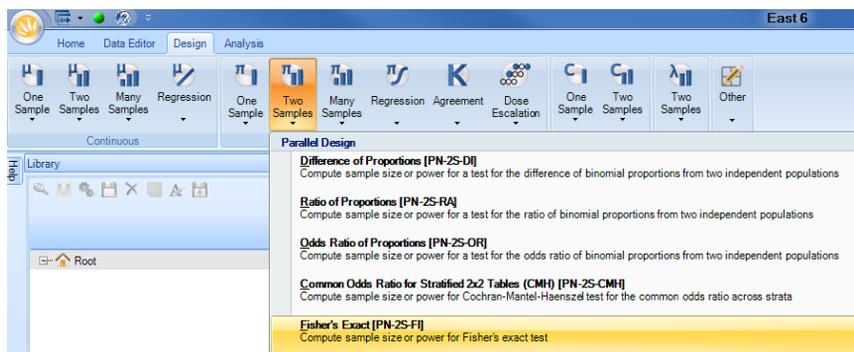
Incorporating a continuity correction factor, the test statistic is

$$Z = \frac{|\hat{\delta}|2/N}{\hat{s}e}. \tag{23.25}$$

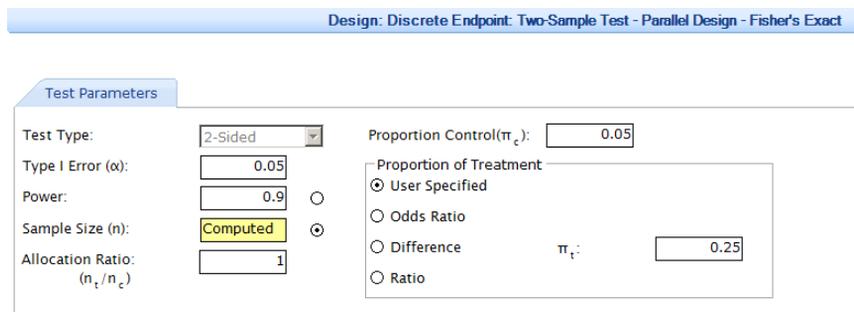
23.5.1 Trial Design

Consider the example where the probability of a response for the control is 5% and it is

hoped that the experimental treatment can increase this rate to 25%. First, in the **Discrete** area, click **Two Samples** on the **Design** tab, and then click **Fisher Exact Test**.



Suppose we want to determine the sample size required to have power of 90% when $\pi_c = 0.05$ and $\pi_t = 0.25$ using a two-sided test with a type-1 error rate of 0.05. Enter the appropriate test parameters so that the dialog box appears as shown. Then click **Compute**.

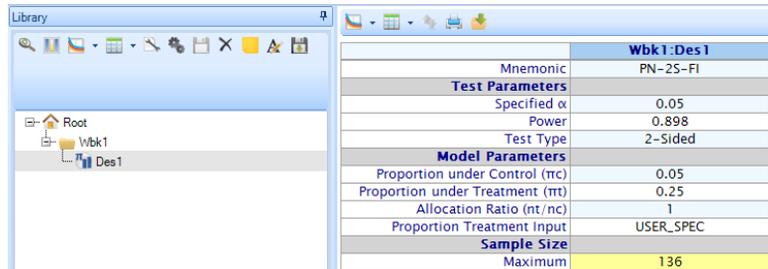


The design is shown as a row in the **Output Preview**, located in the lower pane of this window. This single-look design requires a combined total of 136 subjects from both treatments in order to attain 90% power.

| ID | Specified α | Power | Sample Size | Test Type | π_c | Prop. Treatment (Alt.) | nt / nc | Treat. Input |
|------|--------------------|-------|-------------|-----------|---------|------------------------|---------|--------------|
| Des1 | 0.05 | 0.898 | 136 | 2-Sided | 0.05 | 0.25 | 1 | USER_SPEC |

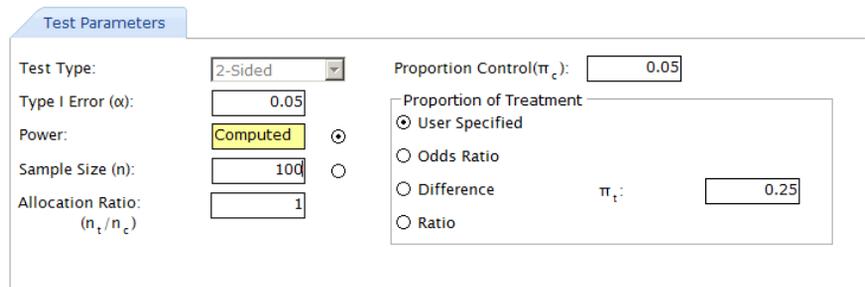
23 Binomial Superiority Two-Sample

You can select this design by clicking anywhere along the row in the **Output Preview**. If you click  icon, some of the design details will be displayed in the upper pane. In the **Output Preview** toolbar, click the  icon to save this design to Workbook1 in the **Library**.



| Wbk1:Des1 | |
|--|-----------|
| Mnemonic | PN-2S-FI |
| Test Parameters | |
| Specified α | 0.05 |
| Power | 0.898 |
| Test Type | 2-Sided |
| Model Parameters | |
| Proportion under Control (π_c) | 0.05 |
| Proportion under Treatment (π_t) | 0.25 |
| Allocation Ratio (n_t/n_c) | 1 |
| Proportion Treatment Input | USER_SPEC |
| Sample Size | |
| Maximum | 136 |

Suppose that this sample size is larger than economically feasible and it is desired to evaluate the power when a total of 100 subjects are enrolled. Create a new design by selecting Des1 in the **Library**, and clicking the  icon. In the input, select the radio button in the box next to **Power**. The box next to **Power** will now say Computed, since we wish to compute power. In the box next to **Sample Size (n)** enter 100.



Test Parameters

Test Type: 2-Sided

Type I Error (α): 0.05

Power: **Computed**

Sample Size (n): 100

Allocation Ratio: 1

Proportion Control (π_c): 0.05

Proportion of Treatment

User Specified

Odds Ratio

Difference π_t : 0.25

Ratio

Click **Compute** to create design Des2. The results of Des2 are shown in the **Output Preview** window. With Des2 selected in the **Output Preview**, click the  icon. In the **Library**, select the rows for both Des1 and Des2, by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the two designs

side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 |
|--|-----------|-----------|
| Mnemonic | PN-2S-FI | PN-2S-FI |
| Test Parameters | | |
| Specified α | 0.05 | 0.05 |
| Power | 0.898 | 0.748 |
| Test Type | 2-Sided | 2-Sided |
| Model Parameters | | |
| Proportion under Control (π_c) | 0.05 | 0.05 |
| Proportion under Treatment (π_t) | 0.25 | 0.25 |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Proportion Treatment Input | USER_SPEC | USER_SPEC |
| Sample Size | | |
| Maximum | 136 | 100 |

Des2 yields a power of approximately 75% as shown. Noting that 100 subjects is economically feasible and yields reasonable power, the question arises as to the sample size required to have 80%, which might still be economically feasible. This can be accomplished by selecting Des1 in the **Library**, and clicking the  icon. In the input, change the **Power** from 0.9 to 0.8. Click **Compute** to generate the output for Des3. The results of Des3 are shown in the **Output Preview** window. With Des3 selected in the **Output Preview**, click the  icon. In the **Library**, select the rows for both Des1, Des2, and Des3 by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the three designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 | Wbk1:Des3 |
|--|-----------|-----------|-----------|
| Mnemonic | PN-2S-FI | PN-2S-FI | PN-2S-FI |
| Test Parameters | | | |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.898 | 0.748 | 0.801 |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Model Parameters | | | |
| Proportion under Control (π_c) | 0.05 | 0.05 | 0.05 |
| Proportion under Treatment (π_t) | 0.25 | 0.25 | 0.25 |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Proportion Treatment Input | USER_SPEC | USER_SPEC | USER_SPEC |
| Sample Size | | | |
| Maximum | 136 | 100 | 110 |

23 Binomial Superiority Two-Sample

Entering 0.8 for the power yields a required sample size of 110 subjects.

23.6 Assurance (Probability of Success)

Assurance, or probability of success, is a Bayesian version of power, which corresponds to the (unconditional) probability that the trial will yield a statistically significant result. Specifically, it is the prior expectation of the power, averaged over a prior distribution for the unknown treatment effect (see O'Hagan et al., 2005). For a given design, East allows you to specify a prior distribution, for which the assurance or probability of success will be computed. First, enter the following values in the Input window: A 3-look design for testing the difference in proportions of two distinct populations with Lan-DeMets(OF) efficacy only boundary, Superiority Trial, 1-sided test, 0.025 type-1 error, 80% power, $\pi_c = 0.15$, and $\pi_t = 0.1$.

Select the **Assurance** checkbox in the Input window. The following options will appear as below.

To address our uncertainty about the treatment proportion, we specify a prior distribution for π_t . In the **Distribution** list, click **Beta**, and in the **Input Method** list, click **Beta Parameters (a and b)**. Enter the values of $a = 11$ and $b = 91$. Recall that the mode of the Beta distribution is $\frac{a-1}{a+b-2}$. Thus, these parameter values generate a Beta distribution that is peaked at 0.1, which matches the assumed treatment

proportion. Click **Compute**.

Assurance (Probability of Success) 0.597

Prior Distribution for: n_t Distribution: Beta

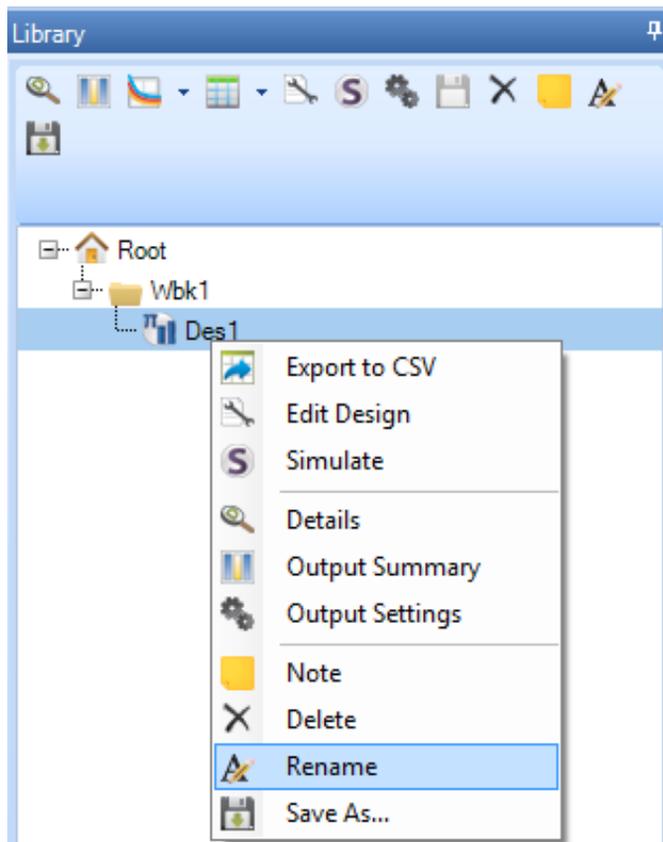
Input Method: Beta Parameters (a and b)

Parameters of Distribution of n_t

a: 11 b: 91

The computed probability of success (0.597) is shown above. Note that for this prior, assurance is very less than the specified power (0.8); incorporating the uncertainty about π_t has yielded a much less optimistic estimate of power. Save this design in the **Library** and rename it as Bayes1.

23 Binomial Superiority Two-Sample



East also allows you to specify an arbitrary prior distribution through a CSV file. In the **Distribution** list, click **User Specified**, and then click **Browse...** to select the CSV file where you have constructed a prior.

Assurance (Probability of Success) 0.735
 Prior Distribution for: n_t Distribution: User Specified
D:\Work\EAST6.3\Assurance_Prior.c Browse...

If you are specifying a prior for one parameter only (either π_c or π_t , but not both), then the CSV file should contain two columns, where the first column lists the grid points

for the parameter of interest, and the second column lists the prior probability assigned to each grid point. If you are specifying priors for both π_c and π_t , the CSV file should contain four columns (from left to right): values of π_c , probabilities for π_c , values of π_t , and probabilities for π_t . The number of points for π_c and number of points for π_t may differ. For example, we consider a 5-point prior for π_t only, with probability = 0.2 at each point.

| | A | B |
|---|------|-----|
| 1 | 0.08 | 0.2 |
| 2 | 0.09 | 0.2 |
| 3 | 0.1 | 0.2 |
| 4 | 0.11 | 0.2 |
| 5 | 0.12 | 0.2 |

Once the CSV filename and path has been specified, click **Compute** to calculate the assurance, which will be displayed in the box below:

Assurance (Probability of Success): 0.735

Prior Distribution for: n_t Distribution: User Specified

D:\Work\East62Manual\Assurance.c Browse...

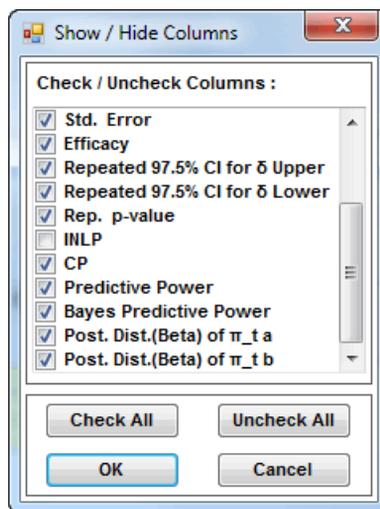
As stated in O’Hagan et al. (2005, p.190): “Assurance is a key input to decision-making during drug development and provides a reality check on other methods of trial design.” Indeed, it is not uncommon for assurance to be much lower than the specified power. The interested reader is encouraged to refer to O’Hagan et al. for further applications and discussions on this important concept.

23.7 Predictive Power and Bayesian Predictive Power

Similar Bayesian ideas can be applied to conditional power for interim monitoring. Rather than calculating conditional power for a single assumed value of the treatment effect, δ , such as at $\hat{\delta}$, we may account for the uncertainty about δ by taking a weighted average of conditional powers, weighted by the posterior distribution for δ . East calculates an average power, called the **predictive power** (Lan, Hu, & Proschan, 2009), assuming a diffuse prior for the drift parameter, η . In addition, if the user specified a beta prior distribution at the design stage to calculate assurance, then East will also calculate the average power, called **Bayesian predictive power**, for the corresponding posterior. We will demonstrate these calculations for the design renamed as Bayes1 earlier.

23 Binomial Superiority Two-Sample

In the **Library**, right-click Bayes1 and click **Interim Monitoring**, then click  in the toolbar of the IM Dashboard.

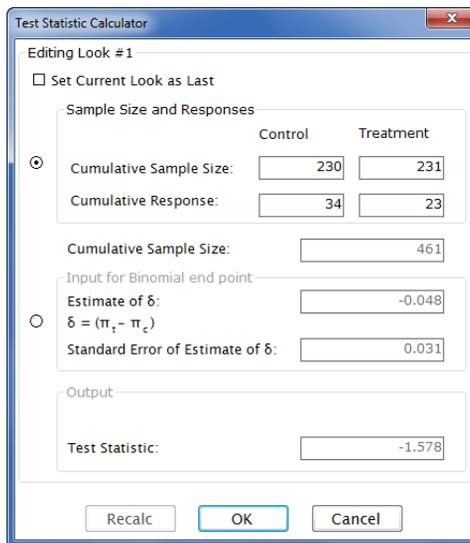


In the Show/Hide Columns window, make sure to show the columns for: CP (Conditional Power), Predictive Power, Bayes Predictive Power, Posterior Distribution of $\pi_t a$, and Posterior Distribution of $\pi_t b$, and click **OK**. The following columns will be added to the main grid of the IM Dashboard.

| CP | Predictive Power | Bayes Predictive Power | Post. Dist.(Beta) of π_t | |
|----|------------------|------------------------|------------------------------|---|
| | | | a | b |
| | | | | |
| | | | | |

In the toolbar of the IM Dashboard, open the Test Statistic Calculator by clicking **Enter Interim Data**. In order to appropriately update the posterior distribution, you will need to use the Test Statistic Calculator to enter the sample size and number of responses for each arm. Enter 34 events out of 230 patients in the control arm, and 23

out of 231 patients in the treatment arm, then click **OK**.



The main grid of the IM Dashboard will be updated as follows. In particular, notice the differing values for CP and the Bayesian measures of power.

| CP | Predictive Power | Bayes Predictive Power | Post. Dist.(Beta) of π_t | |
|-------|------------------|------------------------|------------------------------|-----|
| | | | a | b |
| 0.823 | 0.705 | 0.74 | 34 | 299 |
| | | | | |

24 *Binomial Non-Inferiority Two-Sample*

In a binomial non-inferiority trial the goal is to establish that the response rate of an experimental treatment is **no worse than** that of an active control, rather than attempting to establish that it is superior. A therapy that is demonstrated to be non-inferior to the current standard therapy for a particular indication might be an acceptable alternative if, for instance, it is easier to administer, cheaper, or less toxic. Non-inferiority trials are designed by specifying a non-inferiority margin. The amount by which the response rate on the experimental arm is worse than the response rate on the control arm must fall within this margin in order for the claim of non-inferiority to be sustained. In this chapter, we shall design and monitor non-inferiority trials in which the non-inferiority margin is expressed as either a difference, a ratio, or an odds ratio of two binomial proportions. The difference is examined in Section 24.1. This is followed by two formulations for the ratio: the Wald formulation in Section 24.2 and the Farrington-Manning formulation in Section 24.3. The odds ratio formulation is presented in Section 24.4.

24.1 *Difference of Proportions*

24.1.1 *Trial Design*

24.1.2 *Trial Simulation*

24.1.3 *Interim Monitoring*

Let π_c and π_t denote the response rates for the control and experimental treatments, respectively. Let $\delta = \pi_t - \pi_c$. The null hypothesis is specified as

$$H_0: \delta = \delta_0$$

and is tested against one-sided alternative hypotheses. If the occurrence of a response denotes patient benefit rather than harm, then $\delta_0 > 0$ and the alternative hypothesis is

$$H_1: \delta < \delta_0$$

or equivalently as

$$H_1: \pi_t > \pi_c - \delta_0 .$$

Conversely, if the occurrence of a response denotes patient harm rather than benefit, then $\delta_0 < 0$ and the alternative hypothesis is

$$H_1: \delta > \delta_0$$

or equivalently as

$$H_1: \pi_t < \pi_c - \delta_0 .$$

For any given π_c , the sample size is determined by the desired power at a specified value of $\delta = \delta_1$. A common choice is $\delta_1 = 0$ (or equivalently $\pi_t = \pi_c$) but East permits you to power the study at any value of δ_1 which is consistent with the choice of H_1 .

Let $\hat{\pi}_{tj}$ and $\hat{\pi}_{cj}$ denote the estimates of π_t and π_c based on n_{tj} and n_{cj} observations from the experimental and control treatments, respectively, up to and including j -th look, $j = 1, \dots, K$, where a maximum of K looks are to be made. The test statistic at the j -th look is

$$Z_j = \frac{\hat{\delta}_j - \delta_0}{\text{se}(\hat{\delta}_j)} \tag{24.1}$$

where

$$\hat{\delta}_j = \hat{\pi}_{cj} - \hat{\pi}_{tj} \tag{24.2}$$

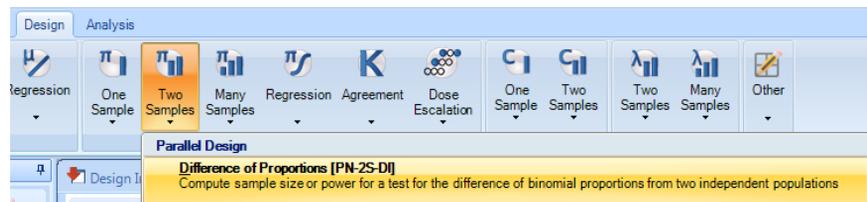
and

$$\text{se}(\hat{\delta}_j) = \sqrt{\frac{\hat{\pi}_{cj}(1 - \hat{\pi}_{cj})}{n_{cj}} + \frac{\hat{\pi}_{tj}(1 - \hat{\pi}_{tj})}{n_{tj}}} \tag{24.3}$$

24.1.1 Trial Design

The 24-week disease-free rate with a standard therapy for HIV is 80%. Suppose that the claim of non-inferiority for an experimental therapy can be sustained if its response rate is greater than 75%; i.e., the non-inferiority margin is $\delta_0 = 0.05$. For studies of this type, we specify inferiority as the null hypothesis, non-inferiority as the alternative hypothesis, and attempt to reject the null hypothesis using a one-sided test. We will specify to East that, under the null hypothesis H_0 , $\pi_c = 0.8$ and $\pi_t = 0.75$. We will test this hypothesis with a one-sided level 0.05 test. Suppose we require 90% power at the alternative hypothesis, H_1 , that both response rates are equal to the null response rate of the control arm, i.e. $\delta_1 = 0$. Thus, under H_1 , we have $\pi_c = \pi_t = 0.8$.

To begin click **Two Samples** on the **Design** tab in the **Discrete** group, and then click **Difference of Proportions**. inxxnon-inferiority,binomial



Single-Look Design Powered at $\delta = 0$ To begin with, suppose we will design a single-look study for rejection of H_0 only, with 90% power at a 0.025 significance level. Enter the relevant parameters into the dialog box as shown below. In the drop

24 Binomial Non-Inferiority Two-Sample

down box next to **Trial** be sure to select Noninferiority.

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

Design Type: Noninferiority Number of Looks: 1

Test Parameters

Test Type: 1-Sided

Type I Error (α): 0.025

Power: 0.9

Sample Size (n): Computed

Allocation Ratio: 1 (n_1/n_2)

Specify Proportion Response
Prop. under Control (π_c): 0.8

Specify Null Hypothesis
Prop. under Treatment (π_{10}): 0.75
Noninferiority Margin (δ_0): -0.05 ($\delta_0 = \pi_{10} - \pi_c$)

Specify Alternative Hypothesis
Prop. under Treatment (π_{11}): 0.8
Diff. in Prop. ($\delta_1 = \pi_{11} - \pi_c$): 0

Perform Exact Computations

Use Casagrande-Pike-Smith Correction (Ignored if alloc. ratio is not 1)

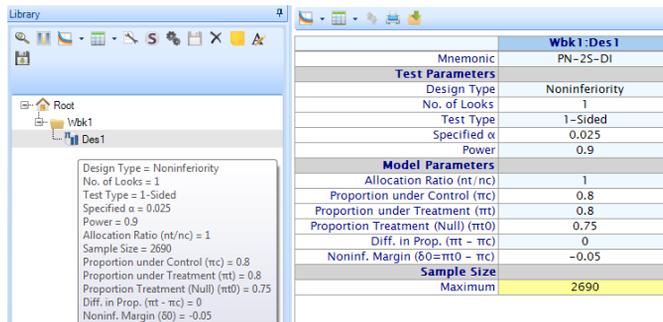
Assurance (Probability of Success)

Now click **Compute**. The design is shown as a row in the Output Preview located in the lower pane of this window. The single-look design requires a combined total of 2690 patients on both arms in order to attain 90% power. We can, however, reduce the expected sample size without any loss of power if we use a group sequential design. This is considered next.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | ttc | Prop. Treatment (Alt.) | Prop. Treatment (Null) | δ_1 | δ_0 |
|------|----------------|--------------|-----------|--------------------|-------|-------|-------------|-----|------------------------|------------------------|------------|------------|
| Des1 | Noninferiority | 1 | 1-Sided | 0.025 | 0.9 | 1 | 2690 | 0.8 | 0.8 | 0.75 | 0 | -0.05 |

Before continuing we will save Design1 to the **Library**. You can select this design by clicking anywhere along the row in the **Output Preview**. Some of the design details will be displayed in the upper pane, labeled **Compare Designs**. In the **Output Preview** toolbar, click the  icon to save this design to Workbook1 in the **Library**. If you hover the cursor over Design1 in the Library, a tooltip will appear that

summarizes the input parameters of the design.



Three-Look Design Powered at $\delta = 0$ For the above study, suppose we wish to take up to two interim looks and one final look at the accruing data. Create a new design by selecting Design1 in the **Library**, and clicking the  icon on the **Library** toolbar. Change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. A new tab **Boundary Info** will appear. Clicking on this tab will reveal the stopping boundary parameters. By default, the **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The left side contains details for the **Efficacy** boundary, and the right side contains details for the **Futility** boundary. By default, there is an efficacy boundary (to reject H0) selected, but no futility boundary (to reject H1). The **Boundary Family** specified is of the Spending Functions type. The default **Spending function** is the Lan-DeMets (Lan & DeMets, 1983), with **Parameter OF** (O’Brien-Fleming), which generates boundaries that are very similar, though not identical, to the classical stopping boundaries of O’Brien and Fleming (1979).

Now suppose, in our example, that the three looks are unequally spaced, with the first look being taken after 50% of the committed accrual, and the second look being taken when after 75% of the committed accrual. Under **Spacing of Looks** in the **Boundary Info** tab, click the **Unequal** radio button. The column titled **Info. Fraction** in the **Look Details** table can be edited to modify the relative spacing of the analyses. The information fraction refers to the proportion of the maximum (yet unknown) sample size. By default, this table displays equal spacing. Enter the new information fraction values as shown below and click **Recalc** to see the updated values of the stopping

24 Binomial Non-Inferiority Two-Sample

boundaries populated in the **Look Details** table.

Design Type: Number of Looks:

Test Parameters **Boundary**

Efficacy
 Boundary Family:
 Spending Function:
 Parameter:
 Type I Error (α): 0.025

Futility
 Boundary Family:

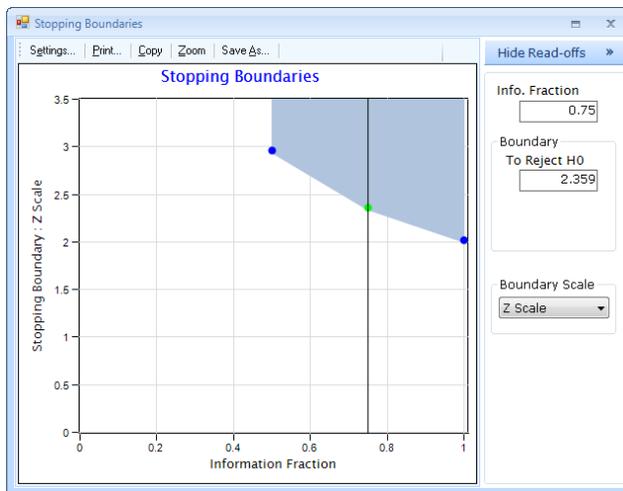
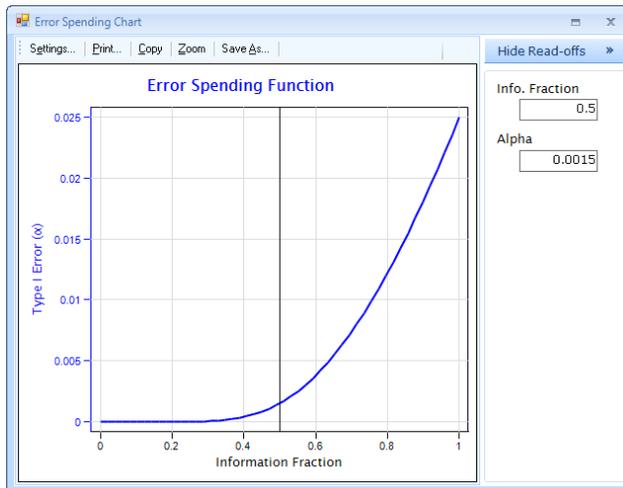
Spacing of Looks Equal Unequal

Efficacy Boundary:  

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.5 | 0.002 | 2.963 |
| 2 | 0.75 | 0.010 | 2.359 |
| 3 | 1 | 0.025 | 2.014 |

On the **Boundary Info** tab, you may also click the  or  icons to view plots

of the error spending functions, or stopping boundaries, respectively.



Click the **Compute** button to generate output for Design2. With Design2 selected in the **Output Preview**, click the  icon to save Design2 to the **Library**. In the **Library**, select the rows for Design1 and Design2, by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the two designs

24 Binomial Non-Inferiority Two-Sample

side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 |
|--|----------------|----------------|
| Mnemonic | PN-2S-DI | PN-2S-DI |
| Test Parameters | | |
| Design Type | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Proportion under Control (π_c) | 0.8 | 0.8 |
| Proportion under Treatment (π_t) | 0.8 | 0.8 |
| Proportion Treatment (Null) (π_{t0}) | 0.75 | 0.75 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0 | 0 |
| Noninf. Margin ($\delta_0 = \pi_{t0} - \pi_c$) | -0.05 | -0.05 |
| Boundary Parameters | | |
| Spacing of Looks | | Unequal |
| Efficacy Boundary | | LD (OF) |
| Sample Size | | |
| Maximum | 2690 | 2740 |
| Expected Under H_0 | | 2732.345 |
| Expected Under H_1 | | 2093.682 |

Let us examine the design output from Design2. The maximum number of subjects that we must commit to this study in order to achieve 90% power is 2740. That is 50 patients more than are needed for Design1. However, since Design1 is a single-look design, there is no prospect of saving resources if indeed H_1 is true and the two treatments have the same response rates. In contrast, Design2 permits the trial to stop early if the test statistic crosses the stopping boundary. For this reason, the expected sample size under H_1 is 2094, a saving of 596 patients relative to Design1. If H_0 is true, the expected sample size is 2732 and there is no saving of patient resources. In order to see the stopping probabilities, as well as other characteristics, select Design2 in the **Library**, and click the  icon. The cumulative boundary stopping

probabilities are shown in the **Stopping Boundaries** table.

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Test Parameters | |
|------------------------------------|----------------|
| Design ID | Des2 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Prop. under Control (π_c) | 0.8 |
| Prop. under Treatment (π_t) | 0.8 |
| Prop. Treatment (Null) (π_0) | 0.75 |
| $\delta = \pi_t - \pi_c$ | |
| Under H0 | -0.05 |
| Under H1 | 0 |
| Allocation Ratio (n_t/n_c) | 1 |
| Casagrande-Pike-Smith Correction | Not Applied |
| Boundary Parameters | |
| Spacing of Looks | Unequal |
| Efficacy Boundary | LD (OF) |

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|-----------------------------------|-------------|---------------|----------|
| Sample Size (n) | | | |
| Maximum | 1370 | 1370 | 2740 |
| Expected H1 | 1047.055 | 1046.627 | 2093.682 |
| Expected H0 | 1366.177 | 1366.169 | 2732.345 |
| Maximum Information (I): 4.281.25 | | | |

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|------------|---|----------|
| | | | | | Under H0 | Under H1 |
| | | | | Efficacy Z | Efficacy | Efficacy |
| 1 | 0.5 | 1370 | 0.002 | 2.963 | 0.002 | 0.258 |
| 2 | 0.75 | 2055 | 0.01 | 2.359 | 0.008 | 0.427 |
| 3 | 1 | 2740 | 0.025 | 2.014 | 0.015 | 0.215 |

To display a chart of average sample number (ASN) versus the effect size, $\pi_t - \pi_c$, select Design2 in the **Library** and click on the  icon and select **Average Sample Number (ASN)**. To display a chart of power versus treatment size, select Design2 in the **Library** and click on the  icon and select **Power vs. Treatment Effect (δ)**.

In Design2, we utilized Lan-DeMets (Lan & DeMets, 1983) spending function, with **Parameter** OF (O'Brien-Fleming) to generate the stopping boundary for early stopping under H_1 . One drawback of Design2 is the large expected sample size if H_0 is true. We can guard against this eventuality by introducing a futility boundary which will allow us to stop early if H_0 is true. A popular approach to stopping early for futility is to compute the conditional power at each interim monitoring time point and stop the study if this quantity is too low. This approach is somewhat arbitrary since there is no guidance as to what constitutes low conditional power. In East, we compute futility boundaries that protect β , the type-2 error, so that the power of the study will not deteriorate. This is achieved by using a β -spending function to generate the futility boundary. Thereby the type-2 error will not exceed β and the power of the study will be preserved. This approach was published by Pampallona and Tsiatis (1994).

Suppose we now wish to include a futility boundary. To design this trial select Design2 in the **Library** and click the  icon. In the **Boundary Info** tab, in the **Futility** box, set **Boundary Family** to Spending Function. Change the **Spending Function** to Gamma Family and change the **Parameter** (Γ) to -8 . This family is parameterized by the single parameter γ which can take all possible non-zero values.

24 Binomial Non-Inferiority Two-Sample

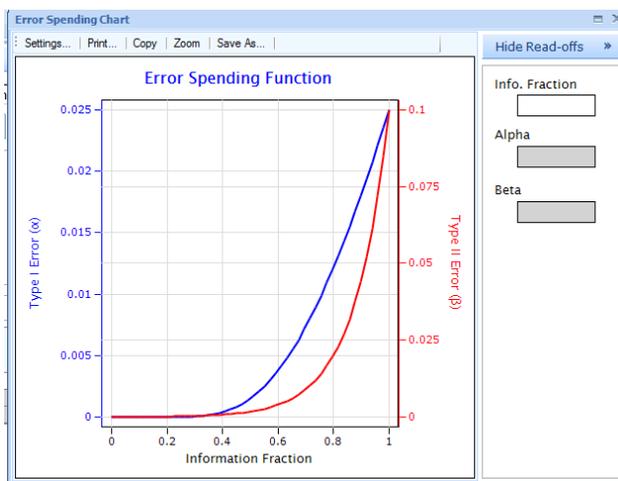
Its functional form is

$$\beta(t) = \frac{\beta(1 - e^{-\gamma t})}{(1 - e^{-\gamma})}. \tag{24.4}$$

Next click **Refresh Boundary**. Your screen should now look like the following:

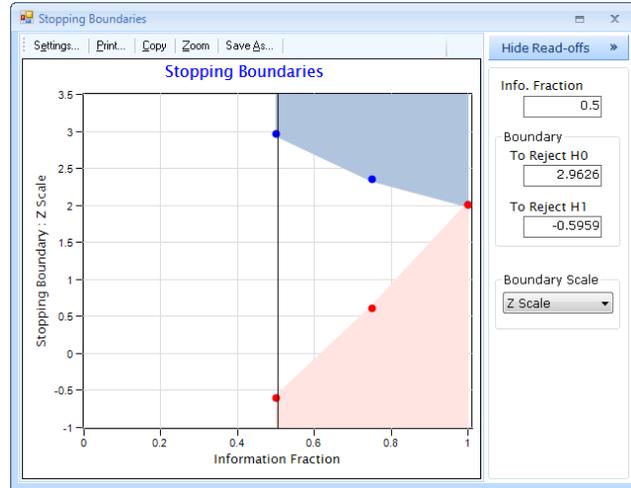
| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Cum. β Spent | Futility Boundary |
|--------|----------------|-------------------------------------|-------------------------------------|--------------|-------------------|--------------|-------------------|
| 1 | 0.500 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.002 | 2.963 | 0.002 | -0.596 |
| 2 | 0.750 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.010 | 2.359 | 0.014 | 0.611 |
| 3 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.025 | 2.014 | 0.100 | 2.014 |

On the **Boundary Info** tab, you may also click the or icons to view plots of the error spending functions, or stopping boundaries, respectively.

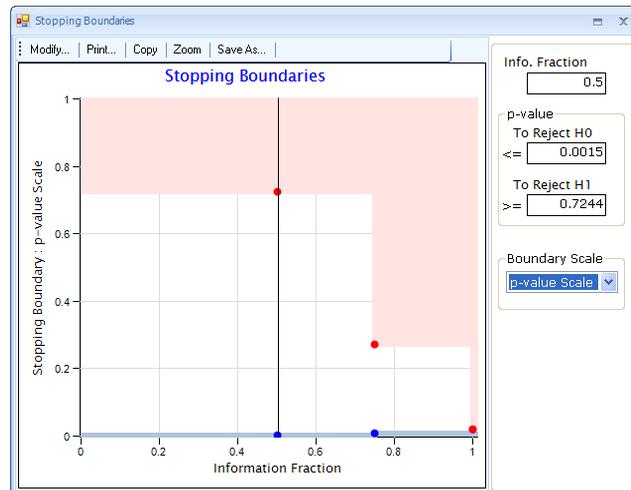


Notice how conservative the β -spending function is compared to the α -spending

function. Its rate of error spending is almost negligible until about 60% of the information has accrued.



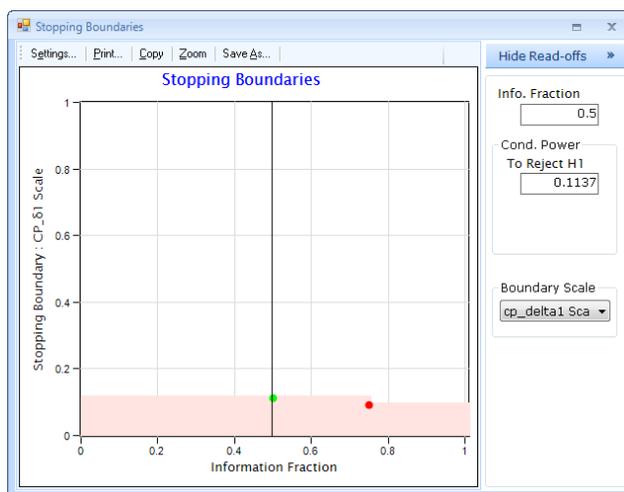
One can view the stopping boundaries on various alternative scales by selecting the appropriate scale from the drop-down list of boundary scales to the right of the chart. It is instructive to view the stopping boundaries on the p-value scale.



24 Binomial Non-Inferiority Two-Sample

By moving the vertical scroll bar from left to right in the above chart, one can observe the p-values required for early stopping at each look. The p-values needed to stop the study and declare non-inferiority at the first, second and third looks are, respectively, 0.0015, 0.0092 and 0.022. The p-values needed to stop the study for futility at the first and second looks are, respectively, 0.7244 and 0.2708.

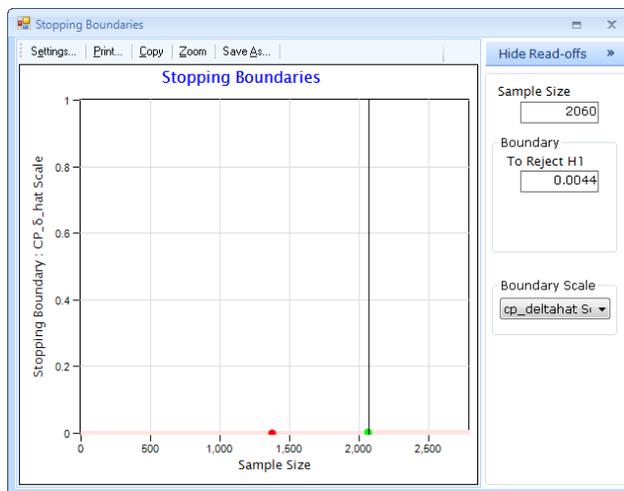
Other useful scales for displaying the futility boundary are the conditional power scales. They are the `cp_delta1` Scale and the `cp_delta1hat` scale. Here ‘cp’ refers to conditional power. The suffix ‘delta1’ implies that we will represent the futility boundary in terms of conditional power evaluated at the value of $\delta = \delta_1$ specified at the design stage under the alternative hypothesis. The suffix ‘delta1hat’ implies that we will represent the futility boundary in terms of conditional power evaluated at the value of $\hat{\delta}$ at which the test statistic $Z = \hat{\delta}/se(\hat{\delta})$ would just hit the futility boundary. The screenshot below represents the first two values of the futility boundary on the `cp_delta1` Scale.



For example, the stopping boundary at the first look is `cp_delta1=0.1137`. This is to be interpreted in the following way: if at the first look the value of the test statistic Z just falls on the futility boundary, then the conditional power, as defined by Section C.3 of Appendix C with $\delta = \delta_1 = 0$, will be 0.1137. This gives us a way to express the futility boundary in terms of conditional power.

The `cp_delta1` Scale might not give one an accurate picture of futility. This is because, on this scale, the conditional power is evaluated at the value of $\delta = \delta_1$

specified at the design stage. However, if the test statistic has actually fallen on the futility boundary, the data are more suggestive of the null than the alternative hypothesis and it is not very likely that $\delta = \delta_1$. Thus it might be more reasonable to evaluate conditional power at the observed value $\delta = \hat{\delta}$. The screenshot below represents the futility boundary on the `cp_deltahat` Scale.



For example, the stopping boundary at the second look is `cp_deltahat=0.0044`. This is to be interpreted in the following way: if at the second look, the value of test statistic Z just falls on the futility boundary, then the conditional power, as defined by Section C.3 of Appendix C with $\delta = \hat{\delta} = Z \times se(\hat{\delta})$, will be 0.0044. It is important to realize that the futility boundary has not changed. It is merely being expressed on a different scale. On the whole, it is probably more realistic to express the futility boundary on the `cp_deltahat` scale than on the `cp_delta1` scale since it is highly unlikely that the true value of δ is equal to δ_1 if Z has hit the futility boundary.

Close this chart before continuing. Click the **Compute** button to generate output for Design3. With Design3 selected in the **Output Preview**, click the  icon. In the **Library**, select the rows for Design1, Design2, and Design3, by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the three

24 Binomial Non-Inferiority Two-Sample

designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 | Wbk1:Des3 |
|--|----------------|----------------|----------------|
| Mnemonic | PN-2S-DI | PN-2S-DI | PN-2S-DI |
| Test Parameters | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 | 3 |
| Test Type | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 | 0.025 |
| Attained α | | | 0.025 |
| Power | 0.9 | 0.9 | 0.9 |
| Model Parameters | | | |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Proportion under Control (π_c) | 0.8 | 0.8 | 0.8 |
| Proportion under Treatment (π_t) | 0.8 | 0.8 | 0.8 |
| Proportion Treatment (Null) (π_{t0}) | 0.75 | 0.75 | 0.75 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0 | 0 | 0 |
| Noninf. Margin ($\delta_0 = \pi_{t0} - \pi_c$) | -0.05 | -0.05 | -0.05 |
| Boundary Parameters | | | |
| Spacing of Looks | | Unequal | Unequal |
| Efficacy Boundary | | LD (OF) | LD (OF) |
| Futility Boundary | | | Gm (-8) (NB) |
| Sample Size | | | |
| Maximum | 2690 | 2740 | 2746 |
| Expected Under H0 | | 2732.345 | 2046.973 |
| Expected Under H1 | | 2093.682 | 2086.492 |

Observe that Design3 will stop with a smaller expected sample size under either H_0 or H_1 compared to Design2.

Three-Look Design Powered at $\delta \neq 0$ The previous designs were all powered to detect the alternative hypothesis that the new treatment and the active control have the same response rate ($\delta_1 = 0$). As is usually the case with non-inferiority trials, the distance between the non-inferiority margin $\delta_0 = 0.05$ and the alternative hypothesis $\delta_1 = 0$ is rather small, thereby resulting in a very large sample size commitment to this trial. Sometimes a new treatment is actually believed to have a superior response rate to the active control. However the anticipated treatment benefit might be too small to make it feasible to run a superiority trial. Suppose, for example, that it is anticipated that the treatment arm could improve upon the 80% response rate of the active control by about 2.5%. A single-look superiority trial designed for 90% power to detect this small of a difference would require over 12000 subjects. In this situation, the sponsor might prefer to settle for a non-inferiority claim. A non-inferiority trial in which the active control has a response probability of $\pi_c = 0.8$, the non-inferiority margin is $\delta_0 = -0.05$, and the alternative hypothesis is $\delta_1 = \pi_c - \pi_t = -0.025$ can be designed as follows.

Create a new design by selecting Design3 in the **Library**, and clicking the  icon on the **Library** toolbar. In the ensuing dialog box, choose the design parameters as

shown below.

Design Type: Number of Looks:

Test Parameters **Boundary**

Test Type:

Type I Error (α):

Power:

Sample Size (n):

Allocation Ratio: (n_1/n_2)

Specify Proportion Response
Prop. under Control (π_c):

Specify Null Hypothesis
Prop. under Treatment (π_{t0}):
Noninferiority Margin (δ_0):
 $(\delta_0 = \pi_{t0} - \pi_c)$

Specify Alternative Hypothesis
Prop. under Treatment (π_{t1}):
Diff. in Prop. ($\delta_1 = \pi_{t1} - \pi_c$):

Use Casagrande-Pike-Smith
 Correction (ignored if alloc. ratio is not 1)

Assurance (Probability of Success):

Click the **Compute** button to generate output for Design4. Notice that this design requires only 1161 subjects. This is 1585 fewer subjects than under Design3.

| | Wbk1-Des1 | Wbk1-Des2 | Wbk1-Des3 | Wbk1-Des4 |
|--|----------------|----------------|----------------|----------------|
| Mnemonic | PN-25-DI | PN-25-DI | PN-25-DI | PN-25-DI |
| Test Parameters | | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 | 3 | 3 |
| Test Type | 1-Sided | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 | 0.025 | 0.025 |
| Attained α | | | 0.025 | 0.025 |
| Power | 0.9 | 0.9 | 0.9 | 0.9 |
| Model Parameters | | | | |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 | 1 |
| Proportion under Control (ttc) | 0.8 | 0.8 | 0.8 | 0.8 |
| Proportion under Treatment (ttt) | 0.8 | 0.8 | 0.8 | 0.825 |
| Proportion Treatment (Null) (ttt0) | 0.75 | 0.75 | 0.75 | 0.75 |
| Diff. in Prop. (tt - ttc) | 0 | 0 | 0 | 0.025 |
| Noninf. Margin ($\delta_0 = \text{ttt0} - \text{ttc}$) | -0.05 | -0.05 | -0.05 | -0.05 |
| Boundary Parameters | | | | |
| Spacing of Looks | | Unequal | Unequal | Unequal |
| Efficacy Boundary | | LD (OF) | LD (OF) | LD (OF) |
| Futility Boundary | | | Gm (-8) (NB) | Gm (-8) (NB) |
| Sample Size | | | | |
| Maximum | 2690 | 2740 | 2746 | 1161 |
| Expected Under H0 | | 2732.345 | 2046.973 | 865.381 |
| Expected Under H1 | | 2093.682 | 2086.492 | 882.038 |

24.1.2 Trial Simulation

You can simulate Design 3 by selecting Design3 in the **Library**, and clicking the  icon from **Library** toolbar. Alternatively, right-click on Design3 and select **Simulate**.

24 Binomial Non-Inferiority Two-Sample

A new Simulation worksheet will appear.

| Look # | Info. Fraction | Cum. α Spent | Efficacy Z | Futility Z |
|--------|----------------|---------------------|------------|------------|
| 1 | 0.500 | 0.002 | 2.963 | -0.596 |
| 2 | 0.750 | 0.010 | 2.359 | 0.612 |
| 3 | 1.000 | 0.025 | 2.014 | 2.014 |

Try different choices for the simulation parameters to verify the operating characteristics of the study. For instance under the **Response Generation Info** tab, set **Prop. Under Control** to 0.8 and **Prop. Under Treatment** to 0.75. You will be simulating under the null hypothesis and should achieve a rejection rate of 2.5%. Now, click on the **Simulate** button.

Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled Simulation1. Select Simulation1 in the **Output Preview**. Note that some of the design details will be displayed in the upper pane, labeled **Compare Designs**. Click the  icon to save it to the **Library**. Double-click on Simulation1

in the **Library**. The simulation output details will be displayed.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Test Parameters | |
|-----------------------------------|----------------|
| Simulation ID | Sim1 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 2746 |
| Noninf. Margin (π_c) | -0.05 |
| Avg. Power at Termination | 0.896 |
| Response Generation Parameters | |
| Prop. under Control (π_c) | 0.8 |
| Prop. under Treatment (π_t) | 0.8 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

⊖ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 1373 |
| 2 | 2060 |
| 3 | 2746 |
| Average | 2079.764 |

⊖ Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|----------|--------------|----------|-------------------|---------|
| | | Efficacy | Futility | Efficacy | Futility | Count | % |
| 1 | 1373 | 2.963 | -0.596 | 2656 | 16 | 2672 | 26.720% |
| 2 | 2060 | 2.359 | 0.612 | 4238 | 126 | 4364 | 43.640% |
| 3 | 2746 | 2.014 | 2.014 | 2068 | 896 | 2964 | 29.640% |
| Total | | | | 8962 | 1038 | 10000 | |
| % | | | | 89.620% | 10.380% | | |

Simulation Seed and Elapsed Time

Starting Seed: 747963
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:05

We see above that we achieved a rejection rate of 2.5%.

Now suppose that the new treatment is actually slightly superior to the control treatment. For example, $\pi_c = 0.8$ and $\pi_t = 0.81$. Since this study is designed for 90% power when $\pi_c = \pi_t = 0.8$, we would expect the simulations to reveal power in excess of 90%.

Select Sim1 node in the **Library**, and click the  icon from **Library** toolbar. Under the **Response Generation Info** tab change the **Prop. Under Treatment** to 0.81. Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled Simulation2. Select Simulation2 in the **Output Preview**. Click the  icon to save it to the **Library**. Double-click on Simulation2 in the **Library**. The simulation output details will be

24 Binomial Non-Inferiority Two-Sample

displayed.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Test Parameters | |
|-----------------------------------|----------------|
| Simulation ID | Sim2 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 2746 |
| Noninf. Margin (π_{ic}) | -0.05 |
| Avg. Power at Termination | 0.974 |
| Response Generation Parameters | |
| Prop. under Control (π_c) | 0.8 |
| Prop. under Treatment (π_t) | 0.81 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 1373 |
| 2 | 2060 |
| 3 | 2746 |
| Average | 1851.222 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|----------|--------------|----------|-------------------|---------|
| | | Efficacy | Futility | Efficacy | Futility | Count | % |
| 1 | 1373 | 2.963 | -0.596 | 4398 | 3 | 4401 | 44.010% |
| 2 | 2060 | 2.359 | 0.612 | 4209 | 26 | 4235 | 42.350% |
| 3 | 2746 | 2.014 | 2.014 | 1135 | 229 | 1364 | 13.640% |
| Total | | | | 9742 | 258 | 10000 | |
| % | | | | 97.420% | 2.580% | | |

Simulation Seed and Elapsed Time

Starting Seed: 1087185
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:04

These results show that the power exceeds 97%.

The power of the study will deteriorate if the response rate of the control arm is less than 0.8, even if $\pi_c = \pi_t$. To see this, let us simulate with $\pi_c = \pi_t = 0.7$. The results

are shown below.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Simulation Parameters | |
|-----------------------------------|----------------|
| Simulation ID | Sim3 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 2746 |
| Noninf. Margin (π_{10}) | -0.05 |
| Avg. Power at Termination | 0.813 |
| Response Generation Parameters | |
| Prop. under Control (π_c) | 0.7 |
| Prop. under Treatment (π_t) | 0.7 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

☉ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 1373 |
| 2 | 2060 |
| 3 | 2746 |
| Average | 2222.683 |

☉ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------------|------------|----------|--------------|----------|-------------------|---------|
| | | Efficacy | Futility | Efficacy | Futility | Count | % |
| 1 | 1373 | 2.963 | -0.596 | 1696 | 43 | 1739 | 17.390% |
| 2 | 2060 | 2.359 | 0.612 | 3864 | 284 | 4148 | 41.480% |
| 3 | 2746 | 2.014 | 2.014 | 2574 | 1539 | 4113 | 41.130% |
| Total | | | | 8134 | 1866 | 10000 | |
| % | | | | 81.340% | 18.660% | | |

☉ Simulation Seed and Elapsed Time

Starting Seed: 2679759
Total Number of Simulations: 10000
Elapsed Time: 00:00:02

Notice that the power has dropped from 90% to 80% even though the new treatment and the control treatment have the same response rates. This is because the lower response rates for π_c and π_t induce greater variability into the distribution of the test statistic. In order to preserve power, the sample size must be increased. This can be achieved without compromising the type-1 error within the group sequential framework by designing the study for a maximum amount of (Fisher) information instead of a maximum sample size. We discuss maximum information studies later, in Chaper 59.

24.1.3 Interim Monitoring

Consider interim monitoring of Design3. Select Design3 in the **Library**, and click the **IM** icon from the Library toolbar. Alternatively, right-click on Design3 and select **Interim Monitoring**. The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs. The bottom section contains four charts, each with a corresponding table to its right. These charts provide graphical and numerical descriptions of the progress of the clinical trial and are useful tools for decision making by a data monitoring committee.

Suppose that the trial is first monitored after accruing 500 subjects on each treatment arm, with 395 responses on the treatment arm and 400 responses on the control arm.

24 Binomial Non-Inferiority Two-Sample

Click on the **Enter Interim Data** icon to invoke the **Test Statistic Calculator**. In the box next to **Cumulative Sample Size** enter 1000. Enter -0.01 in the box next to **Estimate of δ** . In the box next to **Std. Error of δ** enter 0.02553. Next click **Recalc**.

The screenshot shows a window titled "Test Statistic Calculator". It has a "Recalc" button on the left and "OK" and "Cancel" buttons on the right. The main area is divided into sections: "Editing Look #1" with a checkbox "Set Current Look as Last"; "Sample Size and Responses" with radio buttons for "Control" and "Treatment" and input boxes for "Cumulative Sample Size" and "Cumulative Response" (all showing "N/A"); "Input for Binomial end point" with a radio button selected and input boxes for "Estimate of δ " (value: -0.01) and "Standard Error of Estimate of δ " (value: 0.02553); and "Output" with input boxes for "Estimate of $\delta - \delta_0$ " (value: 0.04) and "Test Statistic" (value: 1.567).

Note that the test statistic is computed to be 1.567.

Upon clicking the **OK** button, East will produce the interim monitoring report shown below.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | Futility | 87.5% RCI for δ | | Repeat... p-value | CP | Predicti... Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|----------|------------------------|-------|-------------------|-------|-------------------|
| | | | | | | | | Upper | Lower | | | |
| 1 | 0.364 | 1000 | 1.567 | -0.01 | 0.026 | 3.535 | -1.27 | 0.072 | -0.1 | 0.254 | 0.778 | 0.68 |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |

The stopping boundary for declaring non-inferiority is 3.535 whereas the value of the test statistic is only 1.567. Thus the trial should continue.

Suppose that the next interim look occurs after accruing 1250 patients on each arm with 1000 responses on the control arm and 990 responses on the treatment arm. Click on the second row in the table in the upper section. Then click the **Enter Interim Data** icon. The estimate of δ is -0.008 and the standard error is 0.016118. Enter the appropriate values as shown below and click **Recalc**.

Test Statistic Calculator

Editing Look #2

Set Current Look as Last

Sample Size and Responses

| | Control | Treatment |
|-------------------------|---------|-----------|
| Cumulative Sample Size: | N/A | N/A |
| Cumulative Response: | N/A | N/A |

Cumulative Sample Size: 2500

Input for Binomial end point

Estimate of δ : -0.008

$\delta = (\pi_t - \pi_c)$

Standard Error of Estimate of δ : 0.016118

Output

Estimate of $\delta - \delta_0$: 0.042

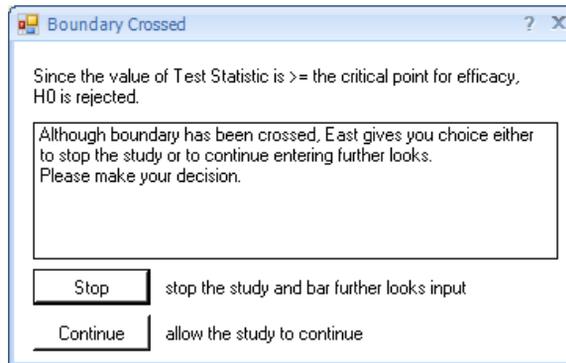
Test Statistic: 2.606

Recalc OK Cancel

Note that the value of the test statistic is now 2.606. Now click the **OK** button. This time the stopping boundary for declaring non-inferiority is crossed. The following

24 Binomial Non-Inferiority Two-Sample

message box appears.



Click the **Stop** button to stop the study. The analysis results are shown below.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | Futility | 87.5% RCI for δ | | Repeat ... p-value | CP | Predicti... Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|----------|------------------------|--------|--------------------|-------|-------------------|
| | | | | | | | | Upper | Lower | | | |
| 1 | 0.364 | 1000 | 1.567 | -0.01 | 0.026 | 3.535 | -1.27 | 0.072 | -0.1 | 0.254 | 0.778 | 0.68 |
| 2 | 0.91 | 2500 | 2.606 | -0.008 | 0.016 | 2.08 | 1.467 | 0.018 | -0.042 | 0.007 | NA | NA |

Click the "Edit Interim Data" button to edit the Look # 2 data.

Stopping Boundaries

| Sample Size | Efficacy | Futility |
|-------------|----------|----------|
| 1000 | 3.535 | -1.27 |
| 2500 | 2.08 | 1.467 |

Final Inference

| | |
|----------------------------|--------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.005 |
| Adj. Pt. Est. for δ | -0.008 |
| Adj. 95% CI for δ | |
| Upper Confidence Bound | 0.024 |
| Lower Confidence Bound | -0.04 |
| Post-Hoc Power | |

Error Spending Function

| Info. Fraction | α | β |
|----------------|----------|---------|
| 0.364 | 0 | 0.001 |
| 0.91 | 0.019 | 0.049 |

Confidence Intervals

| Info. Fraction | RCI Lower | RCI Upper | Naive CI Lower | Naive CI Upper | Adj. CI Upper |
|----------------|-----------|-----------|----------------|----------------|---------------|
| 0.364 | 0.072 | -0.1 | 1 | -0.06 | |
| 0.91 | 0.018 | -0.042 | 1 | -0.04 | 0.024 |

The lower bound on the 87.5% repeated confidence interval is -0.042, comfortably within the non-inferiority margin of -0.05 specified at the design stage.

East also provides a p-value, confidence interval and median unbiased point estimate for $\pi_t - \pi_c$ using stage-wise ordering of the sample space as described in Jennison and Turnbull (2000, page 179). This is located in the **Adjusted Inference Table**, located in the lower section of the IM Worksheet. In the present example, the lower confidence

bound is -0.040, slightly greater than the corresponding bound from the repeated confidence interval.

24.2 Ratio of Proportions: Wald Formulation

24.2.1 Trial Design

24.2.2 Trial Simulation

24.2.3 Interim Monitoring

Let π_c and π_t denote the response rates for the control and the experimental treatments, respectively. Let the difference between the two arms be captured by the ratio

$$\rho = \frac{\pi_t}{\pi_c} .$$

The null hypothesis is specified as

$$H_0: \rho = \rho_0$$

and is tested against one-sided alternative hypotheses. If the occurrence of a response denotes patient benefit rather than harm, then $\rho_0 < 1$ and the alternative hypothesis is

$$H_1: \rho > \rho_0$$

or equivalently as

$$H_1: \pi_t > \rho_0 \pi_c .$$

Conversely, if the occurrence of a response denotes patient harm rather than benefit, then $\rho_0 > 1$ and the alternative hypothesis is

$$H_1: \rho < \rho_0$$

or equivalently as

$$H_1: \pi_t < \rho_0 \pi_c .$$

For any given π_c , the sample size is determined by the desired power at a specified value of $\rho = \rho_1$. A common choice is $\rho_1 = 1$ (or equivalently $\pi_t = \pi_c$), but East permits you to power the study at any value of ρ_1 which is consistent with the choice of H_1 .

Let $\hat{\pi}_{tj}$ and $\hat{\pi}_{cj}$ denote the estimates of π_t and π_c based on n_{tj} and n_{cj} observations from the experimental and control treatments, respectively, up to and including j -th look, $j = 1, \dots, K$, where a maximum of K looks are to be made. It is convenient to express the treatment effect on the logarithm scale as

$$\delta = \ln \rho = \ln \pi_t - \ln \pi_c . \tag{24.5}$$

The test statistic at the j th look is then defined as

$$Z_j = \frac{\hat{\delta}_j - \delta_0}{\text{se}(\hat{\delta}_j)} \tag{24.6}$$

24 Binomial Non-Inferiority Two-Sample

where

$$\hat{\delta}_j = \ln \left(\frac{\hat{\pi}_{tj}}{\hat{\pi}_{cj}} \right), \quad (24.7)$$

$$\delta_0 = \ln(\rho_0) \quad (24.8)$$

and

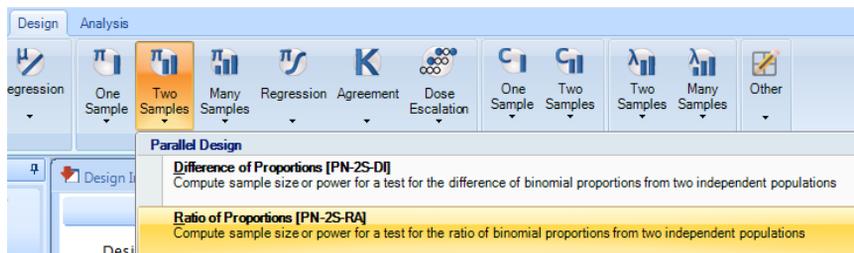
$$se(\hat{\delta}_j) = \sqrt{\frac{1 - \hat{\pi}_{cj}}{n_{cj}\hat{\pi}_{cj}} + \frac{1 - \hat{\pi}_{tj}}{n_{tj}\hat{\pi}_{tj}}}. \quad (24.9)$$

24.2.1 Trial Design

The Coronary Artery Revascularization in Diabetes (CARDia) trial (Kapur et. al., 2005) was designed to compare coronary bypass graft surgery (CABG) and percutaneous coronary intervention (PCI) as strategies for revascularization, with the goal of showing that PCI is noninferior to CABG. We use various aspects of that study to exemplify the methodology to test for inferiority. The endpoint is the one-year event rate, where an event is defined as the occurrence of death, nonfatal myocardial infarction, or cerebrovascular accident.

Suppose that the event rate for the CABG is $\pi_c = 0.125$ and that the claim of non-inferiority for PCI can be sustained if one can demonstrate statistically that the ratio $\rho = \pi_t/\pi_c$ is at most 1.3. In other words, PCI is considered to be non-inferior to CABG as long as $\pi_t < 0.1625$. Thus the null hypothesis $H_0: \rho = 1.3$ is tested against the one-sided alternative hypothesis $H_1: \rho < 1.3$. We want to determine the sample size required to have power of 80% when $\rho = 1$ using a one-sided test with a type-1 error rate of 0.05.

Single Look Design Powered at $\rho = 1$ First we consider a study with only one look and equal sample sizes in the two groups. To begin click **Two Proportions** on the **Design** tab under **Discrete**, and then click **Ratio of Proportions**.



In the ensuing dialog box, next to **Trial**, select *Noninferiority* from the drop down menu. Choose the remaining design parameters as shown below.

Design Type: Number of Looks:

Test Parameters

Test Type:

Type I Error (α):

Power:

Sample Size (n):

Allocation Ratio: (n_1 / n_2)

Assurance (Probability of Success):

Perform Exact Computations

Specify Proportion
Prop. under Control (π_c):

Specify Null Hypothesis
Prop. under Treatment (π_0):
Noninferiority Margin (ρ_0):
 $\rho_0 = \pi_0 / \pi_c$

Specify Alternative Hypothesis
Prop. under Treatment (π_1):
Ratio of Proportions (ρ_1):
 $\rho_1 = \pi_1 / \pi_c$

Test Statistic
 Wald
 Score (Farrington Manning)

Make sure to select the radio button for **Wald** in the **Test Statistic** box. We will discuss the **Score (Farrington Manning)** test statistic in the next section.

Now click **Compute**. The design is shown as a row in the Output Preview located in the lower pane of this window. This single-look design requires a combined total of 2515 subjects from both treatments in order to attain 80% power.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | π_c | Prop. Treatment (Alt.) | Prop. Treatment (Null) | ρ^1 | ρ^0 | Test Statistic |
|------|----------------|--------------|-----------|--------------------|-------|-------|-------------|---------|------------------------|------------------------|----------|----------|----------------|
| Des1 | Noninferiority | 1 | 1-Sided | 0.05 | 0.8 | 1 | 2515 | 0.125 | 0.125 | 0.163 | 1 | 1.3 | Wald |

You can select this design by clicking anywhere along the row in the **Output Preview**. Some of the design details will be displayed in the upper pane, labeled **Compare Designs**. In the **Output Preview** toolbar, click the  icon to save this design to Workbook1 in the **Library**. If you hover the cursor over Design1 in the Library, a

24 Binomial Non-Inferiority Two-Sample

tooltip will appear that summarizes the input parameters of the design.

The screenshot shows a software interface with a 'Library' pane on the left and a parameter table on the right. The 'Library' pane shows a tree structure with 'Root', 'Wbk1', and 'Des1'. A tooltip is displayed over 'Des1', listing design parameters. The parameter table on the right is titled 'Wbk1:Des1' and lists various parameters under 'Test Parameters' and 'Model Parameters'.

| Wbk1:Des1 | |
|--|----------------|
| Mnemonic | PN-2S-RA |
| Test Parameters | |
| Design Type | Noninferiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.8 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.125 |
| Proportion under Treatment (π_t) | 0.125 |
| Proportion Treatment (Null) (π_{t0}) | 0.163 |
| Ratio of Proportions (π_t / π_c) | 1 |
| Noninf. Margin (π_{t0} / π_c) | 1.3 |
| Test Statistic | Wald |
| Sample Size | |
| Maximum | 2515 |

Three-Look Design Powered at $\rho = 1$ For the above study, suppose we wish to take up to two equally spaced interim looks and one final look at the accruing data, using the Lan- DeMets (O'Brien-Fleming) stopping boundary. Create a new design by selecting Design1 in the **Library**, and clicking the  icon on the **Library** toolbar. Change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. A new tab **Boundary Info** will appear. Clicking on this tab will reveal the stopping boundary parameters. By default, the **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The left side contains details for the **Efficacy** boundary, and the right side contains details for the **Futility** boundary. By default, there is an efficacy boundary (to reject H_0) selected, but no futility boundary (to reject H_1). The **Boundary Family** specified is of the Spending Functions type. The default **Spending function** is the Lan-DeMets (Lan & DeMets, 1983), with **Parameter** OF (O'Brien-Fleming), which generates boundaries that are very similar, though not identical, to the classical stopping boundaries of O'Brien and Fleming

(1979). Technical details of these stopping boundaries are available in Appendix F.

Design Type: Number of Looks:

Test Parameters **Boundary**

Efficacy
 Boundary Family:
 Spending Function:
 Parameter:
 Type I Error (α): 0.05

Futility
 Boundary Family:

Spacing of Looks Equal Unequal

Efficacy Boundary:

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.333 | 0.001 | -3.200 |
| 2 | 0.667 | 0.016 | -2.141 |
| 3 | 1.000 | 0.050 | -1.695 |

Click the **Compute** button to generate output for Design2. With Design2 selected in the **Output Preview**, click the icon to save Design2 to the **Library**. In the **Library**, select the rows for Design1 and Design2, by holding the Ctrl key, and then click the icon. The upper pane will display the details of the two designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 |
|--|----------------|----------------|
| Mnemonic | PN-25-RA | PN-25-RA |
| Test Parameters | | |
| Design Type | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.8 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (n_t/n_c) | 1 | 1 |
| Proportion under Control (π_c) | 0.125 | 0.125 |
| Proportion under Treatment (π_t) | 0.125 | 0.125 |
| Proportion Treatment (Null) (π_{t0}) | 0.163 | 0.163 |
| Ratio of Proportions (π_t / π_c) | 1 | 1 |
| Noninf. Margin (π_{t0} / π_c) | 1.3 | 1.3 |
| Test Statistic | Wald | Wald |
| Boundary Parameters | | |
| Spacing of Looks | | Equal |
| Efficacy Boundary | | LD (OF) |
| Sample Size | | |
| Maximum | 2515 | 2566 |
| Expected Under H0 | | 2551.404 |
| Expected Under H1 | | 2133.893 |

Using three planned looks requires an up-front commitment of 2566 subjects, a slight

24 Binomial Non-Inferiority Two-Sample

inflation over the single-look design which required 2515 subjects. However, the three-look design may result in a smaller sample size than that required for the single-look design, with an expected sample size of 2134 subjects under the alternative hypothesis ($\pi_c = 0.125, \rho = 1$), and still ensures that the power is 80%.

By selecting Design2 in the **Library** and clicking on the  icon, East displays the cumulative accrual, the stopping boundary, the type-1 error spent and the boundary crossing probabilities under the null hypothesis $H_0: \rho = 1.3$, and the alternative hypothesis $H_1: \rho = 1$.

☰ Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|------------|---|----------|
| | | | | Efficacy Z | Under H0 | Under H1 |
| | | | | | Efficacy | Efficacy |
| 1 | 0.333 | 855 | 6.852E-4 | -3.201 | 6.852E-4 | 0.04 |
| 2 | 0.667 | 1711 | 0.016 | -2.141 | 0.016 | 0.425 |
| 3 | 1 | 2566 | 0.05 | -1.695 | 0.034 | 0.335 |

Single-Look Design Powered at $\rho \neq 1$ Sample sizes for non-inferiority trials powered at $\rho = 1$ are generally rather large, because regulatory requirements usually impose small non-inferiority margins (see, for example, Wang et. al., 2001). Observe that both Design1 and Design2 were powered at $\rho = 1$ and required sample sizes in excess of 2500 subjects. However, based on Kapur et al (2005), it is reasonable to expect $\pi_t < \pi_c$. We now consider the same design as in Design1, but we will power at the alternative hypothesis $\rho_1 = 0.72$. That is, we will design this study to have 80% power to claim non-inferiority if $\pi_c = 0.125$ and $\pi_t = 0.72 \times 0.125 = 0.09$.

Create a new design by selecting Design1 in the **Library**, and clicking the  icon on the **Library** toolbar. In the ensuing dialog box, change the design parameters as

shown below.

Design Type: Number of Looks:

Test Parameters

Test Type:

Type I Error (α):

Power:

Sample Size (n):

Allocation Ratio:
(n_1/n_2)

Specify Proportion
Prop. under Control (π_c):

Specify Null Hypothesis
Prop. under Treatment (π_0):
Noninferiority Margin (ρ_0):
 $\rho_0 = \pi_0 / \pi_c$

Specify Alternative Hypothesis
Prop. under Treatment (π_1):
Ratio of Proportions (ρ_1):
 $\rho_1 = \pi_1 / \pi_c$

Perform Exact Computations

Test Statistic
 Wald
 Score (Farrington Manning)

Assurance (Probability of Success):

Click the **Compute** button to generate output for Design3. With Design3 selected in the **Output Preview**, click the  icon. In the **Library**, select the rows for Design1, Design2, and Design3, by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the three designs side-by-side:

| | Wbk1.Des1 | Wbk1.Des2 | Wbk1.Des3 |
|--|----------------|----------------|----------------|
| Mnemonic | PN-25-RA | PN-25-RA | PN-25-RA |
| Test Parameters | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 | 1 |
| Test Type | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.8 | 0.8 | 0.801 |
| Model Parameters | | | |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Proportion under Control (π_c) | 0.125 | 0.125 | 0.125 |
| Proportion under Treatment (π_t) | 0.125 | 0.125 | 0.09 |
| Proportion Treatment (Null) (π_0) | 0.163 | 0.163 | 0.163 |
| Ratio of Proportions (π_t / π_c) | 1 | 1 | 0.72 |
| Noninf. Margin (π_0 / π_c) | 1.3 | 1.3 | 1.3 |
| Test Statistic | Wald | Wald | Wald |
| Boundary Parameters | | | |
| Spacing of Looks | | Equal | |
| Efficacy Boundary | | LD (OF) | |
| Sample Size | | | |
| Maximum | 2515 | 2566 | 607 |
| Expected Under H0 | | 2551.404 | |
| Expected Under H1 | | 2133.893 | |

This single-look design requires a combined total of 607 subjects from both treatments in order to attain 80% power. This is a considerable decrease from the 2515 subjects required to attain 80% power using Design1 with $\rho_1 = 1$.

Three-Look Design Powered at $\rho \neq 1$ We now consider the impact of multiple looks on Design3. Suppose we wish to take up to two equally spaced interim looks and one final look at the accruing data, using the Lan-DeMets (O'Brien-Fleming) stopping

24 Binomial Non-Inferiority Two-Sample

boundary.

Create a new design by selecting Design3 in the **Library**, and clicking the  icon on the **Library** toolbar. In the ensuing dialog box, change the **Number of Looks** to 3.

Design Type: Noninferiority Number of Looks: 3

Test Parameters Boundary

Test Type: 1-Sided

Type I Error (α): 0.05

Power: 0.8

Sample Size (n): Computed

Allocation Ratio: 1 (n_t/n_c)

Specify Proportion

Prop. under Control (π_c): 0.125

Specify Null Hypothesis

Prop. under Treatment (π_0): 0.163

Noninferiority Margin (ρ_0): 1.3

$\rho_0 = \pi_0 / \pi_c$

Specify Alternative Hypothesis

Prop. under Treatment (π_1): 0.09

Ratio of Proportions (ρ_1): 0.72

$\rho_1 = \pi_1 / \pi_c$

Test Statistic

Wald

Score (Farrington Manning)

Assurance (Probability of Success):

Click the **Compute** button to generate output for Design4.

| | Wbk1:Des1 | Wbk1:Des2 | Wbk1:Des3 | Wbk1:Des4 |
|--|----------------|----------------|----------------|----------------|
| Mnemonic | PN-25-RA | PN-25-RA | PN-25-RA | PN-25-RA |
| Test Parameters | | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 | 1 | 3 |
| Test Type | 1-Sided | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 | 0.05 | 0.05 |
| Power | 0.8 | 0.8 | 0.801 | 0.8 |
| Model Parameters | | | | |
| Allocation Ratio (n_t/n_c) | 1 | 1 | 1 | 1 |
| Proportion under Control (π_c) | 0.125 | 0.125 | 0.125 | 0.125 |
| Proportion under Treatment (π_t) | 0.125 | 0.125 | 0.09 | 0.09 |
| Proportion Treatment (Null) (π_0) | 0.163 | 0.163 | 0.163 | 0.163 |
| Ratio of Proportions (π_t / π_c) | 1 | 1 | 0.72 | 0.72 |
| Noninf. Margin (π_0 / π_c) | 1.3 | 1.3 | 1.3 | 1.3 |
| Test Statistic | Wald | Wald | Wald | Wald |
| Boundary Parameters | | | | |
| Spacing of Looks | | Equal | | Equal |
| Efficacy Boundary | | LD (OF) | | LD (OF) |
| Sample Size | | | | |
| Maximum | 2515 | 2566 | 607 | 619 |
| Expected Under H0 | | 2551.404 | | 615.477 |
| Expected Under H1 | | 2133.893 | | 514.694 |

Using three planned looks inflates the maximum sample size slightly, from 607 to 619 subjects. However it results in a smaller expected sample size under H_1 . Observe that the expected sample size is only 515 subjects under the alternative hypothesis ($\pi_c = 0.125, \rho = 0.72$), and still ensures the power is 80%.

24.2.2 Trial Simulation

You can simulate Design4 by selecting it from the **Library** and clicking on the  icon. Try different choices for the simulation parameters to verify the operating characteristics of the study. For instance, under the **Response Generation Info** tab set **Prop. Under Control** to 0.125 and **Prop. Under Treatment** to $0.72 \times 0.125 = 0.09$.

Number of Looks:

Test Parameters
Response Generation
Simulation Controls

Specify Proportion

Prop. under Control (π_c):

Prop. under Treatment (π_{t1}):

Click **Simulate** button. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled Simulation1. Select Simulation1 in the **Output Preview**. Note that some of the design details will be displayed in the upper pane, labeled **Compare Designs**. Click the  icon to save it to the **Library**. Double-click on Simulation1 in the **Library**. The simulation output details will be displayed.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Ratio of Proportions

| Test Parameters | |
|--------------------------------------|----------------|
| Simulation ID | Sim1 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 619 |
| Test Statistic | Wald |
| Noninf. Margin (p_0) | 1.3 |
| Avg. Power at Termination | 0.811 |
| Response Generation Parameters | |
| Prop. under Control (π_c) | 0.125 |
| Prop. under Treatment (π_{t1}) | 0.09 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 206 |
| 2 | 413 |
| 3 | 619 |
| Average | 513.906 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 206 | -3.203 | 138 | 138 | 1.380% |
| 2 | 413 | -2.14 | 4825 | 4825 | 48.250% |
| 3 | 619 | -1.695 | 3145 | 5037 | 50.370% |
| Total | | | 8108 | 10000 | |
| % | | | 81.080% | | |

Simulation Seed and Elapsed Time

Starting Seed: 1500486
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:04

24 Binomial Non-Inferiority Two-Sample

We simulated the data under the alternative hypothesis and should achieve a rejection rate of 80%. This is confirmed above (up to Monte Carlo accuracy).

Next, to simulate under the null hypothesis, under the **Response Generation Info** tab set **Prop. Under Treatment** to $1.3 \times 0.125 = 0.1625$. Click **Simulate** button.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Ratio of Proportions

| Test Parameters | |
|-----------------------------------|----------------|
| Simulation ID | Sim2 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 619 |
| Test Statistic | Wald |
| Noninf. Margin (p_0) | 1.3 |
| Avg. Power at Termination | 0.05 |
| Response Generation Parameters | |
| Prop. under Control (π_c) | 0.125 |
| Prop. under Treatment (π_t) | 0.163 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 206 |
| 2 | 413 |
| 3 | 619 |
| Average | 616.013 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 206 | -3.203 | 5 | 5 | 0.050% |
| 2 | 413 | -2.14 | 135 | 135 | 1.350% |
| 3 | 619 | -1.695 | 356 | 9860 | 98.600% |
| Total | | | 496 | 10000 | |
| % | | | 4.960% | | |

Simulation Seed and Elapsed Time

Starting Seed: 1771497
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:03

This time the rejection rate is only 5% (up to Monte Carlo accuracy), as we would expect under the null hypothesis. You may experiment in this manner with different values of π_c and π_t and observe the rejection rates look by look as well as averaged over all looks.

24.2.3 Interim Monitoring

Select Design4 in the **Library**, and click the **IM** icon from the Library toolbar. Alternatively, right-click on Design4 and select **Create IM Dashboard**. The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs. The bottom section contains four charts, each with a corresponding table to its right. These charts provide graphical and numerical descriptions of the progress of the clinical trial and are useful tools for decision making by a data monitoring committee.

Suppose that the trial is first monitored after accruing 125 subjects on each treatment

arm, with 15 responses on the control arm and 13 responses on the treatment arm.

Click on the **Enter Interim Data** icon to invoke the **Test Statistic Calculator**. In the box next to **Cumulative Sample Size** enter 250. Enter -0.143101 in the box next to **Estimate of δ** . In the box next to **Std. Error of δ** enter 0.357197. Next click **Recalc**. Notice that the test statistic is computed to be -1.135 . This value for the test statistic was obtained by substituting the observed sample sizes and responses into equations (24.6) through (24.9).

Upon clicking the **OK** button, East will produce the interim monitoring report shown below.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | Est. of p | Est. of δ | Std. Error of Est. of δ | Efficacy | 95% RCI for p | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|-------------|------------------|--------------------------------|----------|-----------------|-------|------------------|-------|------------------|
| | | | | | | | | Upper | Lower | | | |
| 1 | 0.404 | 250 | -1.135 | 0.867 | -0.143 | 0.357 | -2.872 | 2.417 | 0 | 0.334 | 0.556 | 0.536 |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |

24 Binomial Non-Inferiority Two-Sample

Note - Click on  icon to hide or unhide the columns of your interest.

The stopping boundary for declaring non-inferiority is -2.872 whereas the value of the test statistic is only -1.135. Thus the trial should continue.

This conclusion is supported by the value of the 97.5% upper confidence bound of the repeated confidence interval for $\delta = \ln(\rho)$. The non-inferiority claim could be sustained only if this bound were less than $\ln(1.3) = 0.262$. At the current interim look, however, the upper bound on δ is 0.883, indicating that the non-inferiority claim is not supported by the data.

Suppose that the next interim look occurs after accruing 250 patients on each arm with 31 responses on the control arm and 22 responses on the treatment arm. Click on the second row in the table in the upper section. Then click the **Enter Interim Data** icon. In the box next to **Cumulative Sample Size** enter 500. Enter -0.342945 in the box next to **Estimate of δ** . In the box next to **Std. Error of δ** enter 0.264031. Next click

Recalc. Notice that the test statistic is computed to be -2.293.

Test Statistic Calculator

Editing Look #2
 Set Current Look as Last

Sample Size and Responses

| | Control | Treatment |
|-------------------------|---------|-----------|
| Cumulative Sample Size: | N/A | N/A |
| Cumulative Response: | N/A | N/A |

Cumulative Sample Size:

Input for Binomial end point

Estimate of δ :

$\delta = \ln(\pi_t / \pi_c)$

Standard Error of Estimate of δ :

Output

Estimate of $\delta - \delta_0$:

Test Statistic:

Click the **OK** button. This time the stopping boundary for declaring non-inferiority is crossed. The following message box appears.

Boundary Crossed

Since the value of Test Statistic is \leq the critical point for efficacy, H_0 is rejected.

Although boundary has been crossed, East gives you choice either to stop the study or to continue entering further looks. Please make your decision.

stop the study and bar further looks input

allow the study to continue

24 Binomial Non-Inferiority Two-Sample

Click the **Stop** button to stop the study. The analysis results are shown below.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | Est. of ρ | Est. of δ | Std. Error of Est. of δ | Efficacy | 95% RCI for ρ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|----------------|------------------|--------------------------------|----------|--------------------|-------|------------------|-------|------------------|
| | | | | | | | | Upper | Lower | | | |
| 1 | 0.404 | 250 | -1.135 | 0.867 | -0.143 | 0.357 | -2.872 | 2.417 | 0 | 0.334 | 0.556 | 0.536 |
| 2 | 0.808 | 500 | -2.293 | 0.71 | -0.343 | 0.264 | -1.903 | 1.173 | 0 | 0.022 | NA | NA |

Click the "Edit Interim Data" button to edit the Look # 2 data.

Stopping Boundaries

| Sample Size | Efficacy |
|-------------|----------|
| 250 | -2.872 |
| 500 | -1.903 |

Final Inference

| Final Outputs at Look # | Value |
|----------------------------|--------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.012 |
| Adj. Pt. Est. for ρ | 0.712 |
| Adj. 90% CI for ρ | |
| Upper Confidence Bound | 1.103 |
| Lower Confidence Bound | 0.461 |
| Adj. Pt. Est. for δ | -0.339 |
| Adj. 90% CI for δ | |
| Upper Confidence Bound | 0.098 |

Conf. Intervals for ρ

| Info. Fraction | RCI Upper | RCI Lower | Naive CI Upper | Naive CI Lower |
|----------------|-----------|-----------|----------------|----------------|
| 0.404 | 2.417 | 0 | 1.56 | 0 |
| 0.808 | 1.173 | 0 | 1.096 | 0 |

The upper bound on the 95.0% repeated confidence interval for δ is 0.159. Thus the upper confidence bound on ρ is $\exp(0.159) = 1.172$, comfortably within the non-inferiority margin $\rho_0 = 1.3$ specified at the design stage.

In the **Final Inference** Table in the bottom portion of the IM worksheet, East also provides a p-value, confidence interval and median unbiased point estimate for δ using stage-wise ordering of the sample space as described in Jennison and Turnbull (2000). This approach often yields narrower confidence intervals than the repeated confidence intervals approach although both approaches have the desired 95.0% coverage. In the present example, the upper confidence bound is 0.098, slightly less than the corresponding bound from the repeated confidence interval.

24.3 Ratio of Proportions: Farrington-Manning Formulation

24.3.1 Trial Design

24.3.2 Trial Simulation

24.3.3 Interim Monitoring

An alternative approach to establishing non-inferiority of an experimental treatment to the control treatment with respect to the ratio of probabilities was proposed by Farrington and Manning (1990). Let π_c and π_t denote the response rates for the control and the experimental treatments, respectively. Let the difference between the two arms be expressed by the ratio

$$\rho = \frac{\pi_t}{\pi_c}$$

The null hypothesis is specified as

$$H_0: \rho = \rho_0,$$

or equivalently

$$H_0: \pi_t = \rho_0 \pi_c,$$

which is tested against one-sided alternative hypotheses. If the occurrence of a response denotes patient benefit rather than harm, then $\rho_0 < 1$ and the alternative hypothesis is

$$H_1: \rho > \rho_0$$

or equivalently as

$$H_1: \pi_t > \rho_0 \pi_c .$$

Conversely, if the occurrence of a response denotes patient harm rather than benefit, then $\rho_0 > 1$ and the alternative hypothesis is

$$H_1: \rho < \rho_0$$

or equivalently as

$$H_1: \pi_t < \rho_0 \pi_c .$$

For any given π_c , the sample size is determined by the desired power at a specified value of $\rho = \rho_1$. A common choice is $\rho_1 = 1$ (or equivalently $\pi_t = \pi_c$), but East permits you to power the study at any value of ρ_1 which is consistent with the choice of H_1 .

Let $\hat{\pi}_{tj}$ and $\hat{\pi}_{cj}$ denote the estimates of π_t and π_c based on n_{tj} and n_{cj} observations from the experimental and control treatments, respectively, up to and including the j -th look, $j = 1, \dots, K$, where a maximum of K looks are to be made. The test statistic at the j -th look is defined as

$$Z_j = \frac{\hat{\pi}_{tj} - \rho_0 \hat{\pi}_{cj}}{\sqrt{\left[\frac{\hat{\pi}_{tj}(1-\hat{\pi}_{tj})}{n_{tj}} + \frac{\rho_0^2 \hat{\pi}_{cj}(1-\hat{\pi}_{cj})}{n_{cj}} \right]}} . \quad (24.10)$$

The choice of test statistic is the primary distinguishing feature between the above Farrington-Manning formulation and the Wald formulation of the non-inferiority test discussed in Section 24.2. The Wald statistic (24.6) measures the standardized difference between the observed ratio of proportions and the non-inferiority margin on the natural logarithm scale. The corresponding repeated one-sided confidence bounds displayed in the interim monitoring worksheet estimate $\ln(\pi_t/\pi_c)$ and may be converted to estimates of the ratio of proportions by exponentiation. On the other hand,

24 Binomial Non-Inferiority Two-Sample

the Farrington-Manning formulation focuses on the expression of the null hypothesis as

$$H_0: \pi_t - \rho_0 \pi_c = 0.$$

Thus, we consider

$$\delta = \pi_t - \rho_0 \pi_c \tag{24.11}$$

as the parameter of interest. The test statistic (24.10) is the standardized estimate of this difference obtained at the j -th look. A large difference in the direction of the alternative hypothesis is indicative of non-inferiority. The corresponding repeated one-sided confidence bounds displayed in the interim monitoring worksheet provide estimates of δ rather than directly estimating ρ or $\ln(\rho)$. The Farrington-Manning and Wald procedures are equally applicable for hypothesis testing since the null hypothesis $\delta = 0$ is rejected if and only if the corresponding null hypothesis $\rho = \rho_0$ is rejected.

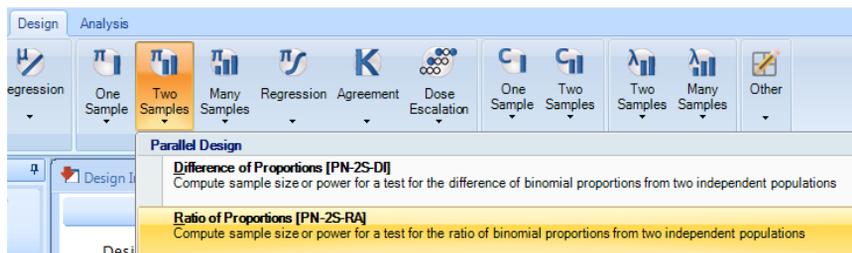
24.3.1 Trial Design

We consider the Coronary Artery Revascularization in Diabetes (CARDia) trial (Kapur et al, 2005) compared coronary bypass graft surgery (CABG) and percutaneous coronary intervention (PCI) as strategies for revascularization, with the goal of showing that PCI is noninferior to CABG, presented in Section 24.2. We use various aspects of that study to exemplify the use of the methodology to test for inferiority with respect to the one-year event rate where an "event" is the occurrence of death, nonfatal myocardial infarction, or cerebrovascular accident, using the Farrington-Manning formulation.

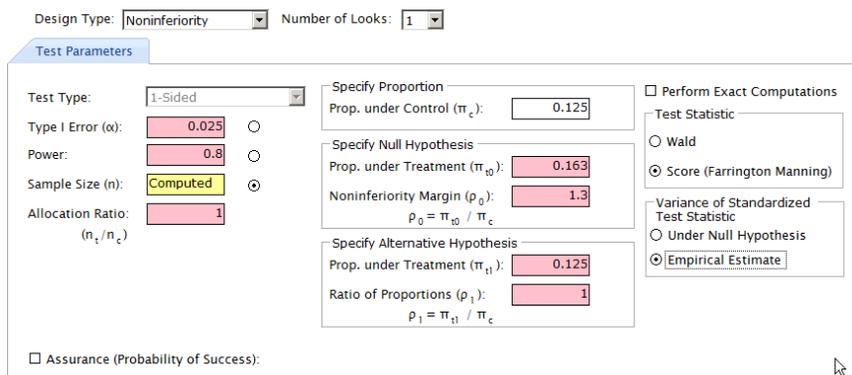
Suppose that the event rate for the CABG is $\pi_c = 0.125$ and that the claim of non-inferiority for PCI can be sustained if the ratio ρ is at most 1.3; that is, the event rate for the PCI (π_t) is at most 0.1625. The null hypothesis $H_0: \rho = 1.3$ is tested against the alternative hypothesis $H_1: \rho < 1.3$. We want to determine the sample size required to have power of 80% when $\rho = 1$ using a one-sided test with a type-1 error rate of 0.05.

Single Look Design Powered at $\rho = 1$ First we consider a study with only one look and equal sample sizes in the two groups. To begin click **Two Proportions** on the

Design tab, and then click **Ratio of Proportions**.



In the ensuing dialog box, next to **Trial**, select *Noninferiority* from the drop down menu. Choose the remaining design parameters as shown below.



Now click **Compute**. The design is shown as a row in the Output Preview located in the lower pane of this window. This single-look design requires a combined total of 2588 subjects from both treatments in order to attain 80% power.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | πc | Prop. Treatment (Alt.) | Prop. Treatment (Null) | ρ1 | ρ0 | Test Statistic | Variance |
|------|----------------|--------------|-----------|-------------|-------|-------|-------------|-------|------------------------|------------------------|----|-----|----------------|--------------------|
| Des1 | Noninferiority | 1 | 1-Sided | 0.025 | 0.8 | 1 | 3285 | 0.125 | 0.125 | 0.163 | 1 | 1.3 | Score (FM) | Empirical Estimate |

You can select this design by clicking anywhere along the row in the **Output Preview**. Some of the design details will be displayed in the upper pane, labeled **Compare Designs**. In the **Output Preview** toolbar, click the  icon to save this design to Workbook1 in the **Library**. If you hover the cursor over Design1 in the Library, a

24 Binomial Non-Inferiority Two-Sample

tooltip will appear that summarizes the input parameters of the design.

| | Des 1 |
|--|--------------------|
| Mnemonic | PN-2S-RA |
| Test Parameters | |
| Design Type | Noninferiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.8 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.125 |
| Proportion under Treatment (π_t) | 0.125 |
| Proportion Treatment (Null) (π_0) | 0.163 |
| Ratio of Proportions (π_t / π_c) | 1 |
| Noninf. Margin (π_0 / π_c) | 1.3 |
| Test Statistic | Score (FM) |
| Variance | Empirical Estimate |
| Sample Size | |
| Maximum | 3285 |

Three-Look Design Powered at $\rho = 1$ For the above study, suppose we wish to take up to two equally spaced interim looks and one final look at the accruing data, using the Lan- DeMets (O'Brien-Fleming) stopping boundary. Create a new design by selecting Design1 in the **Library**, and clicking the  icon on the **Library** toolbar. Change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. A new tab **Boundary Info** will appear. Clicking on this tab will reveal the stopping boundary parameters. By default, the **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The left side contains details for the **Efficacy** boundary, and the right side contains details for the **Futility** boundary. By default, there is an efficacy boundary (to reject H0) selected, but no futility boundary (to reject H1). The **Boundary Family** specified is of the Spending Functions type. The default **Spending function** is the Lan-DeMets (Lan & DeMets, 1983), with **Parameter** OF (O'Brien-Fleming), which generates boundaries that are very similar, though not identical, to the classical stopping boundaries of O'Brien and Fleming

(1979).

Design Type: Number of Looks:

Test Parameters **Boundary**

Test Type:

Type I Error (α):

Power:

Sample Size (n):

Allocation Ratio:
(n_1/n_2)

Assurance (Probability of Success):

Specify Proportion
 Prop. under Control (π_c):

Specify Null Hypothesis
 Prop. under Treatment (π_{10}):
 Noninferiority Margin (ρ_0):
 $\rho_0 = \pi_{10} / \pi_c$

Specify Alternative Hypothesis
 Prop. under Treatment (π_{11}):
 Ratio of Proportions (ρ_1):
 $\rho_1 = \pi_{11} / \pi_c$

Test Statistic
 Wald
 Score (Farrington Manning)

Variance of Standardized Test Statistic
 Under Null Hypothesis
 Empirical Estimate

Click the **Compute** button to generate output for Design2. With Design2 selected in the **Output Preview**, click the  icon to save Design2 to the **Library**. In the **Library**, select the rows for Design1 and Design2, by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the two designs

24 Binomial Non-Inferiority Two-Sample

side-by-side:

| | Des1 | Des2 |
|--|--------------------|--------------------|
| Mnemonic | PN-2S-RA | PN-2S-RA |
| Test Parameters | | |
| Design Type | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.8 | 0.8 |
| Model Parameters | | |
| Proportion under Control (π_c) | 0.125 | 0.125 |
| Proportion under Treatment (π_t) | 0.125 | 0.125 |
| Proportion Treatment (Null) (π_{t0}) | 0.163 | 0.163 |
| Ratio of Proportions (π_t / π_c) | 1 | 1 |
| Noninf. Margin (π_{t0} / π_c) | 1.3 | 1.3 |
| Test Statistic | Score (FM) | Score (FM) |
| Variance | Empirical Estimate | Empirical Estimate |
| Allocation Ratio (n_t/n_c) | 1 | 1 |
| Boundary Parameters | | |
| Efficacy Boundary | | LD (OF) |
| Spacing of Looks | | Equal |
| Sample Size | | |
| Maximum | 2588 | 2640 |
| Expected Under H0 | | 2624.986 |
| Expected Under H1 | | 2195.332 |

Using three planned looks requires an up-front commitment of 2640 subjects, a slight inflation over the single-look design which required only 2588 subjects. However, the three-look design may result in a smaller sample size than that required for the single-look design, with an expected sample size of 2195 subjects under the alternative hypothesis ($\pi_c = 0.125, \rho = 1$), and still ensures that the power is 80%.

By selecting Design2 in the **Library** and clicking on the  icon, East displays the cumulative accrual, the stopping boundary, the type-1 error spent and the boundary crossing probabilities under the null hypothesis $H_0: \rho = 1.3$, and the alternative hypothesis $H_1: \rho = 1$.

⊖ Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|------------|---|----------|
| | | | | | Under H0 | Under H1 |
| | | | | Efficacy Z | Efficacy | Efficacy |
| 1 | 0.333 | 880 | 6.869E-4 | -3.2 | 6.869E-4 | 0.04 |
| 2 | 0.667 | 1760 | 0.016 | -2.141 | 0.016 | 0.425 |
| 3 | 1 | 2640 | 0.05 | -1.695 | 0.034 | 0.335 |

Single-Look Design Powered at $\rho \neq 1$ Sample sizes for non-inferiority trials powered at $\rho = 1$ are generally rather large because regulatory requirements usually impose small non-inferiority margins. Observe that both Design1 and Design2 were powered at $\rho = 1$ and required sample sizes in excess of 2500 subjects. However, based on Kapur et al (2005), it is reasonable to expect $\pi_t < \pi_c$. We now consider the same design as in Design1, but we will power at the alternative hypothesis $\rho_1 = 0.72$. That is, we will design this study to have 80% power to claim non-inferiority if $\pi_c = 0.125$ and $\pi_t = 0.72 \times 0.125 = 0.09$.

Create a new design by selecting Design1 in the **Library**, and clicking the  icon on the **Library** toolbar. In the ensuing dialog box, change the design parameters as shown below.

Design Type: Noninferiority Number of Looks: 3

Test Parameters Boundary

Test Type: 1-Sided

Type I Error (α): 0.025

Power: 0.8

Sample Size (n): Computed

Allocation Ratio: 1
(n_1/n_2)

Specify Proportion
 Prop. under Control (π_c): 0.125

Specify Null Hypothesis
 Prop. under Treatment (π_0): 0.163
 Noninferiority Margin (ρ_0): 1.3
 $\rho_0 = \pi_0 / \pi_c$

Specify Alternative Hypothesis
 Prop. under Treatment (π_{t1}): 0.125
 Ratio of Proportions (ρ_1): 1
 $\rho_1 = \pi_{t1} / \pi_c$

Test Statistic
 Wald
 Score (Farrington Manning)

Variance of Standardized Test Statistic
 Under Null Hypothesis
 Empirical Estimate

Assurance (Probability of Success):

Click the **Compute** button to generate output for Design3. With Design3 selected in the **Output Preview**, click the  icon. In the **Library**, select the rows for Design1, Design2, and Design3, by holding the Ctrl key, and then click the .

24 Binomial Non-Inferiority Two-Sample

icon. The upper pane will display the details of the three designs side-by-side:

| | Wbk1-Des1 | Wbk1-Des2 | Wbk1-Des3 |
|--|--------------------|--------------------|--------------------|
| Mnemonic | PN-2S-RA | PN-2S-RA | FN-2S-RA |
| Test Parameters | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 | 3 |
| Test Type | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.05 | 0.025 |
| Power | 0.8 | 0.8 | 0.8 |
| Model Parameters | | | |
| Allocation Ratio (nt / nc) | 1 | 1 | 1 |
| Proportion under Control (π_c) | 0.125 | 0.125 | 0.125 |
| Proportion under Treatment (π_t) | 0.125 | 0.125 | 0.125 |
| Proportion Treatment (Null) (π_t0) | 0.163 | 0.163 | 0.163 |
| Ratio of Proportions (π_t / π_c) | 1 | 1 | 1 |
| Noninf. Margin (π_t0 / π_c) | 1.3 | 1.3 | 1.3 |
| Test Statistic | Score (FM) | Score (FM) | Score (FM) |
| Variance | Empirical Estimate | Empirical Estimate | Empirical Estimate |
| Boundary Parameters | | | |
| Spacing of Looks | | Equal | Equal |
| Efficacy Boundary | | LD (OF) | LD (OF) |
| Sample Size | | | |
| Maximum | 3285 | 2640 | 3327 |
| Expected Under H0 | | 2624.986 | 3320.178 |
| Expected Under H1 | | 2195.332 | 2843.261 |

This single-look design requires a combined total of 628 subjects from both treatments in order to attain 80% power. This is a considerable decrease from the 2588 subjects required to attain 80% power using Design1, i.e. with $\rho_1 = 1$.

Three-Look Design Powered at $\rho \neq 1$ We now consider the impact of multiple looks on Design3. Suppose we wish to take up to two equally spaced interim looks and one final look at the accruing data, using the Lan- DeMets (O'Brien-Fleming) stopping boundary.

Create a new design by selecting Design3 in the **Library**, and clicking the  icon on the **Library** toolbar. In the ensuing dialog box, change the **Number of Looks** to 3.

Design Type: Noninferiority Number of Looks: 3

Test Parameters **Boundary**

Test Type: 1-Sided

Type I Error (α): 0.05

Power: 0.8

Sample Size (n): Computed

Allocation Ratio: 1 (n_t/n_c)

Specify Proportion

Prop. under Control (π_c): 0.125

Specify Null Hypothesis

Prop. under Treatment (π_{t0}): 0.163

Noninferiority Margin (ρ_0): 1.3
 $\rho_0 = \pi_{t0} / \pi_c$

Specify Alternative Hypothesis

Prop. under Treatment (π_{t1}): 0.125

Ratio of Proportions (ρ_1): 1
 $\rho_1 = \pi_{t1} / \pi_c$

Test Statistic

Wald

Score (Farrington Manning)

Variance of Standardized Test Statistic

Under Null Hypothesis

Empirical Estimate

Assurance (Probability of Success):

Click the **Compute** button to generate output for Design4.

| | Wbk1-Des1 | Wbk1-Des2 | Wbk1-Des3 | Wbk1-Des4 |
|-----------------------------------|--------------------|--------------------|--------------------|--------------------|
| Mnemonic | PN-2S-RA | PN-2S-RA | PN-2S-RA | PN-2S-RA |
| Test Parameters | | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 | 3 | 3 |
| Test Type | 1-Sided | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.05 | 0.025 | 0.05 |
| Power | 0.8 | 0.8 | 0.8 | 0.8 |
| Model Parameters | | | | |
| Allocation Ratio (tt / tc) | 1 | 1 | 1 | 1 |
| Proportion under Control (ttc) | 0.125 | 0.125 | 0.125 | 0.125 |
| Proportion under Treatment (ttt) | 0.125 | 0.125 | 0.125 | 0.125 |
| Proportion Treatment (Null) (tt0) | 0.163 | 0.163 | 0.163 | 0.163 |
| Ratio of Proportions (tt / ttc) | 1 | 1 | 1 | 1 |
| Noninf. Margin (tt0 / ttc) | 1.3 | 1.3 | 1.3 | 1.3 |
| Test Statistic | Score (FM) | Score (FM) | Score (FM) | Score (FM) |
| Variance | Empirical Estimate | Empirical Estimate | Empirical Estimate | Empirical Estimate |
| Boundary Parameters | | | | |
| Spacing of Looks | | Equal | Equal | Equal |
| Efficacy Boundary | | LD (OF) | LD (OF) | LD (OF) |
| Sample Size | | | | |
| Maximum | 3285 | 2640 | 3327 | 2640 |
| Expected Under H0 | | 2624.986 | 3320.178 | 2624.986 |
| Expected Under H1 | | 2195.332 | 2843.261 | 2195.332 |

Using three planned looks inflates the maximum sample size slightly, from 628 to 641 subjects. However it results in a smaller expected sample size under H_1 . Observe that the expected sample size is only 533 subjects under the alternative hypothesis ($\pi_c = 0.125, \rho = 0.72$), and still ensures the power is 80%.

24.3.2 Trial Simulation

You can simulate Design4 by selecting Design4 in the **Library** and clicking on the  icon. Try different choices for the simulation parameters to verify the operating characteristics of the study. For instance, under the **Response Generation Info** tab set **Prop. Under Control** to 0.125 and **Prop. Under Treatment** to $0.72 \times 0.125 = 0.09$.

Number of Looks:

Test Parameters Response Generation Simulation Controls

Specify Proportion

Prop. under Control (π_c):

Prop. under Treatment (π_{t1}):

Click **Simulate** button. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled Simulation1. Select Simulation1 in the **Output Preview**. Note that some of the design details will be displayed in the upper pane, labeled **Compare Designs**. Click the  icon to save it to the **Library**. Double-click on Simulation1 in the **Library**. The simulation output details will be

24 Binomial Non-Inferiority Two-Sample

displayed.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Ratio of Proportions

| Test Parameters | |
|--------------------------------------|----------------------------|
| Simulation ID | Sim1 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 2640 |
| Test Statistic | Score (Farrington Manning) |
| Variance | Empirical Estimate |
| Noninf. Margin (p_0) | 1.3 |
| Avg. Power at Termination | 1 |
| Response Generation Parameters | |
| Prop. under Control (π_0) | 0.125 |
| Prop. under Treatment (π_{11}) | 0.09 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 880 |
| 2 | 1760 |
| 3 | 2640 |
| Average | 1418.648 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | | Efficacy | Count |
| 1 | 880 | -3.2 | 4083 | 4083 | 40.830% |
| 2 | 1760 | -2.141 | 5713 | 5713 | 57.130% |
| 3 | 2640 | -1.695 | 204 | 204 | 2.040% |
| Total | | | 10000 | 10000 | |
| % | | | 100.000% | | |

Simulation Seed and Elapsed Time

Starting Seed: 1940268
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:03

We simulated the data under the alternative hypothesis and should achieve a rejection rate of 80%. This is confirmed above (up to Monte Carlo accuracy).

Next, to simulate under the null hypothesis. Edit the Sim1 node by clicking  icon and under the **Response Generation Info** tab, set **Prop. Under Treatment** to

$1.3 \times 0.125 = 0.1625$. Click **Simulate** button.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Ratio of Proportions

| Test Parameters | |
|-----------------------------------|----------------------------|
| Simulation ID | Sim2 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 2640 |
| Test Statistic | Score (Farrington Manning) |
| Variance | Empirical Estimate |
| Noninf. Margin (p_0) | 1.3 |
| Avg. Power at Termination | 0.05 |
| Response Generation Parameters | |
| Prop. under Control (π_c) | 0.125 |
| Prop. under Treatment (π_t) | 0.163 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

☰ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 880 |
| 2 | 1760 |
| 3 | 2640 |
| Average | 2626.536 |

☰ Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 880 | -3.2 | 5 | 5 | 0.050% |
| 2 | 1760 | -2.141 | 143 | 143 | 1.430% |
| 3 | 2640 | -1.695 | 352 | 9852 | 98.520% |
| Total | | | 500 | 10000 | |
| % | | | 5.000% | | |

Simulation Seed and Elapsed Time

Starting Seed: 2148031
Total Number of Simulations: 10000
Elapsed Time: 00:00:05

This time the rejection rate is only 5% (up to Monte Carlo accuracy), as we would expect under the null hypothesis. You may experiment in this manner with different values of π_c and π_t and observe the rejection rates look by look as well as averaged over all looks.

24.3.3 Interim Monitoring

Select Design4 in the **Library**, and click the **IM** icon from the Library toolbar. Alternatively, right-click on Design4 and select **Interim Monitoring**. The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs. The bottom section contains four charts, each with a corresponding table to its right. These charts provide graphical and numerical descriptions of the progress of the clinical trial and are useful tools for decision making by a data monitoring committee.

Suppose that the trial is first monitored after accruing 125 subjects on each treatment arm, with 15 responses on the control arm and 13 responses on the treatment arm.

Click on the **Enter Interim Data** icon to invoke the **Test Statistic Calculator**. In the box next to **Cumulative Sample Size** enter 250. Enter -0.052 in the box next to **Estimate of δ** . In the box next to **Std. Error of δ** enter 0.046617. Next click **Recalc**.

24 Binomial Non-Inferiority Two-Sample

The test statistic is computed to be -1.115. This value for the test statistic was obtained by substituting the observed sample sizes and responses into equation (24.10).

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Sample Size and Responses

| | Control | Treatment |
|-------------------------|---------|-----------|
| Cumulative Sample Size: | N/A | N/A |
| Cumulative Response: | N/A | N/A |

Cumulative Sample Size: 250

Input for Binomial end point

Estimate of δ : -0.052

$\delta = (\pi_t - \rho_0 * \pi_c)$

Standard Error of Estimate of δ : 0.046617

Output

$\delta - \delta_0$: -0.052

Test Statistic: -1.115

Recalc OK Cancel

Upon clicking the **OK** button, East will produce the interim monitoring report shown below.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 95% RCI for δ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|----------------------|-------|------------------|------|------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | 0.095 | 250 | -1.115 | -0.052 | 0.047 | -6.262 | 0.24 | -1 | 0.643 | 0.98 | 0.738 |
| 2 | | | | | | | | | | | |
| 3 | | | | | | | | | | | |

The stopping boundary for declaring non-inferiority is -2.929 whereas the value of the test statistic is only -1.115. Thus the trial should continue. This conclusion is also

supported by the upper confidence bound on

$$\delta = \pi_t - \rho_0 \pi_c$$

which at present equals 0.085. A necessary and sufficient condition for the stopping boundary to be crossed, and non-inferiority demonstrated thereby, is for this upper confidence bound to be less than zero.

Suppose that the next interim look occurs after accruing 250 patients on each arm with 31 responses on the control arm and 22 responses on the treatment arm. Click on the second row in the table in the upper section. Then click the **Enter Interim Data** icon. In the box next to **Cumulative Sample Size** enter 500. Enter -0.0732 in the box next to **Estimate of δ** . In the box next to **Std. Error of δ** enter 0.032486. Next click **Recalc**. Notice that the test statistic is computed to be -2.253.

The screenshot shows a 'Test Statistic Calculator' dialog box with the following fields and values:

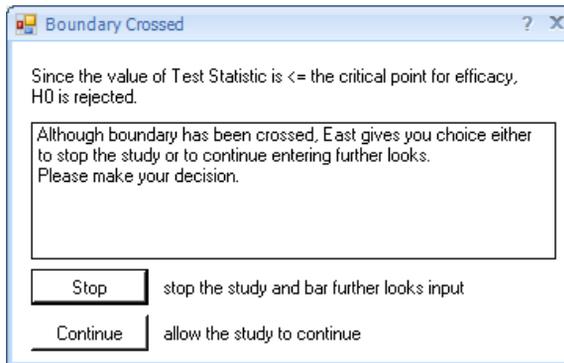
| Sample Size and Responses | | |
|--|----------|-----------|
| | Control | Treatment |
| Cumulative Sample Size: | N/A | N/A |
| Cumulative Response: | N/A | N/A |
| Cumulative Sample Size: | 500 | |
| Input for Binomial end point | | |
| Estimate of δ : | -0.0732 | |
| $\delta = (\pi_t - \rho_0 * \pi_c)$ | | |
| Standard Error of Estimate of δ : | 0.032486 | |
| Output | | |
| $\delta - \delta_0$ | -0.073 | |
| Test Statistic: | -2.253 | |

Buttons: Recalc, OK, Cancel

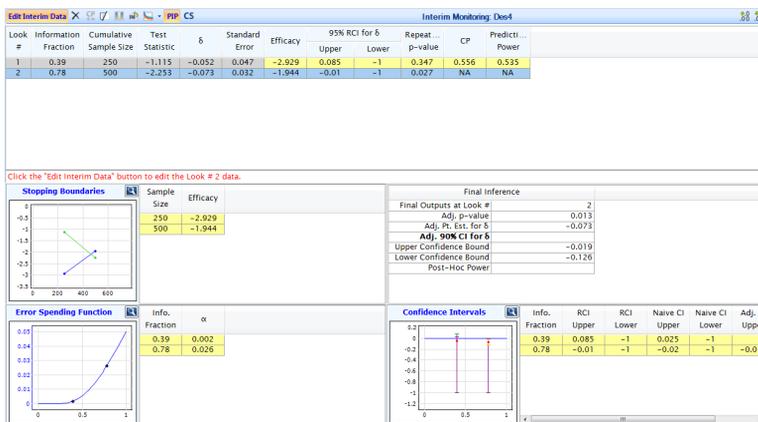
Click the **OK** button. This time the stopping boundary for declaring non-inferiority is

24 Binomial Non-Inferiority Two-Sample

crossed. The following message box appears.



Click the **Stop** button to stop the study. The analysis results are shown below. Notice that the upper confidence bound of the repeated confidence interval for δ excludes zero.



In the **Final Inference** Table in the bottom portion of the IM worksheet, East also provides a p-value, confidence interval and median unbiased point estimate for δ using stage-wise ordering of the sample space as described in Jennison and Turnbull (2000, page 179). The upper confidence bound for δ based on the stage-wise method likewise excludes zero.

24.4 Odds Ratio Test

Let π_t and π_c denote the two binomial probabilities associated with the treatment (t)

and the control (c). Let the difference between the two treatment arms be captured by the odds ratio

$$\psi = \frac{\pi_t/(1 - \pi_t)}{\pi_c/(1 - \pi_c)} = \frac{\pi_t(1 - \pi_c)}{\pi_c(1 - \pi_t)} .$$

The null hypothesis is specified as

$$H_0: \psi = \psi_0$$

and is tested against one-sided alternative hypotheses. If the occurrence of a response denotes patient benefit rather than harm, then $\psi_0 > 1$ and the alternative hypothesis is

$$H_1: \psi > \psi_0 .$$

Conversely, if the occurrence of a response denotes patient harm rather than benefit, then $\psi_0 < 1$ and the alternative hypothesis is

$$H_1: \psi < \psi_0 .$$

For any given π_c , the sample size is determined by the desired power at a specified value $\psi = \psi_1$. A common choice is $\psi_1 = 1$ (or equivalently $\pi_t = \pi_c$), but East permits you to power the study at any value of ψ_1 which is consistent with the choice of H_1 .

Let $\hat{\pi}_{tj}$ and $\hat{\pi}_{cj}$ denote the estimates of π_t and π_c based on n_{tj} and n_{cj} observations from the experimental and control treatments, respectively, up to and including j -th look, $j = 1, \dots, K$, where a maximum of K looks are to be made. It is convenient to express the treatment effect on the logarithmic scale as

$$\delta = \ln \psi . \quad (24.12)$$

The test statistic at the j th look is then defined as

$$Z_j = \frac{\hat{\delta}_j - \delta_0}{\text{se}(\hat{\delta}_j)} = \frac{\ln(\hat{\psi}_j) - \ln(\psi_0)}{\sqrt{\frac{1}{n_{tj}\hat{\pi}_{tj}(1-\hat{\pi}_{tj})} + \frac{1}{n_{cj}\hat{\pi}_{cj}(1-\hat{\pi}_{cj})}} . \quad (24.13)$$

24.4.1 Trial Design

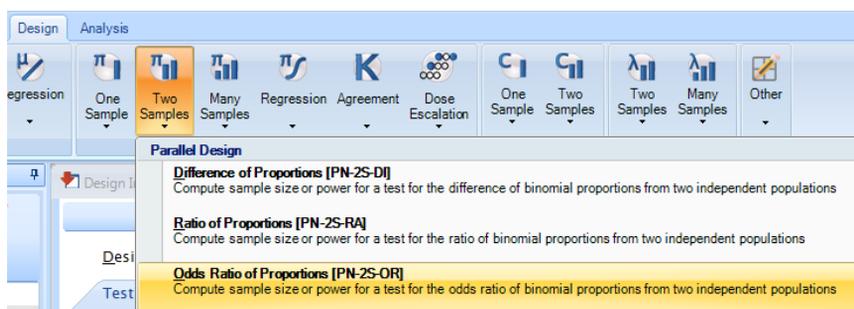
Suppose that the response rate for the control treatment is 90%, where higher response rates imply patient benefit. Assume that a claim of non-inferiority can be sustained if we can demonstrate statistically that the experimental treatment has a response rate of at least 80%. In other words the non-inferiority margin is

$$\psi_0 = \frac{0.8(1 - 0.9)}{0.9(1 - 0.8)} = 0.444 .$$

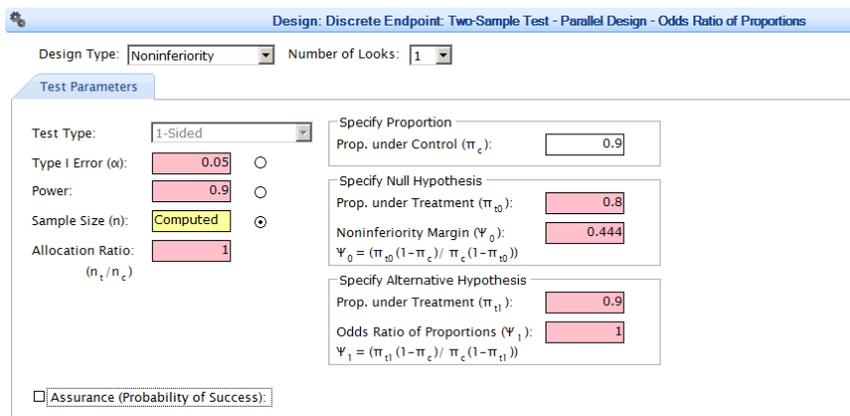
24 Binomial Non-Inferiority Two-Sample

The null hypothesis $H_0: \psi = 0.444$ is to be tested against the one-sided alternative $H_1: \psi > 0.444$. Suppose that we want to determine the sample size required to have power of 90% when $\pi_c = 0.9$ and $\psi_1 = 1$, i.e. $\pi_c = \pi_t$, using a test with a type-1 error rate of 0.05.

Single-Look Design Powered at $\psi = 1$ First we consider a study with only one look and equal sample sizes in the two groups. To begin click **Two Proportions** on the **Design** tab, and then click **Odds Ratio of Proportions**.



In the ensuing dialog box, next to **Trial**, select **Noninferiority** from the drop down menu. Choose the remaining design parameters as shown below.

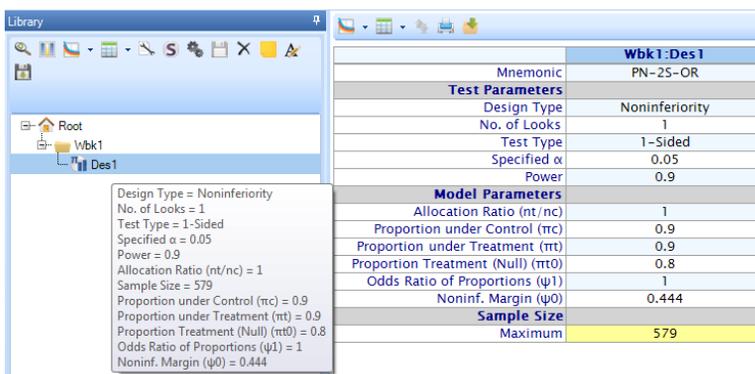


Now click **Compute**. The design is shown as a row in the Output Preview located in the lower pane of this window. This single-look design requires a combined total of

579 subjects from both treatments in order to attain 90% power.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | ttc | Prop. Treatment (Alt.) | Prop. Treatment (Null) | ψ_1 | ψ_0 |
|------|----------------|--------------|-----------|--------------------|-------|-------|-------------|-----|------------------------|------------------------|----------|----------|
| Des1 | Noninferiority | 1 | 1-Sided | 0.05 | 0.9 | 1 | 579 | 0.9 | 0.9 | 0.8 | 1 | 0.444 |

You can select this design by clicking anywhere along the row in the **Output Preview**. Some of the design details will be displayed in the upper pane, labeled **Compare Designs**. In the **Output Preview** toolbar, click the  icon to save this design to Workbook1 in the **Library**. If you hover the cursor over Design1 in the Library, a tooltip will appear that summarizes the input parameters of the design.



The screenshot shows the 'Library' window with a tree view containing 'Root', 'Wbk1', and 'Des1'. A tooltip is displayed over 'Des1' with the following text:

Design Type = Noninferiority
No. of Looks = 1
Test Type = 1-Sided
Specified α = 0.05
Power = 0.9
Allocation Ratio (nt/nc) = 1
Sample Size = 579
Proportion under Control (π_c) = 0.9
Proportion under Treatment (π_t) = 0.9
Proportion Treatment (Null) (π_{t0}) = 0.8
Odds Ratio of Proportions (ψ_1) = 1
Noninf. Margin (ψ_0) = 0.444

To the right, a table titled 'Wbk1:Des1' displays the parameters:

| Mnemonic | PN-25-OR |
|--|----------------|
| Test Parameters | |
| Design Type | Noninferiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.9 |
| Proportion under Treatment (π_t) | 0.9 |
| Proportion Treatment (Null) (π_{t0}) | 0.8 |
| Odds Ratio of Proportions (ψ_1) | 1 |
| Noninf. Margin (ψ_0) | 0.444 |
| Sample Size | |
| Maximum | 579 |

Three-Look Design Powered at $\psi = 1$ For the above study, suppose we wish to take up to two equally spaced interim looks and one final look at the accruing data, using the default Lan- DeMets (O’Brien-Fleming) stopping boundary. Create a new design by selecting Design1 in the **Library**, and clicking the  icon on the **Library** toolbar. Change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. A new tab **Boundary Info** will appear. Clicking on this tab will reveal the stopping boundary parameters. By default, the **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The left side contains details for the **Efficacy** boundary, and the right side contains details for the **Futility** boundary. By default, there is an efficacy boundary (to reject H0) selected, but no futility boundary (to reject H1). The **Boundary Family** specified is of the Spending Functions type. The default **Spending function** is the Lan-DeMets (Lan & DeMets, 1983), with **Parameter** OF (O’Brien-Fleming), which generates boundaries that are very similar, though not identical, to the classical stopping boundaries of

24 Binomial Non-Inferiority Two-Sample

O'Brien and Fleming (1979). Technical details of these stopping boundaries are available in Appendix F.

Design Type: Number of Looks:

Test Parameters **Boundary**

Efficacy
 Boundary Family:
 Spending Function:
 Parameter:
 Type I Error (α): 0.05

Futility
 Boundary Family:

Spacing of Looks Equal Unequal

Efficacy Boundary:  

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.333 | 0.001 | 3.200 |
| 2 | 0.667 | 0.016 | 2.141 |
| 3 | 1.000 | 0.050 | 1.695 |

Click the **Compute** button to generate output for Design2. With Design2 selected in the **Output Preview**, click the  icon to save Design2 to the **Library**. In the **Library**, select the rows for Design1 and Design2, by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the two designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 |
|--|----------------|----------------|
| Mnemonic | PN-2S-OR | PN-2S-OR |
| Test Parameters | | |
| Design Type | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Proportion under Control (π_c) | 0.9 | 0.9 |
| Proportion under Treatment (π_t) | 0.9 | 0.9 |
| Proportion Treatment (Null) (π_{t0}) | 0.8 | 0.8 |
| Odds Ratio of Proportions (ψ_1) | 1 | 1 |
| Noninf. Margin (ψ_0) | 0.444 | 0.444 |
| Boundary Parameters | | |
| Spacing of Looks | | Equal |
| Efficacy Boundary | | LD (OF) |
| Sample Size | | |
| Maximum | 579 | 590 |
| Expected Under H0 | | 586.647 |
| Expected Under H1 | | 457.036 |

Using three planned looks requires an up-front commitment of 590 subjects, a slight inflation over the single-look design which required 579 subjects. However, the

three-look design may result in a smaller sample size than that required for the single-look design, with an expected sample size of 457 subjects under the alternative hypothesis ($\pi_c = 0.9, \psi = 1$), and still ensures that the power is 90%.

Single-Look Design Powered at $\psi \neq 1$ Suppose that it is expected that the new treatment is a bit better than the control, but it is unnecessary and unrealistic to perform a superiority test. The required sample size for $\psi_1 = 1.333$, i.e. $\pi_t = 0.92308$, is determined. Create a new design by selecting Design1 in the

Library, and clicking the  icon on the **Library** toolbar. In the ensuing dialog box, change the design parameters as shown below.

Design Type: Noninferiority Number of Looks: 1

Test Parameters

Test Type: 1-Sided

Type I Error (α): 0.05

Power: 0.9

Sample Size (n): Computed

Allocation Ratio: 1
 (n_t/n_c)

Specify Proportion
 Prop. under Control (π_c): 0.9

Specify Null Hypothesis
 Prop. under Treatment (π_{t0}): 0.8
 Noninferiority Margin (Ψ_0): 0.444
 $\Psi_0 = (\pi_{t0}(1-\pi_c) / \pi_c(1-\pi_{t0}))$

Specify Alternative Hypothesis
 Prop. under Treatment (π_{t1}): 0.923
 Odds Ratio of Proportions (Ψ_1): 1.333
 $\Psi_1 = (\pi_{t1}(1-\pi_c) / \pi_c(1-\pi_{t1}))$

Assurance (Probability of Success):

Click the **Compute** button to generate output for Design3. With Design3 selected in the **Output Preview**, click the  icon. In the **Library**, select the rows for Design1, Design2, and Design3, by holding the Ctrl key, and then click the .

24 Binomial Non-Inferiority Two-Sample

icon. The upper pane will display the details of the three designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 | Wbk1:Des3 |
|--|----------------|----------------|----------------|
| Mnemonic | PN-25-OR | PN-25-OR | PN-25-OR |
| Test Parameters | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 | 1 |
| Test Type | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.9 | 0.9 | 0.9 |
| Model Parameters | | | |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Proportion under Control (mC) | 0.9 | 0.9 | 0.9 |
| Proportion under Treatment (mT) | 0.9 | 0.9 | 0.923 |
| Proportion Treatment (Null) (mT0) | 0.8 | 0.8 | 0.8 |
| Odds Ratio of Proportions (ψ_1) | 1 | 1 | 1.333 |
| Noninf. Margin (ψ_0) | 0.444 | 0.444 | 0.444 |
| Boundary Parameters | | | |
| Spacing of Looks | | Equal | |
| Efficacy Boundary | | LD (OF) | |
| Sample Size | | | |
| Maximum | 579 | 590 | 358 |
| Expected Under H0 | | 586.647 | |
| Expected Under H1 | | 457.036 | |

We observe that a single-look design powered at $\psi_1 = 1.333$ reduces the sample size considerably relative to the single-look design powered at $\psi_1 = 1$. The reduction in maximum sample size for the three-look design is approximately 38% $(=(579-358)/579)$. However, Design3 should be implemented after careful consideration, since its favorable operating characteristics are only applicable to the optimistic situation where $\psi_1 = 1.333$. If $\psi_1 < 1.33$, the power under Design3 decreases and may be too small to establish noninferiority, even if the true value > 1 , but is < 1.333 .

Three-Look Design Powered at $\psi \neq 1$ For the above study (Design3), suppose we wish to take up to two equally spaced interim looks and one final look at the accruing data, using the default Lan- DeMets (O'Brien-Fleming) stopping boundary. Create a new design by selecting Design3 in the **Library**, and clicking the  icon on the **Library** toolbar. In the ensuing dialog box, change the **Number of Looks** to 3. Click the **Compute** button to generate output for Design4.

| | Wbk1:Des1 | Wbk1:Des2 | Wbk1:Des3 | Wbk1:Des4 |
|--|----------------|----------------|----------------|----------------|
| Mnemonic | PN-25-OR | PN-25-OR | PN-25-OR | PN-25-OR |
| Test Parameters | | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 | 1 | 3 |
| Test Type | 1-Sided | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 | 0.05 | 0.05 |
| Power | 0.9 | 0.9 | 0.9 | 0.9 |
| Model Parameters | | | | |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 | 1 |
| Proportion under Control (mC) | 0.9 | 0.9 | 0.9 | 0.9 |
| Proportion under Treatment (mT) | 0.9 | 0.9 | 0.923 | 0.923 |
| Proportion Treatment (Null) (mT0) | 0.8 | 0.8 | 0.8 | 0.8 |
| Odds Ratio of Proportions (ψ_1) | 1 | 1 | 1.333 | 1.333 |
| Noninf. Margin (ψ_0) | 0.444 | 0.444 | 0.444 | 0.444 |
| Boundary Parameters | | | | |
| Spacing of Looks | | Equal | | Equal |
| Efficacy Boundary | | LD (OF) | | LD (OF) |
| Sample Size | | | | |
| Maximum | 579 | 590 | 358 | 365 |
| Expected Under H0 | | 586.647 | | 362.927 |
| Expected Under H1 | | 457.036 | | 282.621 |

Using three planned looks requires an up-front commitment of 365 subjects, a small inflation over the single-look design which required 358 subjects. However, the three-look design may result in a smaller sample size than that required for the single-look design, with an expected sample size of 283 subjects under the alternative hypothesis ($\pi_c = 0.9, \psi = 1.333$), and still ensures that the power is 90%.

24.4.2 Trial Simulation

You can simulate Design4 by selecting Design4 in the **Library** and clicking on the  icon. Try different choices for the simulation parameters to verify the operating characteristics of the study. First, we verify the results under the alternative hypothesis at which the power is to be controlled, namely $\pi_c = 0.9$ and $\pi_t = 0.92308$. Under the **Response Generation Info** tab set **Prop. Under Control** to 0.9 and **Prop. Under Treatment** to 0.92308.

Number of Looks:

Test Parameters Response Generation Simulation Controls

Specify Proportion

Prop. under Control (π_c):

Prop. under Treatment (π_t):

Click **Simulate** button. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled Simulation1. Select Simulation1 in the **Output Preview**. Note that some of the design details will be displayed in the upper pane, labeled **Compare Designs**. Click the  icon to save it to the **Library**. Double-click on Simulation1 in the **Library**. The simulation output details will be

24 Binomial Non-Inferiority Two-Sample

displayed.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Odds Ratio of Proportions

| Test Parameters | |
|--------------------------------------|----------------|
| Simulation ID | Sim1 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 365 |
| Noninf. Margin (ψ_0) | 0.444 |
| Avg. Power at Termination | 0.892 |
| Response Generation Parameters | |
| Prop. under Control (π_c) | 0.9 |
| Prop. under Treatment (π_{t1}) | 0.923 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 122 |
| 2 | 243 |
| 3 | 365 |
| Average | 288.118 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | | Efficacy | Count |
| 1 | 122 | 3.195 | 270 | 270 | 2.700% |
| 2 | 243 | 2.143 | 5764 | 5764 | 57.640% |
| 3 | 365 | 1.695 | 2888 | 3966 | 39.660% |
| Total | | | 8922 | 10000 | |
| % | | | 89.220% | | |

Simulation Seed and Elapsed Time

Starting Seed: 1202067
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:03

We see here that the power is approximately 90%.

Now let's consider the impact if the sample size was determined assuming $\pi_c = 0.9, \psi_1 = 1.333$ when the true values are $\pi_c = 0.9$ and $\psi_1 = 1$. Under the **Response Generation Info** tab set **Prop. Under Treatment** to 0.9. Click **Simulate**

button.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Odds Ratio of Proportions

| Test Parameters | |
|-----------------------------------|----------------|
| Simulation ID | Sim2 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 365 |
| Noninf. Margin (ψ_0) | 0.444 |
| Avg. Power at Termination | 0.741 |
| Response Generation Parameters | |
| Prop. under Control (π_0) | 0.9 |
| Prop. under Treatment (π_1) | 0.9 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

☉ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 122 |
| 2 | 243 |
| 3 | 365 |
| Average | 314.64 |

☉ Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 122 | 3.195 | 136 | 136 | 1.360% |
| 2 | 243 | 2.143 | 3857 | 3857 | 38.570% |
| 3 | 365 | 1.695 | 3419 | 6007 | 60.070% |
| Total | | | 7412 | 10000 | |
| % | | | 74.120% | | |

Simulation Seed and Elapsed Time

Starting Seed: 1426296
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:03

This results in a power of approximately 74%. From this we see that if that optimistic choice is incorrect, then the power to establish noninferiority has decreased to a possibly unacceptable value of 74%.

24.4.3 Interim Monitoring

Select Design4 in the **Library**, and click the **IM** icon from the Library toolbar. Alternatively, right-click on Design4 and select **Interim Monitoring**. The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs. The bottom section contains four charts, each with a corresponding table to its right. These charts provide graphical and numerical descriptions of the progress of the clinical trial and are useful tools for decision making by a data monitoring committee.

Suppose that the trial is first monitored after accruing 60 subjects on each treatment arm, with 50 responses on the control arm and 52 responses on the treatment arm.

Click on the **Enter Interim Data** icon to invoke the **Test Statistic Calculator**. In the box next to **Cumulative Sample Size** enter 120. Enter 0.264231 in the box next to **Estimate of δ** . In the box next to **Std. Error of δ** enter 0.514034. Next click **Recalc**. Notice that the test statistic is computed to be 2.092. This value for the test statistic was

24 Binomial Non-Inferiority Two-Sample

obtained by substituting the observed sample sizes and responses into equation (24.13).

Upon clicking the **OK** button, East will produce the interim monitoring report shown below.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | 95% RCI for γ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|------------------|--------------------------------|----------|----------------------|-------|------------------|-------|------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | 0.329 | 120 | 2.092 | 0.264 | 0.514 | 3.225 | Infinity | 0.248 | 0.176 | 0.993 | 0.922 |
| 2 | | | | | | | | | | | |
| 3 | | | | | | | | | | | |

Note - Click on  icon to hide or unhide the columns of your interest.

The critical value is 3.22, and since the observed value of the test statistic (24.13) is less than this value, the null hypothesis cannot be rejected. Therefore, noninferiority cannot as yet be concluded.

Suppose that the second look is made after accruing 120 subjects on each treatment

arm, with 112 responses on the control arm and 115 responses on the treatment arm. Click on the second row in the table in the upper section. Then click the **Enter Interim Data** icon. In the box next to **Cumulative Sample Size** enter 240. Enter 1.43848 in the box next to **Estimate of δ** . In the box next to **Std. Error of δ** enter 0.801501. Next click **Recalc**. Notice that the test statistic is computed to be 2.808. This value for the test statistic was obtained by substituting the observed sample sizes and responses into equation (24.13).

Test Statistic Calculator

Editing Look #2

Set Current Look as Last

Sample Size and Responses

| | Control | Treatment |
|-------------------------|---------|-----------|
| Cumulative Sample Size: | N/A | N/A |
| Cumulative Response: | N/A | N/A |

Cumulative Sample Size:

Input for Binomial end point

Estimate of δ :

$\delta = \ln(\pi_t(1 - \pi_c)/(\pi_c(1 - \pi_t)))$

Standard Error of Estimate of δ :

Output

Estimate of $\delta - \delta_0$:

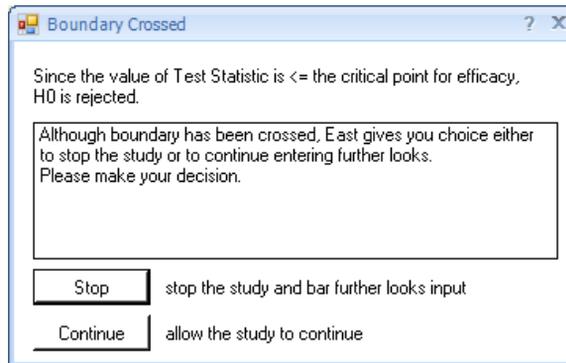
Test Statistic:

Recalc OK Cancel

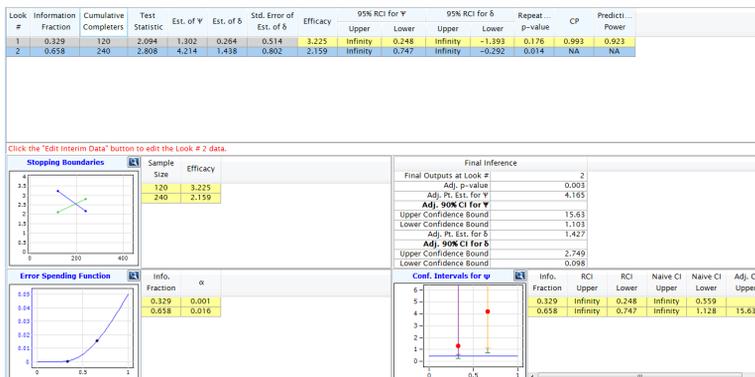
Click the **OK** button. This time the stopping boundary for declaring non-inferiority is

24 Binomial Non-Inferiority Two-Sample

crossed. The following message box appears.



Click the **Stop** button to stop the study. The analysis results are shown below.



The null hypothesis is rejected and we conclude that the treatment is noninferior to the control. In the **Final Inference** Table in the bottom portion of the IM worksheet, East also provides a stage-wise adjusted p-value, median unbiased point estimate and confidence interval for ψ as described in Jennison and Turnbull (2000) and in Appendix C of the East user manual. In the present example the adjusted p-value is 0.003, the point estimate for ψ is $\exp(1.427) = 4.166$ and the upper 95% confidence bound for ψ is $\exp(0.098) = 1.103$.

25 Binomial Equivalence Two-Sample

25.1 Equivalence Test

In some experimental situations, it is desired to show that the response rates for the control and the experimental treatments are "close", where "close" is defined prior to the collection of any data. Examples of this include showing that an aggressive therapy yields a similar rate of a specified adverse event to the established control, such as the bleeding rates associated with thrombolytic therapy or cardiac outcomes with a new stent. Let π_c and π_t denote the response rates for the control and the experimental treatments, respectively, and let $\hat{\pi}_t$ and $\hat{\pi}_c$ denote the estimates of π_t and π_c based on n_t and n_c observations from the experimental and control treatments. Furthermore, let

$$\delta = \pi_t - \pi_c, \tag{25.1}$$

which is estimated by

$$\hat{\delta} = \hat{\pi}_t - \hat{\pi}_c. \tag{25.2}$$

Finally, let the variance of $\hat{\delta}$ be

$$\sigma^2 = \frac{\pi_c(1 - \pi_c)}{n_c} + \frac{\pi_t(1 - \pi_t)}{n_t}, \tag{25.3}$$

which is estimated by

$$\hat{\sigma}^2 = \frac{\hat{\pi}_c(1 - \hat{\pi}_c)}{n_c} + \frac{\hat{\pi}_t(1 - \hat{\pi}_t)}{n_t}. \tag{25.4}$$

The null hypothesis $H_0: |\pi_t - \pi_c| = \delta_0$ is tested against the two-sided alternative hypothesis $H_1: |\pi_t - \pi_c| < \delta_0$, where $\delta_0 (> 0)$ is specified to define equivalence. Following Machin and Campbell (1987), we present the solution to this problem as a one-sided α -level test. The decision rule is to declare equivalence if

$$-\delta_0 + z_\alpha \hat{\sigma} \leq \hat{\pi}_t - \hat{\pi}_c \leq \delta_0 - z_\alpha \hat{\sigma}. \tag{25.5}$$

We see that decision rule (25.5) is the same as declaring equivalence if the $(1 - 2\alpha)$ 100% confidence interval for $\pi_t - \pi_c$ is entirely contained with the interval $(-\delta_0, \delta_0)$. The power or sample size are determined for a single-look study only. The extension to multiple looks is given in the next section. The sample size, or power, is determined at a specified difference $\pi_t - \pi_c$, denoted δ_1 , where $-\delta_0 < \delta_1 < \delta_0$. The probability of declaring equivalence depends on the true values of π_c and π_t . Based on the results of Machin and Campbell (1987), the required total sample size (N) is, for $n_t = rN$ and $n_c = (1 - r)N$,

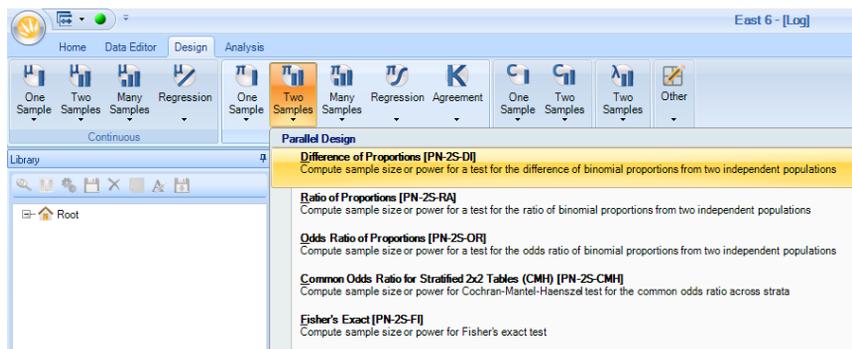
$$N = \frac{(z_\alpha + z_\beta)^2}{(\delta_0 - \delta_1)^2} \left(\frac{\pi_c(1 - \pi_c)}{1 - r} + \frac{(\pi_c + \delta_1)(1 - (\pi_c + \delta_1))}{r} \right). \tag{25.6}$$

25 Binomial Equivalence Two-Sample

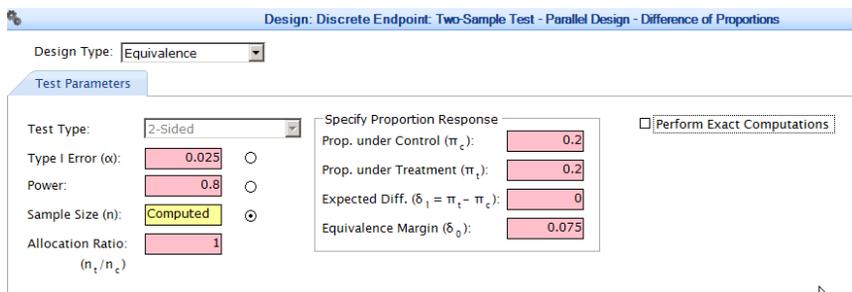
25.1.1 Trial Design

Consider the development of a new stent which is to be compared to the standard stent with respect to target vessel failure (acute failure, target vessel revascularization, myocardial infarction, or death) after one year. The standard stent has an assumed target vessel failure rate of 20%. Equivalence is defined as $\delta_0 = 0.075$. The sample size is to be determined with $\alpha = 0.025$ (one-sided) and power, i.e. probability of declaring equivalence, of $1 - \beta = 0.80$.

To begin click **Two Samples** on the **Design** tab, and then click **Difference of Proportions**.



Suppose that we want to determine the sample size required to have power of 80% when $\delta_1 = 0$. Enter the relevant parameters into the dialog box as shown below. In the drop down box next to **Trial Type** be sure to select **Equivalence**.



Click on the **Compute** button. The design is shown as a row in the Output Preview located in the lower pane of this window. The sample size required in order to achieve the desired 80% power is 1203 subjects.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | π_c | Prop. Treatment (Alt.) | δ_1 | Equivalence Margin | Expected Difference |
|------|-------------|--------------|-----------|--------------------|-------|-------|-------------|---------|------------------------|------------|--------------------|---------------------|
| Des1 | Equivalence | 1 | 2-Sided | 0.025 | 0.801 | 1 | 1203 | 0.2 | 0 | 0.075 | 0 | 0 |

You can select this design by clicking anywhere along the row in the **Output Preview**. If you double click anywhere along the row in the **Output Preview** some of the design details will be displayed in the upper pane, labeled **Output Summary**.

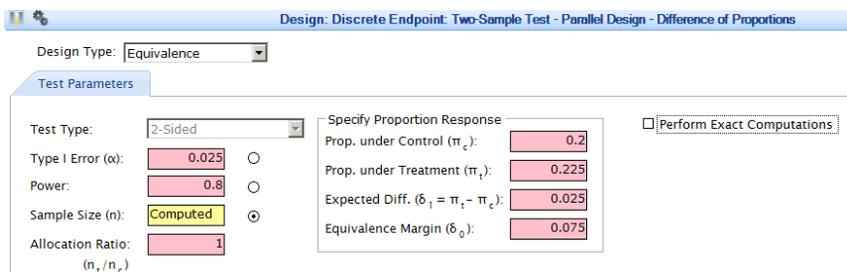
| Des1 | |
|--|-------------|
| Mnemonic | PN-2S-DI |
| Test Parameters | |
| Design Type | Equivalence |
| No. of Looks | 1 |
| Test Type | 2-Sided |
| Specified α | 0.025 |
| Power | 0.801 |
| Model Parameters | |
| Proportion under Control (π_c) | 0.2 |
| Proportion under Treatment (π_t) | 0.2 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0 |
| Equivalence Margin(δ_0) | 0.075 |
| Expected Difference($\pi_t - \pi_c$) | 0 |
| Allocation Ratio (nt/nc) | 1 |
| Sample Size | |
| Maximum | 1203 |

In the **Output Preview** toolbar, click the  icon to save this design to Workbook1 in the **Library**. If you hover the cursor over Des1 in the Library, a tooltip will appear that summarizes the input parameters of the design.

If the assumed difference δ_1 is not zero, it is more difficult to establish equivalence, in the sense that the power is lower and thus the required sample size is larger. Consider $\delta_1 = 0.025$, so that the new stent increases the rate to 22.5%. Create a new design

25 Binomial Equivalence Two-Sample

Des2 by selecting Des1 in the **Library**, and clicking the  icon on the **Library** toolbar. Change the value of **Expected Diff.** from 0 to 0.025 as shown below.



Click on the **Compute** button. The design is shown as a row in the Output Preview located in the lower pane of this window. With Design2 selected in the **Output Preview**, click the  icon. In the **Library**, select the rows for Des1 and Des2, by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the two designs side-by-side:

| | Des1 | Des2 |
|---|-------------|-------------|
| Mnemonic | PN-2S-DI | PN-2S-DI |
| Test Parameters | | |
| Design Type | Equivalence | Equivalence |
| No. of Looks | 1 | 1 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.801 | 0.8 |
| Model Parameters | | |
| Proportion under Control (π_c) | 0.2 | 0.2 |
| Proportion under Treatment (π_t) | 0.2 | 0.225 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0 | 0.025 |
| Equivalence Margin (δ_0) | 0.075 | 0.075 |
| Expected Difference ($\pi_t - \pi_c$) | 0 | 0.025 |
| Allocation Ratio (n_t/n_c) | 1 | 1 |
| Sample Size | | |
| Maximum | 1 203 | 21 20 |

This single-look design requires a combined total of 2120 subjects from both

treatments in order to attain 80% power.

Consider $\delta_1 = -0.025$, so that the new stent decreases the rate to 17.5%. Create a new design, as above, and change the value of **Expected Diff.** to -0.025 . Click the **Compute** button to generate the output for Des3. With Des3 selected in the **Output Preview**, click the  icon. In the **Library**, select the nodes for Des1, Des2, and Des3 by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the three designs side-by-side:

| | Des1 | Des2 | Des3 |
|---|-------------|-------------|-------------|
| Mnemonic | PN-2S-DI | PN-2S-DI | PN-2S-DI |
| Test Parameters | | | |
| Design Type | Equivalence | Equivalence | Equivalence |
| No. of Looks | 1 | 1 | 1 |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.025 | 0.025 | 0.025 |
| Power | 0.801 | 0.8 | 0.8 |
| Model Parameters | | | |
| Proportion under Control (π_c) | 0.2 | 0.2 | 0.2 |
| Proportion under Treatment (π_t) | 0.2 | 0.225 | 0.175 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0 | 0.025 | -0.025 |
| Equivalence Margin (δ_0) | 0.075 | 0.075 | 0.075 |
| Expected Difference ($\pi_t - \pi_c$) | 0 | 0.025 | -0.025 |
| Allocation Ratio (n_t/n_c) | 1 | 1 | 1 |
| Sample Size | | | |
| Maximum | 1203 | 2120 | 1940 |

Des3 yields a required total sample size of 1940 subjects. This asymmetry is due to the fact that the variance is smaller for values of $\pi_c + \delta_1$ further from 0.5.

25.1.2 Extension to Multiple Looks

Although the details presented in the previous section are related to a single-look design only, these results can be used to extend the solution to allow for multiple equally-spaced looks. We can use the General Design Module to generalize the solution to this problem to the study design with multiple looks. Details are given in Chapters 60 and 59.

Let $\hat{\pi}_{tj}$ and $\hat{\pi}_{cj}$ denote the estimates of π_t and π_c based on n_{tj} and n_{cj} observations from the experimental and control treatments, respectively, up to and including the j -th look, $j = 1, \dots, K$, where a maximum of K looks are to be used. Let $n_j = n_{cj} + n_{tj}$ and

$$\hat{\delta}_j = \hat{\pi}_{tj} - \hat{\pi}_{cj} \tag{25.7}$$

25 Binomial Equivalence Two-Sample

denote the estimate of δ , given by (25.1), and let

$$\hat{\sigma}_j^2 = \frac{\hat{\pi}_{c_j}(1 - \hat{\pi}_{c_j})}{n_{c_j}} + \frac{\hat{\pi}_{t_j}(1 - \hat{\pi}_{t_j})}{n_{t_j}} \quad (25.8)$$

denote the estimate of σ^2 , given by (25.3), using the data available at the j -th look.

At the j -th look, the inference is based on

$$Z_j = \frac{\hat{\delta}_j}{\hat{\sigma}_j}. \quad (25.9)$$

Let

$$\eta = \delta \sqrt{I_{max}},$$

where I_{max} is described in Chapter 59. Let $t_j = n_j/n_{max}, j = 1, \dots, K$. Then, using the multivariate normal approximation to the distribution of Z_1, \dots, Z_K , with the expected value of Z_j equal to $t_j^{1/2}\eta$ and the variance of Z_j equal to 1, the $(1 - \alpha)100\%$ repeated confidence intervals for η are

$$\left(\frac{Z_j + C_{Lj}}{t_j^{1/2}}, \frac{Z_j + C_{Uj}}{t_j^{1/2}} \right), \quad (25.10)$$

where C_{Lj} and C_{Uj} are the values specified by the stopping boundary. The corresponding $(1 - \alpha)100\%$ repeated confidence intervals for δ are

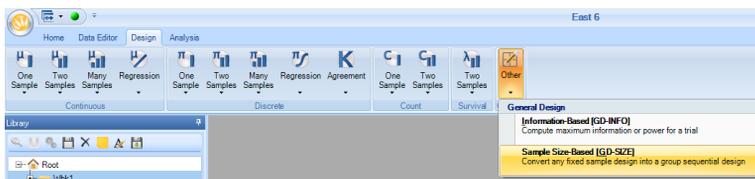
$$(\delta_j + C_{Lj}, \delta_j + C_{Uj}). \quad (25.11)$$

Using the General Design Module, East provides these repeated confidence intervals for η . By considering the decision rule (25.5) as declaring equivalence if the $(1 - 2\alpha)$ 100% confidence interval for $\pi_t - \pi_c$ is entirely contained with the interval $(-\delta_0, \delta_0)$, we generalize the decision rule to a multiple-look design by concluding equivalence and stopping the study the first time one of the repeated $(1 - 2\alpha)$ 100% confidence intervals for η is entirely contained within the interval $(-\eta_{0j}, \eta_{0j})$, where

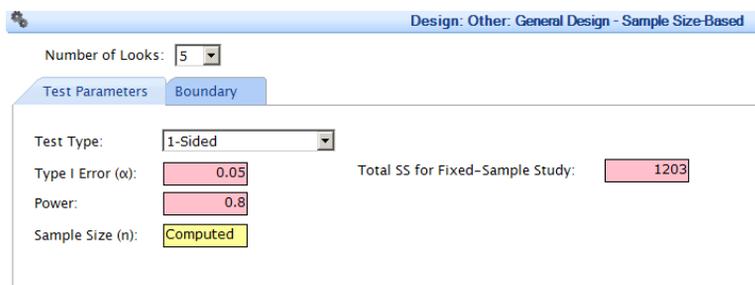
$$\eta_{0j} = \delta_0/t_j^{1/2}\hat{\sigma}_j.$$

Consider Design1 (i.e. $\pi_c = 0.20, \delta_0 = 0.075$, and $\delta_1 = 0$). As we saw above, a total of 1203 subjects are required for decision rule (25.5) to have power of 80% of declaring equivalence, using a 95% confidence interval.

To begin click on the **Other Designs** on the **Design** tab and then click **Sample Size-Based**.



Enter the parameters as shown below. For the **Sample Size for Fixed-Sample Study** enter **1203**, the value obtained from Des1. Also, be sure to set the **Number of Looks** to 5. Recall that the choice here is twice the (one-sided) value specified for the single-look design. The General Design Module is designed for testing the null hypothesis $H_{00}: \eta = 0$. Thus, the specified power of the test pertains to testing H_{00} and is not directly related to the procedure using the confidence interval. The expected sample sizes under H_0 and H_1 depend on the specified value of the power and pertain to the null hypothesis H_{00} and the corresponding alternative hypothesis $H_{10}: \eta \neq 0$ or a corresponding one-sided alternative. These expected sample sizes are not directly applicable to the equivalence problem of testing H_0 against H_1 .



Next click on the **Boundary Info** tab. The repeated confidence intervals for η depend on the choice of spending function boundaries. The sample size for this group sequential study also depends on the choice of the spending function, as well as the choice of the power. Although the boundaries themselves are not used in the decision rule, the width of the repeated confidence intervals for η are determined by the choice of the spending function. Here we will use the Lan- DeMets (O'Brien-Fleming)

25 Binomial Equivalence Two-Sample

stopping boundary, with the looks spaced equally apart, as shown below.

Number of Looks: 5

Test Parameters

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.05

Spacing of Looks: Equal Unequal

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.200 | 0.000 | 4.229 |
| 2 | 0.400 | 0.002 | 2.888 |
| 3 | 0.600 | 0.011 | 2.298 |
| 4 | 0.800 | 0.028 | 1.962 |
| 5 | 1.000 | 0.050 | 1.740 |

Boundary

Futility Boundary Family: None

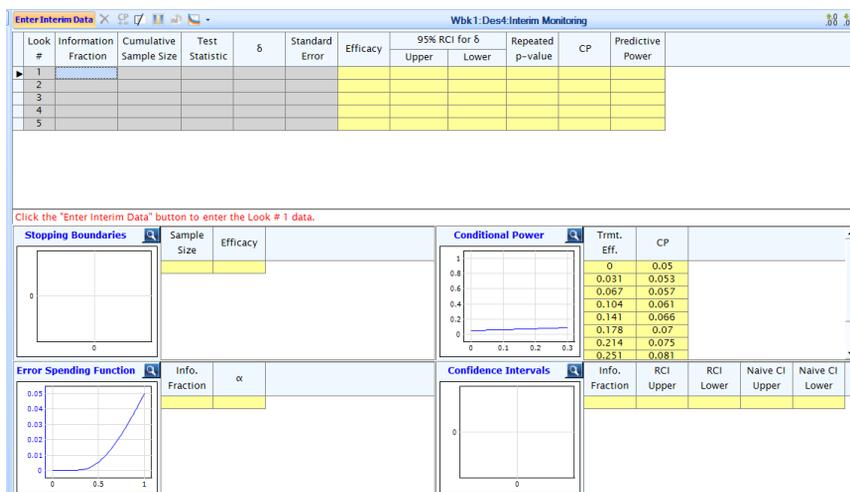
Efficacy Boundary: Z Scale

Click **Compute**. With Des4 selected in the **Output Preview**, click the  icon. In the **Library**, select the rows for Des1 and Des4, by holding the Ctrl key, and then click the  icon. The upper pane will display the summary details of the two designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des4 |
|--|------------------|------------------|
| Mnemonic | PN-2S-DI | GD-SIZE |
| Test Parameters | | |
| Design Type | Equivalence | |
| No. of Looks | | 5 |
| Test Type | 2-Sided | 1-Sided |
| Specified α | 0.025 | 0.05 |
| Power | 0.801 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | |
| Proportion under Control (π_C) | 0.2 | |
| Proportion under Treatment (π_T) | 0.2 | |
| Diff. in Prop. ($\pi_T - \pi_C$) | 0 | |
| Equivalence Margin(δ_0) | 0.075 | |
| Expected Difference($\pi_T - \pi_C$) | 0 | |
| Fixed SS | | 1203 |
| Boundary Parameters | | |
| Spacing of Looks | | Equal |
| Efficacy Boundary | | LD (OF) |
| Sample Size | | |
| Maximum | 1203 | 1246 |
| Expected Under H0 | | 1235.595 |
| Expected Under H1 | | 971.512 |

We see that the extension of Des1 to a five-look design requires a commitment of 1233 subjects, a small inflation over the sample size of 1203 subjects required for Des1.

Select Design4 in the **Library**, and click the **IM** icon from the Library toolbar. Alternatively, right-click on Design4 and select **Create IM Dashboard**. This will invoke the interim monitoring worksheet, from which the repeated 95% confidence intervals will be provided.



The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs. The bottom section contains four charts, each with a corresponding table to its right. These charts provide graphical and numerical descriptions of the progress of the clinical trial and are useful tools for decision making by a data monitoring committee.

We want to perform up to five looks, as data becomes available for every 200 subjects. Suppose that, after 200 subjects, $\hat{\pi}_{c,j} = 18/100 = 0.18$ and $\hat{\pi}_{t,j} = 20/100 = 0.2$. Then, from (25.2) and (25.4), the estimates of δ and the standard error of $\hat{\delta}$ are 0.02 and 0.0555. Click on the  icon to invoke the **Test Statistic Calculator**. Enter the appropriate values as shown below and click **Recalc**. Notice that the test statistic is

25 Binomial Equivalence Two-Sample

computed to be 0.357.

Next click **OK** . The following screen is shown.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 95% RCI for δ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|----------------------|--------|------------------|-------|------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | 0.161 | 200 | 0.36 | 0.02 | 0.056 | 4.754 | Infinity | -0.244 | 0.713 | 0.197 | 0.365 |
| 2 | | | | | | | | | | | |
| 3 | | | | | | | | | | | |
| 4 | | | | | | | | | | | |
| 5 | | | | | | | | | | | |

The first repeated 95% confidence interval for η is (-12.628, 14.402). Since this confidence interval is not contained in the interval (-3.357, 3.357), where

$$\eta_{01} = \frac{\delta_0}{t_1^{1/2} \hat{\sigma}_1} = \frac{0.075}{(0.162)^{1/2} (0.0555)} = 3.357,$$

we take a second look after 400 subjects. Click on the second row in the table in the upper section. Then click the  icon to invoke the **Test Statistic Calculator**. Suppose that $\hat{\pi}_{c_j} = 36/200 = 0.18$ and $\hat{\pi}_{t_j} = 38/200 = 0.19$. Then, from (25.2) and (25.4), the estimates of δ and the standard error of $\hat{\delta}$ are 0.01 and 0.0388. Enter these

values as shown below and click on the **Recalc** button.

Click on the **OK** button and the following values are presented in the interim monitoring worksheet.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 95% RCI for δ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|------|----------------|----------|---------------|--------|------------------|-------|------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | 0.161 | 200 | 0.36 | 0.02 | 0.056 | 4.754 | Infinity | -0.244 | 0.713 | 0.197 | 0.365 |
| 2 | 0.321 | 400 | 0.026 | 0.01 | 0.388 | 3.268 | Infinity | -1.258 | 0.765 | 0.022 | 0.125 |
| 3 | | | | | | | | | | | |
| 4 | | | | | | | | | | | |
| 5 | | | | | | | | | | | |

The second repeated 95% confidence interval for η is (-6.159, 7.064) is not contained in the interval (-3.396, 3.396), where

$$\eta_{02} = \frac{\delta_0}{t_2^{1/2} \hat{\sigma}_2} = \frac{0.075}{(0.324)^{1/2} (0.0388)} = 3.396,$$

so we cannot conclude equivalence. Continue the study and we take a third look after 600 subjects. Click on the third row in the table in the upper section. Then click the  icon to invoke the **Test Statistic Calculator**. Suppose that $\hat{\pi}_{c,j} = 51/300 = 0.17$ and $\hat{\pi}_{t,j} = 60/300 = 0.2$. Then, from (25.2) and (25.4), the estimates of δ and the standard error of $\hat{\delta}$ are 0.03 and 0.0317. Enter these values as

25 Binomial Equivalence Two-Sample

shown below and click on the **Recalc** button. The following screen is shown.

Click on the **OK** button and the following values are presented in the interim monitoring worksheet.

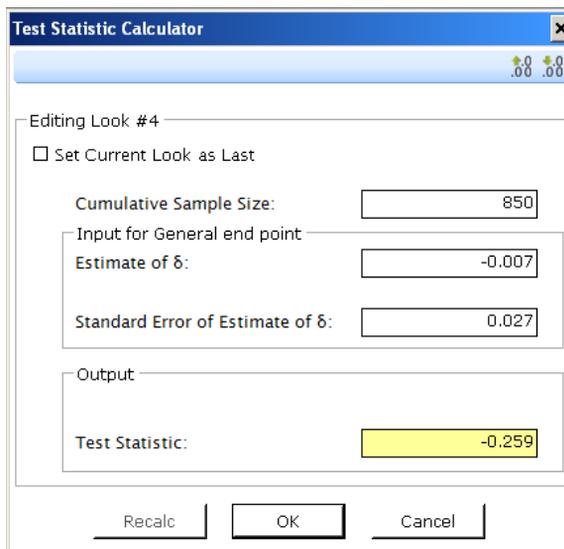
| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 95% RCI for δ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|----------------------|--------|------------------|-------|------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | 0.161 | 200 | 0.36 | 0.02 | 0.056 | 4.754 | Infinity | -0.244 | 0.713 | 0.197 | 0.365 |
| 2 | 0.321 | 400 | 0.026 | 0.01 | 0.388 | 3.268 | Infinity | -1.258 | 0.765 | 0.022 | 0.125 |
| 3 | 0.482 | 600 | 0.946 | 0.03 | 0.032 | 2.608 | Infinity | -0.053 | 0.382 | 0.316 | 0.369 |
| 4 | | | | | | | | | | | |
| 5 | | | | | | | | | | | |

The third repeated 95% confidence interval for η is (-2.965, 5.679) is not contained in the interval (-3.390, 3.390), where

$$\eta_{03} = \frac{\delta_0}{t_3^{1/2} \hat{\sigma}_3} = \frac{0.075}{(0.487)^{1/2} (0.0317)} = 3.390,$$

so we cannot conclude equivalence. Continue the study and we take a fourth look after 850 subjects. Click on the fourth row in the table in the upper section. Then click the  icon to invoke the **Test Statistic Calculator**. Suppose that $\hat{\pi}_{cj} = 91/450 = 0.2022$ and $\hat{\pi}_{tj} = 88/450 = 0.1956$. Then, from (25.2) and (25.4),

the estimates of δ and the standard estimate of δ are -0.007 and 0.027. Enter these values as shown below and click on the **Recalc** button. The following screen is shown.



Click on the **OK** button and the following values are presented in the interim monitoring worksheet.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 95% RCI for δ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|----------------------|--------|------------------|-------|------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | 0.161 | 200 | 0.36 | 0.02 | 0.056 | 4.754 | Infinity | -0.244 | 0.713 | 0.197 | 0.365 |
| 2 | 0.321 | 400 | 0.026 | 0.01 | 0.388 | 3.268 | Infinity | -1.258 | 0.765 | 0.022 | 0.125 |
| 3 | 0.482 | 600 | 0.946 | 0.03 | 0.032 | 2.608 | Infinity | -0.053 | 0.382 | 0.316 | 0.369 |
| 4 | 0.682 | 850 | -0.259 | -0.007 | 0.027 | 2.141 | Infinity | -0.065 | 0.873 | 0 | 0.001 |
| 5 | | | | | | | | | | | |

The fourth confidence interval is (-3.302, 2.678) is entirely contained in the interval (-3.346, 3.346), where

$$\eta_{04} = \frac{\delta_0}{t_4^{1/2} \hat{\sigma}_4} = \frac{0.075}{(0.689)^{1/2} (0.027)} = 3.346$$

and thus we conclude that the two treatments are equivalent. To express the results in terms of the δ , the final confidence interval for η can be transformed to a confidence interval for δ by multiplying the confidence limits by

$$t_4^{1/2} \hat{\sigma}_4 = (0.689)^{1/2} (0.027) = 0.0224,$$

25 *Binomial Equivalence Two-Sample*

resulting in a confidence interval for δ of $(-0.074, 0.060)$, which is entirely contained within the interval $(-0.075, 0.075)$.

26 Binomial Superiority n -Sample

26.1 Chi-Square for Specified Proportions in C Categories
 26.1.1 Trial Design

Let π_{0i} and π_{1i} for $i = 1, 2, \dots, C$ denote the response proportions under null and alternative hypotheses respectively where C denotes the number of categories. The null hypothesis states that the observed frequencies follow multinomial distribution with null proportions as probabilities. The test is performed for only two sided alternative. The sample size, or power, is determined for a specified value of the proportions which is consistent with the alternative hypothesis, denoted by π_{1i} .

Table 26.1: Table: Contingency Table

| Category\ Response | Cured | Not Cured |
|--------------------|----------|-----------|
| Age Group A | n_{11} | n_{21} |
| Age Group B | n_{12} | n_{22} |
| Age Group C | n_{13} | n_{23} |
| Marginal | $n_{1.}$ | $n_{2.}$ |

The null hypothesis is $H_0 : \pi_i = \pi_{0i}, i = 1, 2, 3, \dots, C$ and is tested against two-sided alternative.

The test statistic is given as,

$$\chi^2 = \sum_i \frac{(n_{1i} - \mu_i)^2}{\mu_i} \tag{26.1}$$

where $\mu_i = n_{1.}\pi_{0i}$

Let χ_0^2 be the observed value of χ^2 . For large samples, χ^2 has approximately Chi-squared distribution with d.f. $C - 1$. The p-value is approximated by $P(\chi_{c-1}^2 \geq \chi_0^2)$, where χ_{c-1}^2 denotes a Chi-squared random variable with d.f. = $C - 1$.

26.1.1 Trial Design

Consider the design of a single-arm trial with binary response - Cured and Not Cured. The responses for Cured population for three categories are of interest - Age group A, Age group B and Age group C. We wish to determine whether the proportion of cured in the three age groups are 0.25, 0.25, and 0.50 respectively. Thus it is desired to test $H_0 : \pi_A = 0.25, \pi_B = 0.25, \pi_C = 0.50$. We wish to design the trial with a two-sided

26 Binomial Superiority n-Sample

test that achieves 90% power at $H_1 : \pi_A = 0.3, \pi_B = 0.4, \pi_C = 0.3$ at level of significance 0.05.

Start East. Click **Design** tab, then click **Many Samples** in the **Discrete** group, and then click **Chi-Square Test of Specified Proportions in C Categories**.

In the upper pane of this window is the Input dialog box, which displays default input values.

Enter the **Number of Categories (C)** as 3. Under **Table of Proportion of Response**, enter the values of proportions under **Null Hypothesis** and **Alternative Hypothesis** for each category except the last one such that the sum of values in a row equals to 1. Enter the inputs as shown below and click **Compute**.

Number of Categories(C):

Test Parameters

Type I Error (α):

Power:

Sample Size (n):

| Table of Proportion of Response | | | |
|---------------------------------|------|------|------|
| Categories | Cat1 | Cat2 | Cat3 |
| Null Hypothesis | 0.25 | 0.25 | 0.5 |
| Alternative Hypothesis | 0.3 | 0.4 | 0.3 |

The design output will be displayed in the Output Preview, with the computed sample size highlighted in yellow. A total of 71 subjects must be enrolled in order to achieve 90% power under the alternative hypothesis. Besides sample size, one can also compute the power and the level of significance for this **Chi-Square Test of Specified Proportions in C Categories** study design.

| ID ▲ | Specified α | Power | Sample Size | No. of Categories | Δ |
|------|--------------------|-------|-------------|-------------------|------|
| Des1 | 0.05 | 0.903 | 71 | 3 | 0.18 |

You can select this design by clicking anywhere on the row in the **Output Preview**. If you click icon, some of the design details will be displayed in the upper pane. In the Output Preview toolbar, click the icon, to save this design to workbook Wbk1 in the **Library**. If you hover the cursor over Des1 in the **Library**, a tooltip will appear

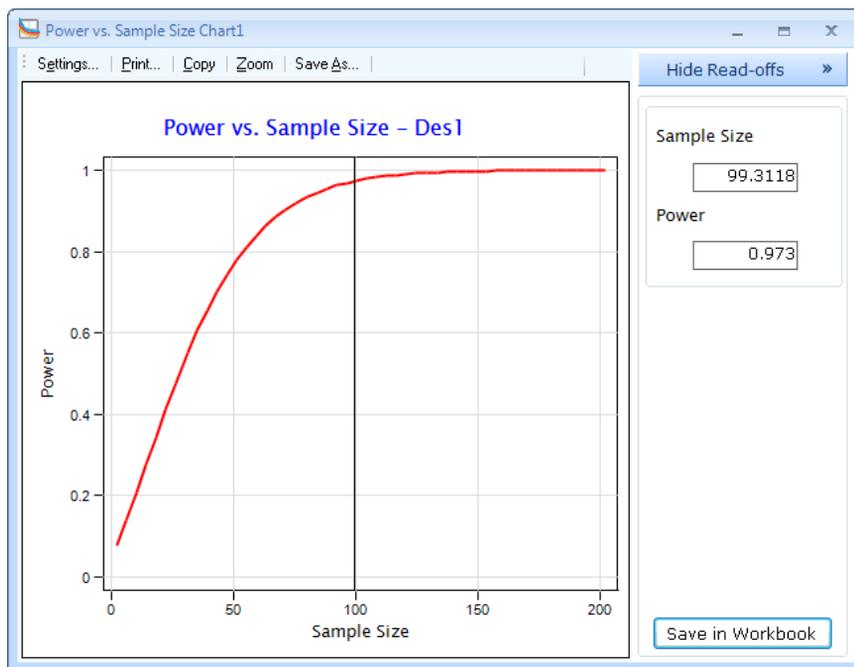
that summarizes the input parameters of the design.



| | Wbk1:Des1 |
|--------------------------|------------------|
| Mnemonic | PN-nS-CH1C |
| Test Parameters | |
| Specified α | 0.05 |
| Power | 0.903 |
| Model Parameters | |
| No. of Categories | 3 |
| Effect Size (Δ) | 0.18 |
| Sample Size | |
| Maximum | 71 |

With Des1 selected in the **Library**, click  icon on the **Library** toolbar, and then click **Power vs. Sample Size**. The resulting power curve for this design is shown. You can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...** For now, you may close the chart before continuing.

26 Binomial Superiority n -Sample



26.2 Two-Group Chi-square for Proportions in C Categories

26.2.1 Trial Design

Let π_{1j} and π_{2j} denote the response proportions of group 1 and group 2 respectively for the j -th category, where $j = 1, 2, \dots, C$.

The null hypothesis $H_0 : \pi_{1j} = \pi_{2j} \forall j = 1, 2, \dots, C$ is tested against the alternative hypothesis that for at least one j , π_{1j} differs from π_{2j} .

Table 26.2: Table: Contingency Table

| Categories \ Groups | Group 1 | Group 2 | Marginal |
|---------------------|----------|----------|----------|
| A | n_{11} | n_{21} | n_{01} |
| B | n_{12} | n_{22} | n_{02} |
| C | n_{13} | n_{23} | n_{03} |
| Marginal | n_{10} | n_{20} | n |

The test statistic is given as,

$$\chi^2 = \sum_{ij} \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}} \quad (26.2)$$

where $\mu_{ij} = \frac{n_{oj}n_{io}}{n}$, $j = 1, 2, \dots, C$ and $i = 1, 2$.

Let χ_0^2 be the observed value of χ^2 . For large samples, χ^2 has approximately Chi-squared distribution with d.f. $C - 1$. The p-value is approximated by $P(\chi_{C-1}^2 \geq \chi_0^2)$, where χ_{C-1}^2 denotes a Chi-squared random variable with d.f. = $C - 1$.

26.2.1 Trial Design

Suppose researchers want to investigate the relationship between different dose levels (level 1, level 2 and level 3) of a drug and the type adverse events (serious or not serious). The proportions who were treated with different dose levels will be compared using a Chi-square test. Suppose the expected proportions of patients for three different dose levels are 0.30, 0.35 and 0.35 where patients had no serious adverse events and the expected proportions are 0.20, 0.30 and 0.50 where patients had serious adverse events. We wish to design the trial with a two-sided test that achieves 90% power at level of significance 0.05.

Start East. Click **Design** tab, then click **Many Samples** in the **Discrete** group, and then click **Two-Group Chi-square for Proportions in C Categories**.

The Input dialog box, with default input values will appear in the upper pane.

Enter the **Number of Categories (C)** as 3. Under **Table of Proportion of Response**, enter the values of proportions under **Control** and **Treatment** for each category except the last one such that the sum of values in a row equals to 1. Enter the inputs as shown below and click **Compute**.

26 Binomial Superiority n-Sample

Number of Categories(C):

Test Parameters

Type I Error (α):

Power:

Sample Size (n):

Allocation Ratio:
(n_t/n_c)

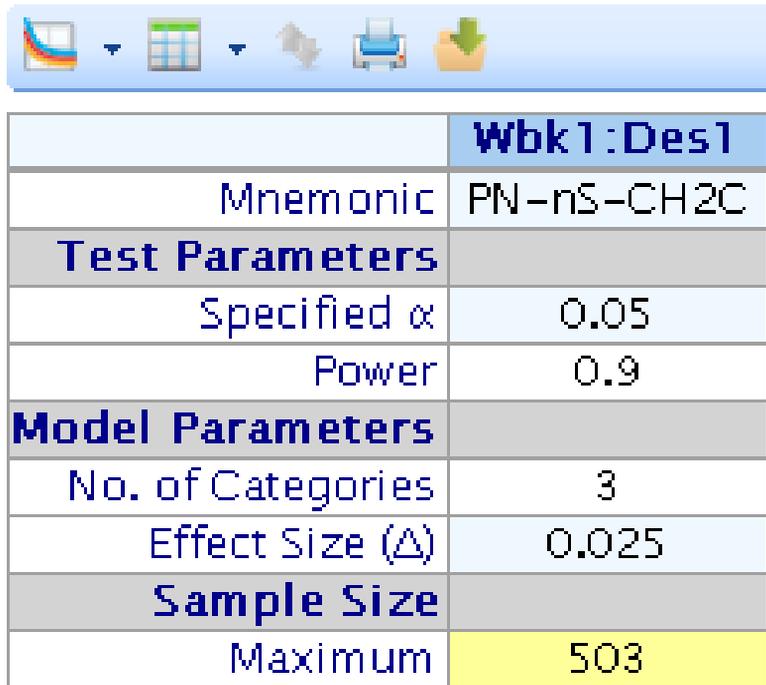
Table of Proportion of Response

| Categories | Cat1 | Cat2 | Cat3 |
|------------|------|------|------|
| Control | 0.3 | 0.35 | 0.35 |
| Treatment | 0.2 | 0.3 | 0.5 |

The design output will be displayed in the Output Preview, with the computed sample size highlighted in yellow. A total of 503 subjects must be enrolled in order to achieve 90% power under the alternative hypothesis. Besides sample size, one can also compute the power and the level of significance for this **Chi-Square Test of Specified Proportions in C Categories** study design.

| ID ▲ | Specified α | Power | Sample Size | No. of Categories | Δ |
|--|--------------------|-------|-------------|-------------------|-------|
|  Des1 | 0.05 | 0.9 | 503 | 3 | 0.025 |

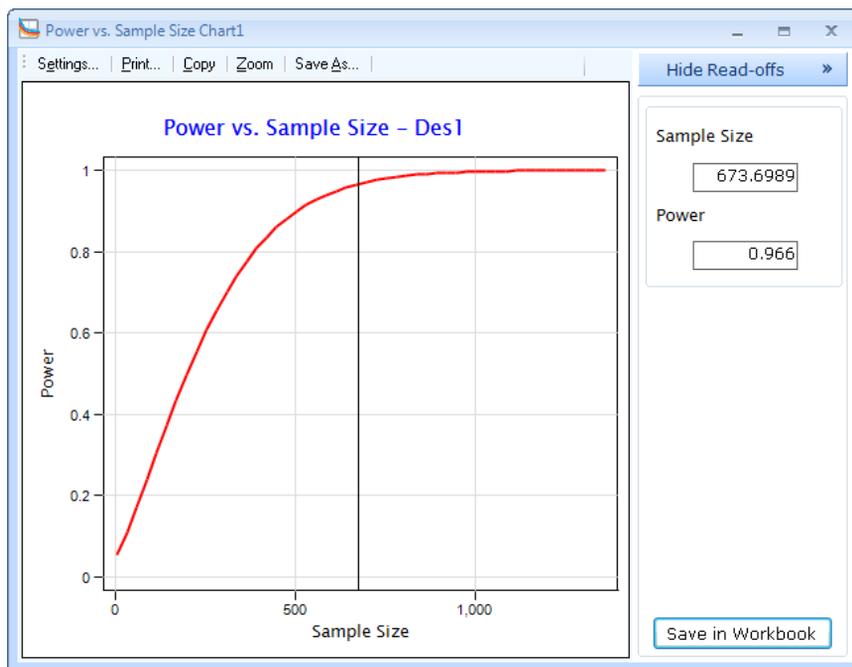
You can select this design by clicking anywhere on the row in the Output Preview. If you click  icon, some of the design details will be displayed in the upper pane. In the Output Preview toolbar, click  icon to save this design to Wbk1 in the **Library**. If you hover the cursor over Des1 in the **Library**, a tooltip will appear that summarizes the input parameters of the design.



| | Wbk1:Des1 |
|--------------------------|------------------|
| Mnemonic | PN-nS-CH2C |
| Test Parameters | |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| No. of Categories | 3 |
| Effect Size (Δ) | 0.025 |
| Sample Size | |
| Maximum | 503 |

With Des1 selected in the **Library**, click  icon on the **Library** toolbar, and then click **Power vs. Sample Size**. The resulting power curve for this design is shown. You can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...** For now, you may close the chart before continuing.

26 Binomial Superiority n -Sample



26.3 Nonparametric: Wilcoxon Rank Sum for Ordered Categorical Data

26.3.1 Trial Design

When we compare two treatments with respect to signs and symptoms associated with a disease, we may base the comparison on a variable that assesses degree of response or the degree of severity, using an ordinal categorical variable. For example, investigators may report the severity of an adverse event, or other abnormality, using a specified grading system or using a simple scale, such as "none", "mild", "moderate", or "severe". The latter rating scale might be used in an analgesia study to report the severity of pain. Although this four-point scale is often used and intuitively appealing, additional categories, such as "very mild" and "very severe", may be added. In other situations, the efficacy of the treatment is best assessed by the subject reporting response to therapy using a similar scale. The Wilcoxon test for ordered categories is a nonparametric test for use in such situations. East provides the power for a specified sample size for a single-look design using the constant proportional odds ratio model. Let π_{cj} and π_{tj} denote the probabilities for category j , $j = 1, 2, \dots, J$ for the control c and the treatment t respectively. Let $\gamma_{ci} = \sum_{j=1}^i \pi_{cj}$ and $\gamma_{ti} = \sum_{j=1}^i \pi_{tj}$. We assume that

$$\frac{\gamma_{ci}}{1-\gamma_{ci}} = e^{\psi} \frac{\gamma_{ti}}{1-\gamma_{ti}}, i = 1, 2, \dots, J - 1,$$

or, equivalently,

$$\psi = \ln(\gamma_{ci}) - \ln(1 - \gamma_{ci}) - (\ln(\gamma_{ti}) - \ln(1 - \gamma_{ti})) \quad (26.3)$$

We compare the two distributions by focusing on the parameter ψ . Thus we test the null hypothesis $H_0 : \psi = 0$ against the two-sided alternative $H_1 : \psi \neq 0$ or a one-sided alternative hypothesis $H_1 : \psi > 0$. East requires the specified value of ψ to be positive. Technical details can be found in Rabbee et al.,2003.

26.3.1 Trial Design

We consider here a placebo-controlled parallel-group study where subjects report the response to treatment as "none", "slight", "considerable", or "total". We expect that most of the subjects in the placebo group will report no response.

Start East. Click **Design** tab, then click **Many Samples** in the **Discrete** group, and then click **Non Parametric: Wilcoxon Rank Sum for Ordered Categorical Data**.

The Input dialog box, with default input values will appear in the upper pane.

We want to determine the power, using a two-sided test with a type-1 error rate of 0.05, with a total of 100 subjects, and equal sample sizes for the two groups. Enter **Number of Categories** as 4. We will use **User Specified for Specify Pop 1 Probabilities** and **Proportional Odd Model for Pop2 Probabilities** here. Click **Proportional Odds Model** radio button. A new field for **Shift** will appear. Enter 1.5 in this field. Based on the results of a pilot study, the values of 0.55, 0.3, 0.1, and 0.05 are used as **Pop 1 probabilities**. Enter the inputs as shown below and click **Compute**.

Number of Categories(C):

Test Parameters

Test Type:

Type I Error (α):

Power:

Sample Size (n):

Allocation Ratio:
(n_1/n_2)

Specify Pop 1 Probabilities

User Specified Equal Probabilities

Specify Pop 2 Probabilities

User Specified Proportional Odds Model

Log $\frac{Y_{2j}(1 - Y_{1j})}{Y_{1j}(1 - Y_{2j})} =$ Shift

Y_{ij} is the right tail cumulative prob. for Pop i upto and including categories j.

Specify Table of Probabilities

| | Pop 1 | Pop 2 |
|------|-------|-------|
| Cat1 | 0.55 | 0.214 |
| Cat2 | 0.3 | 0.344 |
| Cat3 | 0.1 | 0.251 |
| Cat4 | 0.05 | 0.191 |

26 Binomial Superiority n-Sample

The design output will be displayed in the Output Preview, with the computed power highlighted in yellow. This design results in a power of approximately 98% for a total sample size of 100 subjects.

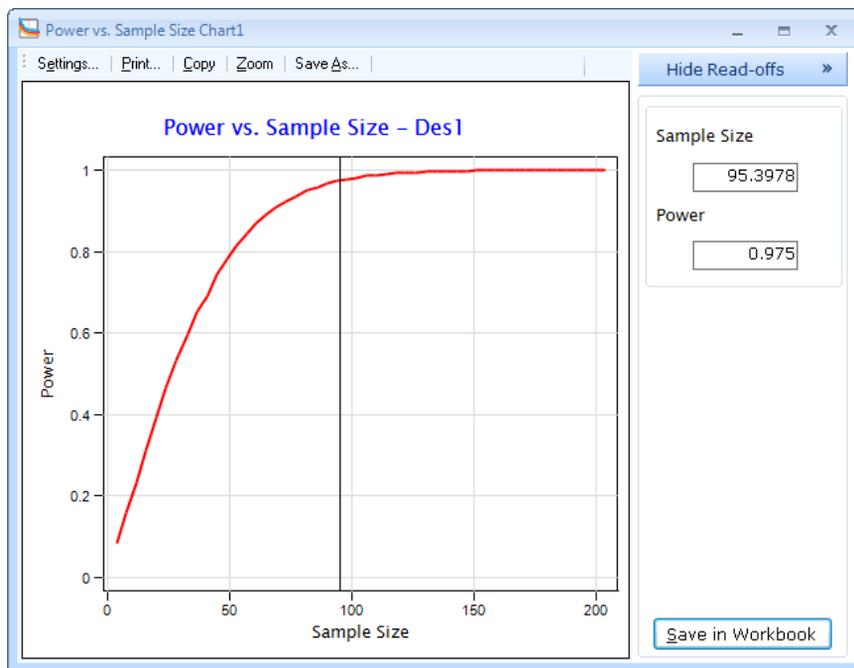
| ID ▲ | Test Type | Specified α | Power | Sample Size | nt/nc | No. of Categories | Shift |
|--|-----------|--------------------|-------|-------------|-------|-------------------|-------|
|  Des1 | 2-Sided | 0.05 | 0.98 | 100 | 1 | 4 | 1.5 |

You can select this design by clicking anywhere on the row in the Output Preview. If you click  icon, some of the design details will be displayed in the upper pane. In the Output Preview toolbar, click  icon, to save this design to workbook Wbk1 in the **Library**. If you hover the cursor over Des1 in the **Library**, a tooltip will appear that summarizes the input parameters of the design.



| Wbk1:Des1 | |
|--------------------------|-----------|
| Mnemonic | PN-nS-WRS |
| Test Parameters | |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.98 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| No. of Categories | 4 |
| shift | 1.5 |
| Sample Size | |
| Maximum | 100 |

With Des1 selected in the **Library**, click  icon, on the **Library** toolbar, and then click **Power vs. Sample Size**. The resulting power curve for this design is shown. You can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As....** For now, you may close the chart before continuing.



With such high power, a total sample size of 100 subjects may be an inefficient use of resources. We are willing to use a smaller sample size to achieve a lower power. Change the maximum sample size to 50 in the previous design. Leave all other values as defaults, and click **Compute**.

This design results in approximately 80% power using a total sample size of 50 subjects.

26 Binomial Superiority n -Sample

| | ID ▼ | Test Type | Specified α | Power | Sample Size | nt/nc | No. of Categories | Shift |
|---|------|-----------|--------------------|-------|-------------|-------|-------------------|-------|
|  | Des2 | 2-Sided | 0.05 | 0.8 | 50 | 1 | 4 | 1.5 |
|  | Des1 | 2-Sided | 0.05 | 0.98 | 100 | 1 | 4 | 1.5 |

26.4 Trend in R Ordered Binomial Proportions

26.4.1 Trial Design

In some experimental situations, there are several binomial distributions indexed by an ordinal variable and we want to examine changes in the probabilities of success as the levels of the indexing variable changes. Examples of this include the examination of a dose-related presence of a response or a particular side effect, dose-related tumorigenicity, or presence of fetal malformations relative to levels of maternal exposure to a particular toxin, such as alcohol, tobacco, or environmental factors.

The test for trend in R ordered proportions is based on the Cochran Armitage trend test. Let π_j denote the probability of interest for the j -th category of the ordinal variable, $j = 1, 2, \dots, R$ and let scores be denoted by $\omega_1, \omega_2, \dots, \omega_R$. It is assumed that the odds ratio relating to j -th category to the $(j - 1)$ -th category satisfies

$$\frac{\pi_j}{1 - \pi_j} = \psi^{\omega_j - \omega_{j-1}} \frac{\pi_{j-1}}{1 - \pi_{j-1}} \quad (26.4)$$

or equivalently,

$$\ln\left(\frac{\pi_j}{1 - \pi_j}\right) = (\omega_j - \omega_{j-1}) \ln(\psi) + \ln\left(\frac{\pi_{j-1}}{1 - \pi_{j-1}}\right) \quad (26.5)$$

This assumption can also be equivalently expressed as a relationship between the odds ratio for the j -th category to that of the first category; namely,

$$\frac{\pi_j}{1 - \pi_j} = \psi^{\omega_j - \omega_1} \frac{\pi_1}{1 - \pi_1} \quad (26.6)$$

or equivalently,

$$\ln\left(\frac{\pi_j}{1 - \pi_j}\right) = (\omega_j - \omega_1) \ln(\psi) + \ln\left(\frac{\pi_1}{1 - \pi_1}\right) \quad (26.7)$$

It is assumed that $\pi_1 < \dots < \pi_R$ with $\psi > 1$ or $\pi_1 > \dots > \pi_R$ with $\psi < 1$.

We want to test the null hypothesis $H_0 : \psi = 1$ against the two sided alternative $H_1 : \psi \neq 1$ or against a one-sided alternative $H_1 : \psi > 1$ or $H_1 : \psi < 1$. The sample size required to achieve a specified power or the power for a specified sample size is determined for a single-look design with the specified parameters. The sample size calculation is conducted using the methodology presented below, which is similar to that described in Nam, 1987.

Let $n_j = r_j N$ denote the sample size for the j -th category where r_j is the j -th sample fraction and N is the total sample size. The determination of the sample size required to control the power of the test of H_0 is based on

$$W = \sum_{j=1}^R r_j (\omega_j - \bar{\omega}) \hat{\pi}_j \tag{26.8}$$

with $\bar{\omega} = \sum_{j=1}^R r_j \omega_j$

The expected value of W is

$$E(W) = \sum_{j=1}^R r_j (\omega_j - \bar{\omega}) \pi_j \tag{26.9}$$

and the variance of W is

$$V(W) = \sum_{j=1}^R r_j (\omega_j - \bar{\omega})^2 \pi_j (1 - \pi_j) \tag{26.10}$$

The expected value of W under H_0 is

$$E_0(W) = \pi \sum_{j=1}^R r_j (\omega_j - \bar{\omega}) \tag{26.11}$$

and the variance of W under H_0 is

$$V_0(W) = \pi(1 - \pi) \sum_{j=1}^R r_j (\omega_j - \bar{\omega})^2 \tag{26.12}$$

26 Binomial Superiority n-Sample

Where,

$$\pi = \sum_{j=1}^R r_j \pi_j \quad (26.13)$$

The test statistic used to determine the sample size is

$$Z = \frac{W - E_0(W)}{V_0(W)^{\frac{1}{2}}} \quad (26.14)$$

The total sample size required for a two-sided test with type-1 error rate of α to have power $1 - \beta$ when $\psi = \psi_1$ is

$$N = \frac{[z_{\alpha/2} V_0(W)^{\frac{1}{2}} + z_{\beta} V(W)^{\frac{1}{2}}]^2}{E(W)^2} \quad (26.15)$$

The total sample size required for a one-sided test with type-1 error rate of α to have power $1 - \beta$ when $\psi = \psi_1$ is determined from (1.11) with $\alpha/2$ replaced by α .

26.4.1 Trial Design

Consider the problem of comparing three durations of therapy for a specific disorder. We want to have sufficiently large power when 10% of subjects with shorter duration, 25% of subjects with intermediate duration and 50% of subjects with extensive duration will respond by the end of therapy. These parameters result in an odds ratio of $\psi = 3$ or equivalently $\ln(\psi) = 1.1$. We would like to determine the sample size to achieve 90% power when $\ln(\psi) = 1.1$ based on a two-sided test at significance level 0.05.

Start East. Click **Design** tab, then click **Many Samples** in the **Discrete** group, and then click **Trend in R Ordered Binomial Proportions**.

The Input dialog box, with default input values will appear in the upper pane.

Response probabilities can be specified in one of the two ways, selected from **Response Probabilities**: (1) User Specified Probabilities or (2) Model Based Probabilities. User can specify probabilities for each population if he or she chooses User Specified Probabilities whereas Model Based Probabilities are based on logit

transformation. We will use **Model Based Probabilities** here. Under **Response Probabilities**, click **Model Based Probabilities** radio button. A new field for **log of Common odds Ratio** will appear. Enter **1.1** in this field. Enter **0.1** in **Prop. of Response** field. One can specify the **Scores (W(i))** also in monotonically increasing order. We will use **Equally Spaced** here. Enter the inputs as shown below and click **Compute**.

Number of Populations(R):

Test Parameters

Test Type:

Type I Error (α):

Power:

Sample Size (n):

Population Fraction: Equal
 Unequal

Scores W(i): Equally Spaced
 User Specified

Response Probabilities: User Specified Probabilities
 Model Based Probabilities

Log of Common odds Ratio $\pi_1(1 - \pi_{1-1})/(1 - \pi_1)\pi_{1-1}$: * (W₁ - W_{i-1})

Specify Table of Scores

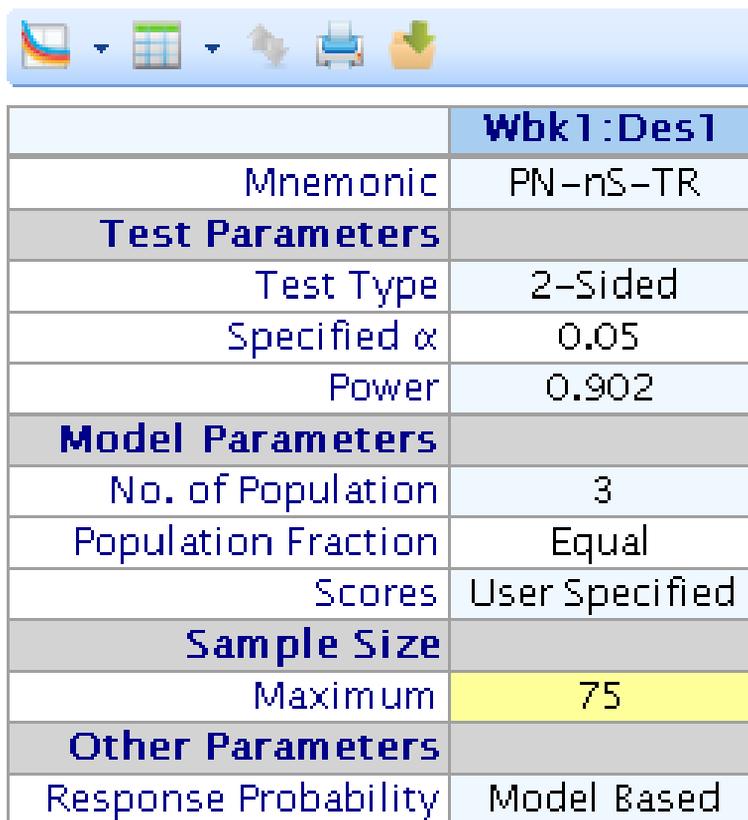
| Population # (i) | Population Fraction | Scores W(i) | Prop. Of Response |
|------------------|---------------------|-------------|-------------------|
| 1 | 0.333 | 1 | 0.1 |
| 2 | 0.333 | 2 | 0.25 |
| 3 | 0.333 | 3 | 0.501 |

The design output will be displayed in the Output Preview, with the computed sample size highlighted in yellow. A total of **75** subjects must be enrolled in order to achieve **90%** power under the alternative hypothesis. Besides sample size, one can also compute the power and the level of significance for this design.

| ID ▲ | Test Type | Specified α | Power | Sample Size | No. of Population | Population Fraction | Scores | Response Probability |
|--|-----------|-------------|-------|-------------|-------------------|---------------------|----------------|----------------------|
|  Des1 | 2-Sided | 0.05 | 0.902 | 75 | 3 | Equal | User Specified | Model Based |

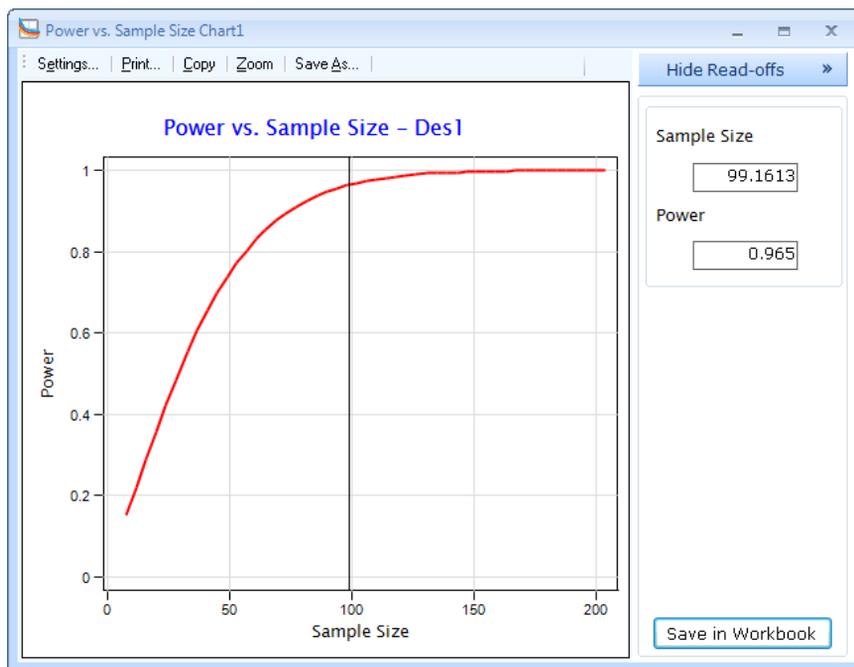
You can select this design by clicking anywhere on the row in the Output Preview. If you click on  icon, some of the design details will be displayed in the upper pane. In the Output Preview toolbar, click  icon, to save this design to Wbk1 in the **Library**. If you hover the cursor over Des1 in the **Library**, a tooltip will appear that summarizes the input parameters of the design.

26 Binomial Superiority n-Sample



| Wbk1:Des1 | |
|-------------------------|----------------|
| Mnemonic | PN-nS-TR |
| Test Parameters | |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.902 |
| Model Parameters | |
| No. of Population | 3 |
| Population Fraction | Equal |
| Scores | User Specified |
| Sample Size | |
| Maximum | 75 |
| Other Parameters | |
| Response Probability | Model Based |

With Des1 selected in the **Library**, click  icon on the **Library** toolbar, and then click **Power vs. Sample Size**. The resulting power curve for this design is shown. You can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...** For now, you may close the chart before continuing.



The default specification of equally spaced scores is useful when the categories are ordinal, but not numerical. If the categories are numerical, such as doses of a therapy, then the numerical value will be more appropriate. Consider three doses of 10, 20, and 30. One must exhibit care in specification of $\log(\psi)$ when the differences between scores for adjacent categories are equal, but this common difference is not equal to one. Although the differences are equal, user defined scores must be used. If the common difference is equal to a positive value A, then equating $\log(\psi)$ to $1/A$ of that for the default of equally spaced scores, with a common difference of one, will provide identical results. With three doses of (Scores $W(i)$) of 10, 20, and 30 and \log of Common odds Ratio = 0.11, the results are the same as those shown above. This is shown in the following screenshot.

Log of Common odds Ratio $\pi_i(1 - \pi_{i-1}) / (1 - \pi_i)\pi_{i-1}$: * $(W_i - W_{i-1})$

Specify Table of Scores

| Population #(i) | Population Fraction | Scores $W(i)$ | Prop. Of Response |
|-----------------|---------------------|---------------|-------------------|
| 1 | 0.333 | 10 | 0.1 |
| 2 | 0.333 | 20 | 0.25 |
| 3 | 0.333 | 30 | 0.501 |

26 Binomial Superiority n-Sample

The design output will be displayed in the Output Preview, with the computed sample size highlighted in yellow. A total of 75 subjects must be enrolled in order to achieve 90% power under the alternative hypothesis when $\log(\psi) = .11$ and $\pi_1 = 0.1$.

| ID ▲ | Test Type | Specified α | Power | Sample Size | No. of Population | Population Fraction | Scores | Response Probability |
|------|-----------|--------------------|-------|-------------|-------------------|---------------------|----------------|----------------------|
| Des1 | 2-Sided | 0.05 | 0.902 | 75 | 3 | Equal | User Specified | Model Based |
| Des2 | 2-Sided | 0.05 | 0.902 | 75 | 3 | Equal | User Specified | Model Based |

Similarly, if the differences between scores for adjacent categories are not equal, user defined scores must be used. Consider three doses of 10, 20, and 50, with **log of Common odds Ratio**= 0.11. Change the scores (**Scores W(i)**) to 10, 20, and 50 in the previous design. This is shown in the following screenshot.

Log of Common odds Ratio $\pi_1(1 - \pi_{i-1})/(1 - \pi_i)\pi_{i-1}$: * (W_i - W_{i-1})

Specify Table of Scores

| Population #(i) | Population Fraction | Scores W(i) | Prop. Of Response |
|-----------------|---------------------|-------------|-------------------|
| 1 | 0.333 | 10 | 0.1 |
| 2 | 0.333 | 20 | 0.25 |
| 3 | 0.333 | 50 | 0.9 |

The design output will be displayed in the Output Preview, with the computed sample size highlighted in yellow. A total of 16 subjects must be enrolled in order to achieve 90% power under the alternative hypothesis when $\log(\psi) = .11$ and $\pi_1 = 0.1$.

| ID ▲ | Test Type | Specified α | Power | Sample Size | No. of Population | Population Fraction | Scores | Response Probability |
|------|-----------|--------------------|-------|-------------|-------------------|---------------------|----------------|----------------------|
| Des1 | 2-Sided | 0.05 | 0.902 | 75 | 3 | Equal | User Specified | Model Based |
| Des2 | 2-Sided | 0.05 | 0.902 | 75 | 3 | Equal | User Specified | Model Based |
| Des3 | 2-Sided | 0.05 | 0.914 | 16 | 3 | Equal | User Specified | Model Based |

Although, a small sample size is usually desirable, here it may be due to a value of $\pi_3 (= 0.90)$ which may be too large to be meaningful. Then the power should be controlled at a smaller value of $\log(\psi)$. Consider $\log(\psi) = 0.07$. Change the **log of Common odds Ratio** value to 0.07. This is shown in the following screenshot.

Log of Common odds Ratio $\pi_1(1 - \pi_{i-1})/(1 - \pi_1)\pi_{i-1}$: * $(W_i - W_{i-1})$

Specify Table of Scores

| Population #(i) | Population Fraction | Scores W(i) | Prop. Of Response |
|-----------------|---------------------|-------------|-------------------|
| 1 | 0.333 | 10 | 0.1 |
| 2 | 0.333 | 20 | 0.183 |
| 3 | 0.333 | 50 | 0.646 |

The design output will be displayed in the Output Preview, with the computed sample size highlighted in yellow. A total of 37 subjects must be enrolled in order to achieve 90% power under the alternative hypothesis when $\log(\psi) = .07$ and $\pi_1 = 0.1$.

| ID ▲ | Test Type | Specified α | Power | Sample Size | No. of Population | Population Fraction | Scores | Response Probability |
|------|-----------|--------------------|-------|-------------|-------------------|---------------------|----------------|----------------------|
| Des1 | 2-Sided | 0.05 | 0.902 | 75 | 3 | Equal | User Specified | Model Based |
| Des2 | 2-Sided | 0.05 | 0.902 | 75 | 3 | Equal | User Specified | Model Based |
| Des3 | 2-Sided | 0.05 | 0.914 | 16 | 3 | Equal | User Specified | Model Based |
| Des4 | 2-Sided | 0.05 | 0.902 | 37 | 3 | Equal | User Specified | Model Based |

The trend test is particularly useful in situations where there are several categories. Consider now an example of a dose-ranging study to examine the safety of a therapy, with respect to the occurrence of a specified adverse event (AE), such as a dose-limiting toxicity (DLT). Six doses (1, 2, 4, 8, 12, 16) have been selected. It is expected that approximately 5% on the lowest dose will experience the AE. The study is to be designed to have power of 90% if approximately 20% on the highest dose experience the AE. This suggests that the study should be designed with $\log(\psi)$ approximately $(\log(0.20) - \log(0.05))/15 = 0.092$. Enter **log of Common odds Ratio** as 0.1, **Prop. Of Response** as 0.05 and **Number of Populations** as 6. Enter the **Scores W(i)** as 1, 2, 4, 8, 12, and 16. Leave all other values as defaults, and click **Compute**.

26 Binomial Superiority n-Sample

Number of Populations(R):

Test Parameters

Test Type:

Type I Error (α):

Power:

Sample Size (n):

Population Fraction
 Equal
 Unequal

Scores W(i)
 Equally Spaced
 User Specified

Response Probabilities
 User Specified Probabilities
 Model Based Probabilities

Log of Common odds Ratio $\pi_1(1 - \pi_{1-1})/(1 - \pi_1)\pi_{1-1}$: * (W₁ - W₁₋₁)

Specify Table of Scores

| Population # (i) | Population Fraction | Scores W(i) | Prop. Of Response |
|------------------|---------------------|-------------|-------------------|
| 1 | 0.167 | 1 | 0.05 |
| 2 | 0.167 | 2 | 0.055 |
| 3 | 0.167 | 4 | 0.066 |
| 4 | 0.167 | 8 | 0.096 |
| 5 | 0.167 | 12 | 0.137 |
| 6 | 0.167 | 16 | 0.191 |

The design output will be displayed in the Output Preview, with the computed sample size highlighted in yellow. A total of 405 subjects must be enrolled in order to achieve 90% power under the alternative hypothesis when $\log(\psi) = .1$ and $\pi_1 = 0.05$.

| ID ▲ | Test Type | Specified α | Power | Sample Size | No. of Population | Population Fraction | Scores | Response Probability |
|------|-----------|--------------------|-------|-------------|-------------------|---------------------|----------------|----------------------|
| Des1 | 2-Sided | 0.05 | 0.902 | 75 | 3 | Equal | User Specified | Model Based |
| Des2 | 2-Sided | 0.05 | 0.902 | 75 | 3 | Equal | User Specified | Model Based |
| Des3 | 2-Sided | 0.05 | 0.914 | 16 | 3 | Equal | User Specified | Model Based |
| Des4 | 2-Sided | 0.05 | 0.902 | 37 | 3 | Equal | User Specified | Model Based |
| Des5 | 2-Sided | 0.05 | 0.9 | 405 | 6 | Equal | User Specified | Model Based |

This sample size may not be economically feasible, so we instead select the sample size to achieve a power of 80%. Selecting **Power(1- β)** as 0.8 yields the result shown in the following screen shot. This design requires a combined total of 298 subjects from all groups to attain 80% power when $\log(\psi) = 0.1$ and $\pi_1 = 0.05$.

| ID ▲ | Test Type | Specified α | Power | Sample Size | No. of Population | Population Fraction | Scores | Response Probability |
|------|-----------|--------------------|-------|-------------|-------------------|---------------------|----------------|----------------------|
| Des1 | 2-Sided | 0.05 | 0.902 | 75 | 3 | Equal | User Specified | Model Based |
| Des2 | 2-Sided | 0.05 | 0.902 | 75 | 3 | Equal | User Specified | Model Based |
| Des3 | 2-Sided | 0.05 | 0.914 | 16 | 3 | Equal | User Specified | Model Based |
| Des4 | 2-Sided | 0.05 | 0.902 | 37 | 3 | Equal | User Specified | Model Based |
| Des5 | 2-Sided | 0.05 | 0.9 | 405 | 6 | Equal | User Specified | Model Based |
| Des6 | 2-Sided | 0.05 | 0.801 | 298 | 6 | Equal | User Specified | Model Based |

26.5 Chi-Square for R Unordered Binomial Proportions
26.5.1 Trial Design

Let π_{ij} denote proportions of response in i -th group and j -th category with $i = 1, 2, \dots, R$ and $j = 1, 2$ where R denotes the number of groups. The null hypothesis of equality of proportions in all groups for every category is tested against the alternative that at least one proportion is different across all groups for any category.

The null hypothesis is defined as,

$$H_0 : \pi_{i1} = \pi_0 \forall i$$

The alternative is defined as,

$$H_1 : \pi_{i1} \neq \pi_0 \text{ for any } i = 1, 2, \dots, R$$

Table 26.3: Table: $R \times 2$ Contingency Table

| Rows | Col 1 | Col 2 | Row Total |
|-----------|----------|----------|-----------|
| Row 1 | n_{11} | n_{12} | m_1 |
| Row 2 | n_{21} | n_{22} | m_2 |
| . | . | . | . |
| . | . | . | . |
| Row R | n_{R1} | n_{R2} | m_R |
| Col Total | n_1 | n_2 | N |

The test statistic is given as,

$$\chi^2 = \sum_{i=1}^R \sum_{j=1}^2 \frac{(n_{ij} - \frac{m_i n_j}{N})^2}{\frac{m_i n_j}{N}} \tag{26.16}$$

Let χ_0^2 be the observed value of χ^2 . For large samples, χ^2 has approximately Chi-squared distribution with d.f. $R - 1$. The p-value is approximated by $P(\chi_{R-1}^2 \geq \chi_0^2)$, where χ_{R-1}^2 denotes a Chi-squared random variable with d.f. = $R - 1$.

26.5.1 Trial Design

Consider a 3-arm trial with treatments A, B and C. The response is the reduction in

26 Binomial Superiority n-Sample

blood pressure (BP). From historical data it is known that the response rates of treatment A, B and C are 37.5%, 59% and 40% respectively. That is, out of 40 individuals under treatment A, 15 had a reduction in BP, out of 68 individuals under treatment B, 40 had a reduction in BP and out of 30 individuals under treatment C, 12 had a reduction in BP. Based on these data we can fill the entries in the table of proportions.

Table 26.4: Table: Proportion of Response

| Groups\Categories: | Reduction in BP | No Reduction | Marginal |
|--------------------|-----------------|--------------|----------|
| Treatment A | 0.375 | 0.625 | 1 |
| Treatment B | 0.59 | 0.41 | 1 |
| Treatment C | 0.4 | 0.6 | 1 |

This can be posed as a two-sided testing problem for testing $H_0 : \pi_A = \pi_B = \pi_C (= \pi_0, \text{ say})$ against $H_1 : \pi_i \neq \pi_0$ (for at least any $i = A, B, C$) at 0.05 level. We wish to determine the sample size to have 90% power for the values displayed in the above table.

Start East. Click **Design** tab, then click **Many Samples** in the **Discrete** group, and then click **Chi-Square Test for Unordered Binomial Proportions**.

The Input dialog box, with default input values will appear in the upper pane.

Enter the values of **Response Proportion** in each group and **Alloc.Ratio** $r_i = n_i/n_1$ where **Alloc.Ratio** $r_i = n_i/n_1$ is the corresponding weights relative to the first group . Enter the inputs as shown below and click **Compute**.

Number of Groups(R):

Test Parameters

Type I Error (α):

Power:

Sample Size (n):

Table of Proportion of Response

| Group | Response Proportion | Alloc.Ratio $r_i = n_i/n_1$ |
|--------|---------------------|-----------------------------|
| Group1 | 0.375 | 1 |
| Group2 | 0.59 | 1.7 |
| Group3 | 0.4 | 0.75 |

The design output will be displayed in the Output Preview, with the computed sample size highlighted in yellow. A total of 301 subjects must be enrolled in order to achieve

90% power under the alternative hypothesis. Besides sample size, one can also compute the power and the level of significance for this **Chi-Square test for $R \times 2$ Table** study design.

| ID ▲ | Specified α | Power | Sample Size | No. of Groups | Δ | Average Proportion | Variance of Proportions |
|--|--------------------|-------|-------------|---------------|----------|--------------------|-------------------------|
|  Des1 | 0.05 | 0.9 | 301 | 3 | 0.042 | 0.486 | 0.011 |

You can select this design by clicking anywhere on the row in the Output Preview. If you click  icon, some of the design details will be displayed in the upper pane. In the Output Preview toolbar, click  icon to save this design to Wbk1 in the **Library**. If you hover the cursor over Des1 in the **Library**, a tooltip will appear that summarizes the input parameters of the design.

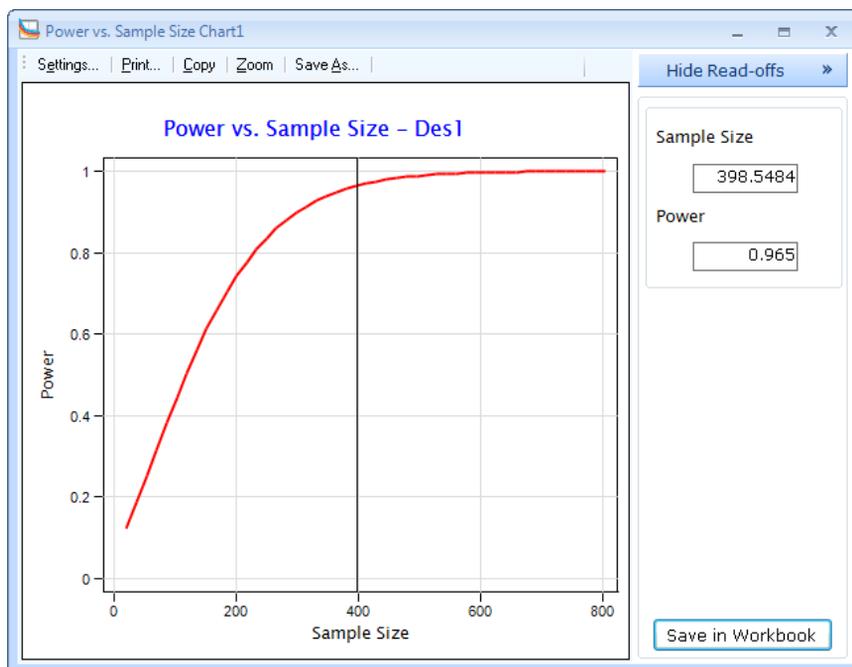


| Wbk1:Des1 | |
|--------------------------|------------|
| Mnemonic | PN-nS-CHR2 |
| Test Parameters | |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| No. of Groups | 3 |
| Effect Size (Δ) | 0.042 |
| Average Proportion | 0.486 |
| Variance of Proportions | 0.011 |
| Sample Size | |
| Maximum | 301 |

With Des1 selected in the **Library**, click  icon on the **Library** toolbar, and then

26 Binomial Superiority n-Sample

click **Power vs Sample Size**. The resulting power curve for this design is shown. You can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As....** For now, you may close the chart before continuing



26.6 Chi-Square for R Unordered Multinomial Proportions

Let π_{ij} denote the response proportion in i -th group and j -th category. The null hypothesis $H_0 : \pi_{1j} = \pi_{2j} = \dots = \pi_{Rj} \forall j = 1, 2, \dots, C$ is tested against the alternative hypothesis that for at least one category, the response proportions in all groups are not same.

The test statistic is given as,

$$\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(n_{ij} - \frac{m_i n_j}{N})^2}{\frac{m_i n_j}{N}} \quad (26.17)$$

Let χ_0^2 be the observed value of χ^2 . For large samples, χ^2 has approximately

Table 26.5: Table: Contingency Table

| Rows | Col 1 | Col 2 | . | . | Col C | Row Total |
|-----------|----------|----------|---|---|----------|-----------|
| Row 1 | n_{11} | n_{12} | . | . | n_{1C} | m_1 |
| Row 2 | n_{21} | n_{22} | . | . | n_{2C} | m_2 |
| . | . | . | . | . | . | . |
| . | . | . | . | . | . | . |
| Row R | n_{R1} | n_{R2} | . | . | n_{RC} | m_R |
| Col Total | n_1 | n_2 | . | . | n_C | m_N |

Chi-squared distribution with d.f. $(R - 1)(C - 1)$. The p-value is approximated by $P(\chi^2_{(R-1)(C-1)} \geq \chi_0^2)$, where $\chi^2_{(R-1)(C-1)}$ denotes a Chi-squared random variable with d.f. $= (R - 1)(C - 1)$.

26.6.1 Trial Design

Consider a 3-arm oncology trial with treatments A, B and C. The responses in 4 categories - CR (complete response), PR (partial response), SD (stable disease) and PD (disease progression) are of interest. We wish to determine whether the response proportion in each of the 4 categories is same for the three treatments. From historical data we get the following proportions for each category for the three treatments. Out of 100 patients, 30 were treated with treatment A, 35 were treated with treatment B and 35 were treated with treatment C. The response proportion information for each treatment is given below. Assuming equal allocation in each treatment arm, we wish to design a two-sided test which achieves 90% power at significance level 0.05.

Table 26.6: Table: Contingency Table

| Categories \ Treatment | Treatment A | Treatment B | Treatment C |
|------------------------|-------------|-------------|-------------|
| CR | 0.019 | 0.158 | 0.128 |
| PR | 0.001 | 0.145 | 0.006 |
| SD | 0.328 | 0.154 | 0.003 |
| PD | 0.652 | 0.543 | 0.863 |
| Marginal | 1 | 1 | 1 |

Start East. Click **Design** tab, then click **Many Samples** in the **Discrete** group, and then click **Chi-Square R Unordered Multinomial Proportions**.

26 Binomial Superiority n-Sample

The Input dialog box with default input values will appear in the upper pane of this window.

Enter **Number of Categories (C)** as 4. Enter the values of **Proportion of Response** and $r_i = n_i/n_1$ where $r_i = n_i/n_1$ is the corresponding weights relative to the first group. Enter the inputs as shown below and click **Compute**.

Number of Groups(R): Number of Categories(C):

Test Parameters

Type I Error (α):

Power:

Sample Size (n):

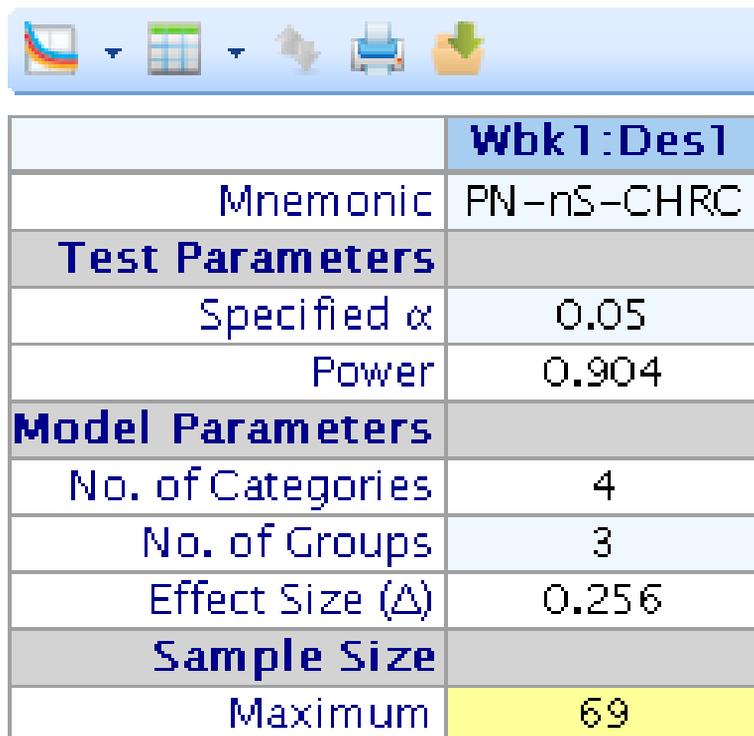
| Categories | Cat1 | Cat2 | Cat3 | Cat4 | |
|------------|-------|-------|-------|-------|--|
| Group1 | 0.019 | 0.001 | 0.328 | 0.652 | |
| Group2 | 0.158 | 0.145 | 0.154 | 0.543 | |
| Group3 | 0.128 | 0.006 | 0.003 | 0.863 | |

| $r_i = n_i/n_1$ |
|-----------------|
| 1 |
| 1.166 |
| 1.166 |

The design output will be displayed in the Output Preview, with the computed sample size highlighted in yellow. A total of 69 subjects must be enrolled in order to achieve 90% power under the alternative hypothesis. Besides sample size, one can also compute the power and the level of significance for this **Chi-Square Test of Comparing Proportions in R by C Table** study design.

| ID ▲ | Specified α | Power | Sample Size | No. of Categories | No. of Groups | Δ |
|--|--------------------|-------|-------------|-------------------|---------------|----------|
|  Des1 | 0.05 | 0.904 | 69 | 4 | 3 | 0.256 |

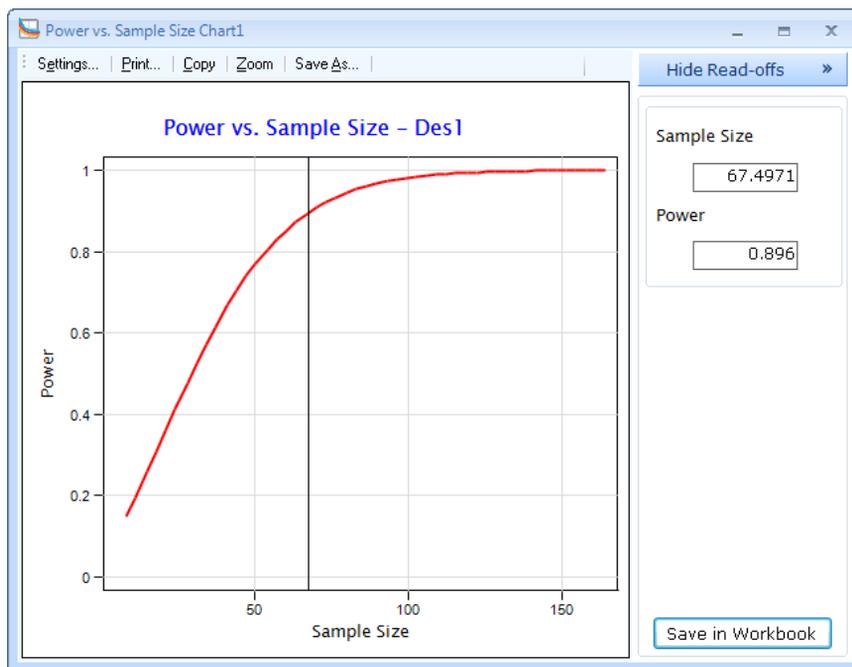
You can select this design by clicking anywhere on the row in the Output Preview. If you click  icon, some of the design details will be displayed in the upper pane. In the Output Preview toolbar, click  icon, to save this design to Wbk1 in the **Library**. If you hover the cursor over Des1 in the **Library**, a tooltip will appear that summarizes the input parameters of the design.



| | Wbk1:Des1 |
|--------------------------|------------|
| Mnemonic | PN-nS-CHRC |
| Test Parameters | |
| Specified α | 0.05 |
| Power | 0.904 |
| Model Parameters | |
| No. of Categories | 4 |
| No. of Groups | 3 |
| Effect Size (Δ) | 0.256 |
| Sample Size | |
| Maximum | 69 |

With Des1 selected in the **Library**, click  icon on the **Library** toolbar, and then click **Power vs Sample Size**. The resulting power curve for this design is shown. You can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As....** For now, you may close the chart before continuing

26 Binomial Superiority n -Sample



27 Multiple Comparison Procedures for Discrete Data

Sometime it might be the case that multiple treatment arms are compared with a placebo or control arm in one single trial on the basis of a primary endpoint that is binary. These objectives are formulated into a family of hypotheses. Formal statistical hypothesis tests can be performed to see if there is strong evidence to support clinical claims. Type I error is inflated when one considers the inferences together as a family. Failure to compensate for multiplicities can have adverse consequences. For example, a drug could be approved when actually it is not better than placebo. Multiple comparison (MC) procedures provides a guard against inflation of type I error due to multiple testing. The probability of making at least one type I error is known as family wise error rate (FWER). East supports following MC procedures based on binary endpoint.

| Procedure | Reference |
|---------------------|------------------------------------|
| Bonferroni | Bonferroni CE (1935, 1936) |
| Sidak | Sidak Z (1967) |
| Weighted Bonferroni | Benjamini Y and Hochberg Y (1997) |
| Holm’s Step Down | Holm S (1979) |
| Hochberg’s Step Up | Hochberg Y (1988) |
| Hommel’s Step Up | Hommel G (1988) |
| Fixed Sequence | Westfall PH and Krishen A (2001) |
| Fallback | Wiens B, Dmitrienko A (2005) |

In this chapter we explain how to design a study using a MC procedure.

In East, one can calculate the power from the simulated data under different MC procedures. With the information on power, one can choose the right MC procedure that provides maximum power yet strongly maintains the FWER. MC procedures included in East strongly control FWER. Strong control of FWER refers to preserving the probability of incorrectly claiming at least one null hypothesis. To contrast strong control with weak control of FWER, the latter controls the FWER under the assumption that all hypotheses are true.

27.1 Bonferroni Procedure

27.1.1 Example: HIV Study

Bonferroni procedure is described below with an example.

Assume that there are k arms including the control where the treatments arms will be compared with placebo on the basis of a binary response variable X . Let n_i be the

27 Multiple Comparison Procedures for Discrete Data

number of subjects for i -th treatment arm ($i = 0, 2, \dots, k - 1$). Let $N = \sum_{i=0}^{k-1} n_i$ be the total sample size and the arm 0 refers to control. Also, assume π_i be the response probabilities in i -th arm. We are interested in the following hypotheses:

- For the right tailed test: $H_i : \pi_i - \pi_0 \leq 0$ vs $K_i : \pi_i - \pi_0 > 0$
- For the left tailed test: $H_i : \pi_i - \pi_0 \geq 0$ vs $K_i : \pi_i - \pi_0 < 0$

For the global null hypothesis at least one of the H_i is rejected in favor of K_i after controlling for FWER. Here H_i and K_i refer to the null and alternative hypotheses, respectively, for comparison of i -th arm with the control arm.

Let $\hat{\pi}_i$ be the sample proportion for treatment arm i and $\hat{\pi}_0$ be the sample proportion for the control arm. For unpooled variance case, the test statistic to compare i -th arm with control (i.e., H_i vs K_i) is defined as

$$T_i = \frac{\hat{\pi}_i - \hat{\pi}_0}{\sqrt{\frac{1}{n_i} \hat{\pi}_i (1 - \hat{\pi}_i) + \frac{1}{n_0} \hat{\pi}_0 (1 - \hat{\pi}_0)}} \quad (i = 0, 2, \dots, k - 1) \quad (27.1)$$

For the pooled variance case, one need to replace $\hat{\pi}_i$ and $\hat{\pi}_0$ by the pooled sample proportion $\hat{\pi}$. Pooled sample proportion $\hat{\pi}$ is defined as

$$\hat{\pi} = \frac{n_i \hat{\pi}_i + n_0 \hat{\pi}_0}{n_i + n_0} \quad (i = 0, 2, \dots, k - 1) \quad (27.2)$$

Let t_i be the observed value of T_i and these observed values for $K - 1$ treatment arms can be ordered as $t_{(1)} \geq t_{(2)} \geq \dots \geq t_{(k-1)}$. For the right tailed test the marginal p-value for comparing the i -th arm with placebo is calculated as $p_i = P(Z > t_i) = \Phi(-t_i)$ and for left tailed test $p_i = P(Z < t_i) = \Phi(t_i)$, where Z is distributed as standard normal and $\Phi(\cdot)$ is the the cumulative distribution function of a standard normal variable. Let $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k-1)}$ be the ordered p-values.

East supports three single step MC procedures for comparing proportions- Bonferroni procedure, Sidak procedure and weighted Bonferroni procedure. For the Bonferroni procedure, H_i is rejected if $p_i < \frac{\alpha}{k-1}$ and the adjusted p-value is given as $\min(1, (k - 1)p_i)$.

27.1.1 Example: HIV Study

This is a randomized, double-blind, parallel-group, placebo-controlled, multi-center study to assess the efficacy and safety of 125mg, 250 mg, and 500 mg orally twice daily of a new drug for a treatment of HIV associated diarrhea. The primary efficacy endpoint is clinical response, defined as two or less watery bowel movements per

week, during at least two of the four weeks of the 4-week efficacy assessment period. The efficacy will be evaluated by comparing the proportion of responders in the placebo group to the proportion of responders in the three treatment groups at a one-sided alpha of 0.025. The estimated response rate in placebo group is 35%. The response rates in the treatment groups are expected to be 40% for 125mg, 45% for 250mg and 55% for 500 mg.

| Dose (mg) | Estimated proportion |
|-----------|----------------------|
| Placebo | 0.35 |
| 125 | 0.40 |
| 250 | 0.45 |
| 500 | 0.55 |

With the above underlying scenario, we would like to calculate the power for a total sample size of 500. This will be a balanced study with a one-sided 0.025 significance level to detect at least one dose with significant difference from placebo. We will show how to simulate the power of such a study using the multiple comparison procedures listed above.

Designing the Study

Start East. Click **Design** tab, then click **Many Samples** in the **Discrete** group, and then click **Single Look** under **Multiple Pairwise Comparisons to Control - Differences of Proportions**.

This will launch a new window which asks the user to specify the values of a few design parameters including the number of arms, overall type I error, total sample size and multiple comparison procedure. For our example, we have 3 treatment groups plus a placebo. So enter **4** for **Number of Arms**. Under the **Test Parameters** tab, there are several fields which we will fill in. First, there is a box with the label **Test Type**. Here you need to specify whether you want a one-sided or two-sided test. Currently, only one-sided tests are available. The next dropdown box has the label **Rejection Region**. If left tail is selected, the critical value for the test is located in the left tail of the distribution of the test statistic. Likewise, if right tail is selected the critical value for the test is located in the right tail of the distribution of the test statistic. For our example, we will select **Right Tail**. Under that, there is a box with the label **Type - 1 Error (α)**. This is where you need to specify the FWER. For our example, enter 0.025. Now go to the box with the label **Sample Size (n)**. Here we input the total number of subjects, including those in the placebo arm. For this example, enter 500. To the right, there will be a heading with the title **Multiple Comparison Procedures**.

27 Multiple Comparison Procedures for Discrete Data

Check the box next to **Bonferroni**, as this is the multiple comparison procedure we are illustrating in this subsection. After entering these parameters your screen should now look like this:

Number of Arms:

Test Parameters
Response Generation
Simulation Controls

Test Type:

Rejection Region:

Type I Error (α):

Sample Size (n):

Multiple Comparisons Procedures

p-value

Bonferroni Hochberg's step up

Sidak Hommel's step up

Weighted Bonferroni Fixed Sequence

Holm's step down Fallback

Now click on **Response Generation** tab. You will see a table titled **Table of Proportions**. In this table we can specify the labels for treatment arms. Also you have to specify the dose level if you want to generate proportions through dose-response curve.

There are two fields in this tab above the table. The first one is labeled as **Variance** and this has drop down list with two options - **Pooled** and **Unpooled**. Here you have to select whether you are considering pooled variance or unpooled variance for the calculation of test statistics for each test. For this example, select **Unpooled** for **Variance**.

Variance:

Generate Proportions Through DR Curve

Table of Proportions

| Arm | Response Rate |
|---------|---------------|
| Control | 0.1 |
| 1 | 0.3 |
| 2 | 0.5 |
| 3 | 0.7 |

Next to the **Variance** there is check box labeled **Generate Proportions Through DR Curve**. If you want to generate response rate for each arm according to dose-response

curve, you need to check this box. Check the box **Generate Proportions Through DR Curve**. Once you check this box you will notice two things. First, an additional column with label **Dose** will appear in the table. Here you need to enter the dose levels for each arm. For this example, enter 0, 125, 250 and 500 for Placebo, Dose1, Dose2 and Dose3 arms, respectively. Secondly, you will notice an additional section will appear to the right which provides the option to generate the response rate from four families of parametric curves which are Four Parameter Logistic, Emax, Linear and Quadratic. The technical details about each curve can be found in the Appendix H. Here you need to choose the appropriate parametric curve from the drop-down list under **Dose Response Curve** and then you have to specify the parameters associated with these curves. Suppose the response rate follows the following four parameter logistic curve:

$$E(\pi|D) = \beta + \frac{\delta}{1 + \exp\left(\frac{\theta - D}{\tau}\right)} \quad (27.3)$$

where D indicates dose. The parameter for the logistic dose-response curve should be chosen with care. We want to parameterize the above logistic model such that the proportions from logistic model agrees as close as possible to the estimated proportions stated at the beginning of the example. We will consider a situation where the response rate at dose 0 is very close to the parameter β . In other words, β indicates the placebo effect. For this to hold, $\frac{\delta}{1 + \exp\left(\frac{\theta - D}{\tau}\right)}$ should be very close to 0 at $D = 0$. For now, assume that it holds and we will return to this later. We have assumed 35% response rate in placebo arm. Therefore, we specify β as 0.35. The parameter $\beta + \delta$ indicates the maximum response rate. Since the response rate cannot exceed 1, δ should be chosen such a way that $\beta + \delta \leq 1$. The situation where the 100% response rate can never be achieved, δ would be even less. For this example, the response rate for the highest dose of 550 mg is 55%. Therefore, we assume that maximum response rate with the new drug could be achieved as only 60%. Therefore, we specify the δ as 0.60 - 0.35 or 0.25. The parameter θ indicates the median dose that can produce 50% of maximum improvement in response rate or a response that is equal to $\beta + \frac{\delta}{2}$. With $\beta = 0.35$ and $\delta = 0.25$, $\beta + \frac{\delta}{2}$ is 0.475. Note that we have assumed the dose 250 mg can provide response rate of 45%. Therefore, we assume θ as 300. τ need to be selected in such a way that $\frac{\delta}{1 + \exp\left(\frac{\theta - D}{\tau}\right)}$ should be very close to 0 at $D = 0$. We can assure this condition by choosing any small value of τ . However, a very small τ is an indicator of sharp improvement in response rate around the median dose and negligible improvement for almost other doses. In the HIV example, the estimated response rates indicate improvement in all the dose levels. With τ as 75, $\frac{\delta}{1 + \exp\left(\frac{\theta - D}{\tau}\right)}$ is 0.0045 and the proportions from the logistic regression are close to the estimated proportions for the chosen doses. Therefore, $\beta = 0.35$, $\delta = 0.25$, $\theta = 300$ and $\tau = 75$ seems to be a reasonable for our example.

27 Multiple Comparison Procedures for Discrete Data

Select **Four Parameter Logistic** from drop-down list of **Dose Response Curve**. To the right of this dropdown box, Now we need to specify the 4 parameter values in the **Parameters** box. Enter 0.35 for β , 0.25 for δ , 250 for θ and 75 for τ . You can verify that the values in **Response Rate** column is changed to 0.359, 0.39, 0.475 and 0.591 for the four arms, respectively. These proportions are very close to the estimated proportions stated at the beginning of the example.

Variance:

Generate Proportions Through DR Curve

Dose Response Curve:

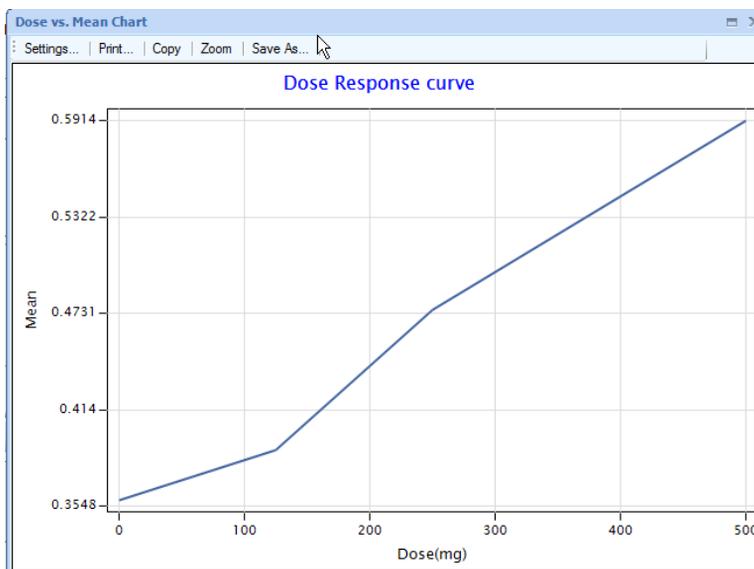
Parameters

| | | | |
|---------|----------|----------|--------|
| β | δ | θ | τ |
| 0.35 | 0.25 | 250 | 75 |

Table of Proportions

| Arm | Dose | Resp... Rate |
|---------|------|--------------|
| Control | 0 | 0.359 |
| 1 | 125 | 0.39 |
| 2 | 250 | 0.475 |
| 3 | 500 | 0.591 |

Now click **Plot DR Curve** located below the parameters to see the dose-response curve.



You will see the logistic dose response curve that intersects the Y-axis at 0.359. Close this plot. Since the response rates from logistic regression is close but not exactly

similar to the estimated proportions stated at the beginning of the example. Therefore, we will specify directly the estimated response rates in the **Table of Proportions**. In order to do this first uncheck **Generate Proportions Through DR Curve**. You will notice two things. First, the column with label **Dose** will disappear in the table. Second, the section in right will disappear as well. Now enter the estimated proportions in the **Response Rate** column. Enter 0.35, 0.40, 0.45 and 0.55 in this column. Now the **Response Generation** tab should appear as below.

Variance:

Generate Proportions Through DR Curve

Table of Proportions

| Arm | Res... Rate | |
|---------|----------------|--|
| Control | 0.35 | |
| 1 | 0.4 | |
| 2 | 0.45 | |
| 3 | 0.55 | |

Click on the **Include Options** button located in the right-upper corner in the Simulation window and check **Randomized**. This will add **Randomization** tab. Now click on the **Randomization** tab. Second column of the **Table of Allocation** table displays the allocation ratio of each treatment arm to that of control arm. The cell for the control arm is always one and is not editable. Only those cells for treatment arms other than control need to be filled in. The default value for each treatment arm is one which represents a balanced design. For the HIV study example, we consider a balanced design and leave the default values for the allocation ratios unchanged. Your

27 Multiple Comparison Procedures for Discrete Data

screen should now look like this:

Randomization Method: Fixed Allocation

Table of Allocation

| Arm | $r_i = n_i/n_1$ |
|---------|-----------------|
| Control | 1.000 |
| 1 | 1.000 |
| 2 | 1.000 |
| 3 | 1.000 |

The last tab is **Simulation Control**. Specify 10000 as **Number of Simulations** and 1000 as **Refresh Frequency** in this tab. The box labeled **Random Number Seed** is where you can set the seed for the random number generator. You can either use the clock as the seed or choose a fixed seed (in order to replicate past simulations). The default is the **clock** and we will use that. The box besides that is labeled **Output Options**. This is where you can choose to save summary statistics for each simulation run and/or to save the subject level data for a specific number of simulation runs. To save the output for each simulation, check the box with label **Save summary statistics for every simulation run**.

Number of Simulations: 10000

Refresh Frequency: 1000

Random Number Seed

Clock

Fixed [100]

Suppress All Intermediate Output

Pause after Refresh

Stop At End

Output Options

Output Type: Case Data

Save summary statistics for every simulation run

Save subject level data for [1] simulation runs

Note: Max. 100,000 records will be saved.

Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim1.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. of Pairwise Comps. | Sample Size | Rejection Region | Variance | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|------|-------------------------------|-----------|--------------------|--------------|------------------------|-------------|------------------|-----------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.807 | 0.035 | 0.807 | 500 |

Select Sim1 in the **Output Preview** and click  icon. Now double-click on Sim1 in the **Library**. The simulation output details will be displayed in the right pane.

Discrete Endpoint: Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Differences of Proportions

Hypothesis:

$H_1: \pi_1 - \pi_0 \leq 0$ Vs. $K_1: \pi_1 - \pi_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

| Test Parameters | |
|-------------------------------|------------|
| Simulation ID | Sim1 |
| Test Type | 4-Sided |
| Rejection Region | Right-Tail |
| Nominal Type-1 Error (α) | 0.025 |
| Sample Size (n) | 500 |
| Multiple Comparison Procedure | Bonferroni |
| Total Number of Arms (k) | 4 |
| Variance type | Un-pooled |
| Simulation Control Parameters | |
| Clock | Clock |
| Number of Simulations | 10000 |

Overall Powers

| | |
|--|-------|
| Global (Reject any H_i) | 0.807 |
| Conjunctive (Reject all H_i where $\pi_i > \pi_0$) | 0.035 |
| Disjunctive (Reject at least one H_i where $\pi_i > \pi_0$) | 0.807 |
| POWER (Reject any H_i where $\pi_i \leq \pi_0$) | 0 |

Marginal Powers

| | |
|-------|-------|
| Arm 1 | 0.059 |
| Arm 2 | 0.222 |
| Arm 3 | 0.797 |

Treatment Parameters

| Arm | Response Rate | Sample Size Per Group | Allocation Ratio |
|---------|---------------|-----------------------|------------------|
| Control | 0.35 | 125 | 1 |
| 1 | 0.4 | 125 | 1 |
| 2 | 0.45 | 125 | 1 |
| 3 | 0.55 | 125 | 1 |

Simulation Seed and Elapsed Time

| | |
|-----------------------------|----------|
| Starting Seed | 2184857 |
| Total Number of Simulations | 10000 |
| Elapsed Time | 00:00:04 |

Summary

Using Bonferroni test, this study has 0.807 global power to detect at least one treatment arm which is different from control arm given a total sample size 500. This test has conjunctive power 0.035 to detect all treatment arms which are truly different from control arm. This test provides disjunctive power 0.807 to detect at least one treatment arm which is truly different from control arm.

The first section in the output is the **Hypothesis** section. In our situation, we are testing 3 hypotheses. We are comparing the estimated response rate of each dose group with that of placebo. That is, we are testing the 3 hypotheses:

$$H_1 : \pi_1 = \pi_0 \quad \text{vs} \quad K_1 : \pi_1 > \pi_0$$

$$H_2 : \pi_2 = \pi_0 \quad \text{vs} \quad K_2 : \pi_2 > \pi_0$$

$$H_3 : \pi_3 = \pi_0 \quad \text{vs} \quad K_3 : \pi_3 > \pi_0$$

Here, π_0, π_1, π_2 and π_3 represent the population response rate for the placebo, 125 mg, 250 mg and 500 mg dose groups, respectively. Also, H_i and K_i are the null and alternative hypotheses, respectively, for the i -th test.

The **Input Parameters** section provides the design parameters that we specified earlier. The next section **Overall Power** gives us estimated power based on the simulation. The second line gives us the global power, which is 0.807. Global power indicates the power to reject global null $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_0$. Thus, the global power of 0.807 indicates that 80.7% of times the global null will be rejected. In other words, at least one of the H_1, H_2 and H_3 is rejected in 81.2% of the occasions. Global

27 Multiple Comparison Procedures for Discrete Data

power is useful to show the existence of dose-response relationship and the dose-response may be claimed if any of the doses in the study is significantly different from placebo.

The next line displays the conjunctive power. Conjunctive power indicates the proportion of cases in the simulation where all the H_i 's, which are truly false, were rejected. In this example, all the H_i 's are false. Therefore, for this example, conjunctive power is the proportion of cases where all of the H_1 , H_2 and H_3 were rejected. For this simulation conjunctive power is only 0.035 which means that only in 3.5% of time, all of the H_1 , H_2 and H_3 were rejected.

Disjunctive power indicates the proportion of rejecting at least one of those H_i 's where H_i is truly false. The main distinction between global and distinctive power is that the former finds any rejection whereas the latter looks for rejection only among those H_i 's which are false. Since here all of the H_1 , H_2 and H_3 are false, therefore, global and disjunctive power ought to be the same.

The next section gives us the marginal power for each hypothesis. Marginal power finds the proportion of times when a particular hypothesis is rejected. Based on simulation results, H_1 is rejected about 6% of times, H_2 is rejected about 22% of times and H_3 is rejected about 80% of times.

Recall that we have asked East to save the simulation results for each simulation run—. Open this file by clicking on **SummaryStat** in the library and you will see that it contains 10,000 rows - each rows represents results for a single simulation. Find the 3 columns with labels **Rej_Flag_1**, **Rej_Flag_2** and **Rej_Flag_3**, respectively. These columns represents the rejection status for H_1 , H_2 and H_3 , respectively. A value of 1 is indicator of rejection on that particular simulation, otherwise the null is not rejected. Now the proportion of 1's in **Rej_Flag_1** indicates the marginal power to reject H_1 . Similarly we can find out the marginal power for H_2 and H_3 from **Rej_Flag_2** and **Rej_Flag_3**, respectively. To obtain the global and disjunctive power, count the total number of cases where at least one of the H_1 , H_2 and H_3 have been rejected and then divide by the total number of simulations of 10,000. Similarly, to obtain the conjunctive power count the total number of cases where all of the H_1 , H_2 and H_3 have been rejected and then divide by the total number of simulations of 10,000.

Next we will consider an example to show how global and disjunctive power are different from each other. Select Sim 1 in **Library** and click . Now go to the the **Response Generation** tab and enter 0.35, 0.35, 0.38 and 0.42 in the 4 cells in second

column labeled as **Response Rate**.

Variance:

Generate Proportions Through DR Curve

Table of Proportions

| Arm | Res... Rate |
|---------|----------------|
| Control | 0.35 |
| 1 | 0.35 |
| 2 | 0.38 |
| 3 | 0.42 |

Here we are generating response for placebo from distribution $\text{Bin}(125, 0.35)$, for Dose1 from distribution $\text{Bin}(125, 0.35)$, for Dose2 from distribution $\text{Bin}(125, 0.38)$ and for Dose3 from distribution $\text{Bin}(125, 0.42)$. Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 2.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. of Pairwise Comps. | Sample Size | Rejection Region | Variance | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|------|-------------------------------|-----------|--------------------|--------------|------------------------|-------------|------------------|-----------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.807 | 0.035 | 0.807 | 500 |
| Sim2 | Bonferroni | 1-Sided | 0.025 | 0.009 | 3 | 500 | Right-Tail | Un-pooled | 0.121 | 0.015 | 0.118 | 500 |

For Sim 2, the global power and disjunctive power are close to 12%. To understand why, click on **SummaryStat** in the library for Sim 2. The total number of cases where at least one of H_1 , H_2 and H_3 are rejected is about 1270 and dividing this by total number of simulation 10,000 gives the global power of 12.7%. Again, the total number of cases where at least one of H_2 and H_3 are rejected is close to 1230 and dividing this by total number of simulation 10,000 gives the disjunctive power of 12.3%. The exact result of the simulations may differ slightly, depending on the seed.

Now, delete the Sim 2 from the **Output Preview** because we have modified the design in HIV example to explain the difference between global power and disjunctive power. In order to do this, select row corresponding to Sim 2 in **Output Preview** and click  in the toolbar.

27 Multiple Comparison Procedures for Discrete Data

27.2 Weighted Bonferroni procedure

In this section we will cover the weighted Bonferroni procedure with the same HIV example.

For the weighted Bonferroni procedure, H_i is rejected if $p_i < w_i\alpha$ and the adjusted p-value is given as $\min(1, \frac{p_i}{w_i})$. Here w_i denotes the proportion of α allocated to the H_i such that $\sum_{i=1}^{k-1} w_i = 1$. Note that, if $w_i = \frac{1}{k-1}$, then the Bonferroni procedure is reduced to the regular Bonferroni procedure.

Since the other design specifications remain same except that we are using weighted Bonferroni procedure in place of Bonferroni procedure, we can design simulation in this section with only little effort. Select Sim 1 in **Library** and click . Now go to the **Test Parameters** tab. In the **Multiple Comparison Procedures** box, uncheck the **Bonferroni** box and check the **Weighted Bonferroni** box.

Next click on **Response Generation** tab and look at the **Table of Proportions**. You will see an additional column with label **Proportion of Alpha** is added. Here you have to specify the proportion of total alpha you want to spend in each test. Ideally, the values in this column should add up to 1; if not, then East will normalize it to add them up to 1. By default, East distributes the total alpha equally among all tests. Here we have 3 tests in total, therefore each of the tests have proportion of alpha as $1/3$ or 0.333. You can specify other proportions as well. For this example, keep the equal

proportion of alpha for each test.

Variance: Un-pooled

Generate Proportions Through DR Curve

Table of Proportions

| Arm | Resp... Rate | Proportion of Alpha |
|---------|--------------|---------------------|
| Control | 0.35 | |
| 1 | 0.4 | 0.333 |
| 2 | 0.45 | 0.333 |
| 3 | 0.55 | 0.333 |

Now click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 2.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. of Pairwise Comps. | Sample Size | Rejection Region | Variance | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|------|-------------------------------|-----------|--------------------|--------------|------------------------|-------------|------------------|-----------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.807 | 0.035 | 0.807 | 500 |
| Sim2 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.81 | 0.034 | 0.81 | 500 |

The weighted Bonferroni MC procedure has global and disjunctive power of 81% and conjunctive power of 3.4%. Note that, the powers in the weighted Bonferroni procedure is quite close to the Bonferroni procedure. This is because the weighted Bonferroni procedure with equal proportion is equivalent to the simple Bonferroni procedure. The difference in power between Bonferroni test in previous section and the weighted Bonferroni power in this section attributed to simulation error. The exact result of the simulations may differ slightly, depending on the seed. Now select Sim2 in the **Output Preview** and click the  icon. This will save Sim2 in Wbk1 in **Library**.

27 Multiple Comparison Procedures for Discrete Data

27.3 Sidak procedures

Sidak procedures are described below using the same HIV example from the section 27.1. For the Sidak procedure, H_i is rejected if $p_i < 1 - (1 - \alpha)^{\frac{1}{k-1}}$ and the adjusted p-value is given as $1 - (1 - p_i)^{k-1}$.

Select Sim1 in **Library** and click . Now go to the **Test Parameters** tab. In the **Multiple Comparison Procedures** box, uncheck the **Bonferroni** box and check the **Sidak** box.

Number of Arms:

Test Parameters
Response Generation
Randomization
Simulation Controls

Test Type:

Rejection Region:

Type I Error (α):

Sample Size (n):

Multiple Comparisons Procedures

p-value

Bonferroni Hochberg's step up

Sidak Hommel's step up

Weighted Bonferroni Fixed Sequence

Holm's step down Fallback

Now click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional rows to the **Output Preview** labeled as Sim3.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. of Pairwise Comps. | Sample Size | Rejection Region | Variance | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|------|-------------------------------|-----------|--------------------|--------------|------------------------|-------------|------------------|-----------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.807 | 0.035 | 0.807 | 500 |
| Sim2 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.81 | 0.034 | 0.81 | 500 |
| Sim3 | Sidak | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.812 | 0.038 | 0.812 | 500 |

Sidak procedure has disjunctive and global powers of 81% and conjunctive powers of 3.8%. The exact result of the simulations may differ slightly, depending on the seed.

Now select Sim 3 in the **Output Preview** using the Ctrl key and click the  icon. This will save Sim 3 in the Wbk1 in **Library**.

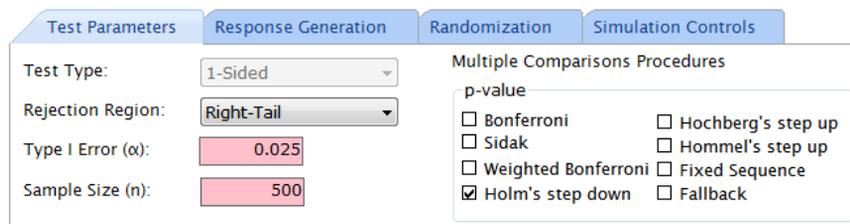
27.4 Holm's step-down procedure

In the single step MC procedures, the decision to reject any hypothesis does not depend on the decision to reject other hypotheses. On the other hand, in the stepwise procedures decision of one hypothesis test can influence the decisions on the other tests of hypotheses. There are two types of stepwise procedures. One type of procedures proceeds in data-driven order. The other type proceeds in a fixed order set a priori. Stepwise tests in a data-driven order can proceed in step-down or step-up manner. East supports Holm step down MC procedure which start with the most significant comparison and continue as long as tests are significant until the test for

certain hypothesis fails. The testing procedure stops at the first time a non-significant comparison occurs and all remaining hypotheses will be retained. In i -th step, $H_{(i)}$ is rejected if $p_{(i)} \leq \frac{\alpha}{k-i}$ and goes to the next step.

Holm's step down

As before we will use the same HIV example to illustrate Holm's step down procedure. Select Sim1 in **Library** and click . Now go to the the **Test Parameters** tab. In the **Multiple Comparison Procedures** box, uncheck the **Bonferroni** box and check the **Holm's Step down** box.



Now click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim4.

| ID ▲ | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. of Pairwise Comps. | Sample Size | Rejection Region | Variance | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|------|-------------------------------|-----------|--------------------|--------------|------------------------|-------------|------------------|-----------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.807 | 0.035 | 0.807 | 500 |
| Sim2 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.81 | 0.034 | 0.81 | 500 |
| Sim3 | Sidak | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.812 | 0.038 | 0.812 | 500 |
| Sim4 | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.811 | 0.09 | 0.811 | 500 |

Holm's step down procedure has global and disjunctive power close to 81% and conjunctive power close to 9%. The exact result of the simulations may differ slightly, depending on the seed. Now select Sim4 in the **Output Preview** and click the  icon. This will save Sim4 in Wbk1 in **Library**.

27.5 Hocheberg and Hommel procedures

Step-up tests start with the least significant comparison and continue as long as tests are not significant until the first time when a significant comparison occurs and all remaining hypotheses will be rejected. East supports two such MC procedures - Hochberg step-up and Hommel step-up procedures. In the Hochberg step-up procedure, in i -th step $H_{(k-i)}$ is retained if $p_{(k-i)} > \frac{\alpha}{i}$. In the Hommel step-up procedure, in i -th step $H_{(k-i)}$ is retained if $p_{(k-j)} > \frac{i-j+1}{i} \alpha$ for $j = 1, \dots, i$. Fixed

27 Multiple Comparison Procedures for Discrete Data

sequence test and fallback test are the types of tests which proceed in a prespecified order.

Hochberg's and Hommel's step up procedures are described below using the same HIV example from the section 27.1 on Bonferroni procedure.

Since the other design specifications remain same except that we are using Dunnett's step down in place of single step Dunnett's test we can design simulation in this section with only little effort. Select Sim1 in **Library** and click . Now go to the **Test Parameters** tab. In the **Multiple Comparison Procedures** box, uncheck the **Bonferroni** box and check the **Hochberg's step up** and **Hommel's step up** boxes.

Number of Arms:

Test Parameters
Response Generation
Randomization
Simulation Controls

Test Type:

Rejection Region:

Type I Error (α):

Sample Size (n):

Multiple Comparisons Procedures

p-value

Bonferroni Hochberg's step up

Sidak Hommel's step up

Weighted Bonferroni Fixed Sequence

Holm's step down Fallback

Now click **Simulate** to obtain power. Once the simulation run has completed, East will add two additional rows to the **Output Preview** labeled as Sim 5 and Sim 6.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. of Pairwise Comps. | Sample Size | Rejection Region | Variance | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|------|-------------------------------|-----------|--------------------|--------------|------------------------|-------------|------------------|-----------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.807 | 0.035 | 0.807 | 500 |
| Sim2 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.81 | 0.034 | 0.81 | 500 |
| Sim3 | Sidak | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.812 | 0.038 | 0.812 | 500 |
| Sim4 | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.811 | 0.09 | 0.811 | 500 |
| Sim5 | Hochberg's step up | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.812 | 0.097 | 0.812 | 500 |
| Sim6 | Hommel's step up | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.814 | 0.097 | 0.814 | 500 |

The Hocheberg and Hommel procedures have disjunctive and global powers of 81.2% and 81.4%, respectively and conjunctive powers close to 10%. The exact result of the simulations may differ slightly, depending on the seed. Now select Sim5 and Sim6 in the **Output Preview** using Ctrl key and click the  icon. This will save Sim5 and Sim6 in Wbk1 in **Library**.

27.6 Fixed-sequence testing procedure

In data-driven stepwise procedures, we don't have any control on the order of the hypotheses to be tested. However, sometimes based on our preference or prior knowledge we might want to fix the order of tests a priori. Fixed sequence test and fallback test are the types of tests which proceed in a pre-specified order. East supports both of these procedures.

Assume that H_1, H_2, \dots, H_{k-1} are ordered hypotheses and the order is prespecified so that H_1 is tested first followed by H_2 and so on. Let p_1, p_2, \dots, p_{k-1} be the associated raw marginal p-values. In the fixed sequence testing procedure, for $i = 1, \dots, k - 1$, in i -th step, if $p_i < \alpha$, reject H_i and go to the next step; otherwise retain H_i, \dots, H_{k-1} and stop.

Fixed sequence testing strategy is optimal when early tests in the sequence have largest treatment effect and performs poorly when early hypotheses have small treatment effect or are nearly true (Westfall and Krishen 2001). The drawback of fixed sequence test is that once a hypothesis is not rejected no further testing is permitted. This will lead to lower power to reject hypotheses tested later in the sequence.

As before we will use the same HIV example to illustrate fixed sequence testing procedure. Select Sim 1 in **Library** and click . Now go to the the **Test Parameters** tab. In the **Multiple Comparison Procedures** box, uncheck the **Bonferroni** box and check the **Fixed Sequence** box.

Number of Arms:

| Test Parameters | Response Generation | Randomization | Simulation Controls |
|--|--|---------------|---------------------|
| Test Type: <input type="text" value="1-Sided"/> Rejection Region: <input type="text" value="Right-Tail"/> Type I Error (α): <input type="text" value="0.025"/> Sample Size (n): <input type="text" value="500"/> | Multiple Comparisons Procedures p-value <input type="checkbox"/> Bonferroni <input type="checkbox"/> Hochberg's step up <input type="checkbox"/> Sidak <input type="checkbox"/> Hommel's step up <input type="checkbox"/> Weighted Bonferroni <input checked="" type="checkbox"/> Fixed Sequence <input type="checkbox"/> Holm's step down <input type="checkbox"/> Fallback | | |

Next click on **Response Generation** tab and look at the **Table of Proportions**. You will see an additional column with label **Test Sequence** is added. Here you have to specify the order in which the hypotheses will be tested. Specify 1 for the test that will be tested first, 2 for the test that will be tested next and so on. By default East specifies 1 to the first test, 2 to the second test and so on. For now we will keep the default

27 Multiple Comparison Procedures for Discrete Data

which means that H_1 will be tested first followed by H_2 and finally H_3 will be tested.

Variance: Un-pooled

Generate Proportions Through DR Curve

Table of Proportions

| Arm | Resp... Rate | Test Sequence |
|---------|-----------------|------------------|
| Control | 0.35 | |
| 1 | 0.4 | 1 |
| 2 | 0.45 | 2 |
| 3 | 0.55 | 3 |

Now click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim7.

| ID ▲ | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. of Pairwise Comps. | Sample Size | Rejection Region | Variance | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|------|-------------------------------|-----------|--------------------|--------------|------------------------|-------------|------------------|-----------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.807 | 0.035 | 0.807 | 500 |
| Sim2 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.81 | 0.034 | 0.81 | 500 |
| Sim3 | Sidak | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.812 | 0.038 | 0.812 | 500 |
| Sim4 | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.811 | 0.09 | 0.811 | 500 |
| Sim5 | Hochberg's step up | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.812 | 0.097 | 0.812 | 500 |
| Sim6 | Hommel's step up | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.814 | 0.097 | 0.814 | 500 |
| Sim7 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.133 | 0.097 | 0.133 | 500 |

The fixed sequence procedure with the specified sequence has global and disjunctive power close to 13% and conjunctive power close to 10%. The reason for small global and disjunctive power is due to the smallest treatment effect is tested first and the magnitude of treatment effect increases gradually for the remaining tests. For optimal power in fixed sequence procedure, the early tests in the sequence should have larger treatment effects. In our case, Dose3 has largest treatment effect followed by Dose2 and Dose1. Therefore, to obtain optimal power, H_3 should be tested first followed by H_2 and H_1 .

Select Sim7 in the **Output Preview** and click the  icon. Now, select Sim7 in **Library**, click  and go to the the **Response Generation** tab. In **Test Sequence**

column in the table, specify 3 for Dose1, 2 for Dose2 and 1 for Dose3.

Variance:

Generate Proportions Through DR Curve

Table of Proportions

| Arm | Resp... Rate | Test Sequence |
|---------|--------------|---------------|
| Control | 0.35 | |
| 1 | 0.4 | 3 |
| 2 | 0.45 | 2 |
| 3 | 0.55 | 1 |

Now click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional rows to the **Output Preview** labeled as Sim8.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. of Pairwise Comps. | Sample Size | Rejection Region | Variance | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|------|-------------------------------|-----------|--------------------|--------------|------------------------|-------------|------------------|-----------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.807 | 0.035 | 0.807 | 500 |
| Sim2 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.81 | 0.034 | 0.81 | 500 |
| Sim3 | Sidak | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.812 | 0.038 | 0.812 | 500 |
| Sim4 | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.811 | 0.09 | 0.811 | 500 |
| Sim5 | Hochberg's step up | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.812 | 0.097 | 0.812 | 500 |
| Sim6 | Hommel's step up | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.814 | 0.097 | 0.814 | 500 |
| Sim7 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.133 | 0.097 | 0.133 | 500 |
| Sim8 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.892 | 0.097 | 0.892 | 500 |

Now the fixed sequence procedure with the pre-specified sequence (H_3, H_2, H_1) has global and disjunctive power close to 89% and conjunctive power of 9.7%. This example illustrates that fixed sequence procedure is powerful provided the hypotheses are tested in a sequence of descending treatment effects. Fixed sequence procedure controls the FWER because for each hypothesis, testing is conditional upon rejecting all hypotheses earlier in sequence. The exact result of the simulations may differ slightly, depending on the seed. Select Sim8 in the **Output Preview** and click the  icon to save it in **Library**.

27.7 Fallback procedure

Fallback test alleviates the above undesirable feature for fixed sequence test. Let w_i be the proportion of α for testing H_i such that $\sum_{i=1}^{k-1} w_i = 1$. In the fixed sequence

27 Multiple Comparison Procedures for Discrete Data

testing procedure, in i -th step ($i = 1, \dots, k - 1$), test H_i at $\alpha_i = \alpha_{i-1} + \alpha w_i$ if H_{i-1} is rejected and at $\alpha_i = \alpha w_i$ if H_{i-1} is retained. If $p_i < \alpha_i$, reject H_i ; otherwise retain it. Unlike the fixed sequence testing approach, the fallback procedure can continue testing even if a non-significant outcome is encountered by utilizing the fallback strategy. If a hypothesis in the sequence is retained, the next hypothesis in the sequence is tested at the level that would have been used by the weighted Bonferroni procedure. With $w_1 = 1$ and $w_2 = \dots = w_{k-1} = 0$, the fallback procedure simplifies to fixed sequence procedure.

Again we will use the same HIV example to illustrate the fallback procedure. Select Sim 1 in **Library** and click . Now go to the **Test Parameters** tab. In the **Multiple Comparison Procedures** box, uncheck the **Dunnett's single step** box and check the **Fallback** box.

Number of Arms:

| Test Parameters | Response Generation | Randomization | Simulation Controls |
|--|--|---------------|---------------------|
| Test Type: <input type="text" value="1-Sided"/> Rejection Region: <input type="text" value="Right-Tail"/> Type I Error (α): <input type="text" value="0.025"/> Sample Size (n): <input type="text" value="500"/> | Multiple Comparisons Procedures p-value <input type="checkbox"/> Bonferroni <input type="checkbox"/> Hochberg's step up <input type="checkbox"/> Sidak <input type="checkbox"/> Hommel's step up <input type="checkbox"/> Weighted Bonferroni <input type="checkbox"/> Fixed Sequence <input type="checkbox"/> Holm's step down <input checked="" type="checkbox"/> Fallback | | |

Next click on **Response Generation** tab and look at the **Table of Proportions**. You will see two additional columns with label **Test Sequence** and **Proportion of Alpha**. In the column **Test Sequence**, you have to specify the order in which the hypotheses will be tested. Specify 1 for the test that will be tested first, 2 for the test that will be tested next and so on. By default East specifies 1 to the first test, 2 to the second test and so on. For now we will keep the default which means that H_1 will be tested first followed by H_2 and finally H_3 will be tested.

In the column **Proportions of Alpha**, you have to specify the proportion of total alpha you want to spend in each test. Ideally, the values in this column should add up to 1; if not, then East will normalize it to add them up to 1. By default East distributes the total alpha equally among the all tests. Here we have 3 tests in total, therefore each of the tests have proportion of alpha as $1/3$ or 0.333 . You can specify other proportions as

well. For this example, keep the equal proportion of alpha for each test.

Variance: Un-pooled

Generate Proportions Through DR Curve

Table of Proportions

| Arm | Resp... Rate | Test Sequence | Proportion of Alpha |
|---------|--------------|---------------|---------------------|
| Control | 0.35 | | |
| 1 | 0.4 | 1 | 0.333 |
| 2 | 0.45 | 2 | 0.333 |
| 3 | 0.55 | 3 | 0.333 |

Now click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional rows to the **Output Preview** labeled as Sim9.

| ID ▲ | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. of Pairwise Comps. | Sample Size | Rejection Region | Variance | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|------|-------------------------------|-----------|--------------------|--------------|------------------------|-------------|------------------|-----------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.807 | 0.035 | 0.807 | 500 |
| Sim2 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.81 | 0.034 | 0.81 | 500 |
| Sim3 | Sidak | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.812 | 0.038 | 0.812 | 500 |
| Sim4 | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.811 | 0.09 | 0.811 | 500 |
| Sim5 | Hochberg's step up | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.812 | 0.097 | 0.812 | 500 |
| Sim6 | Hommel's step up | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.814 | 0.097 | 0.814 | 500 |
| Sim7 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.133 | 0.097 | 0.133 | 500 |
| Sim8 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.892 | 0.097 | 0.892 | 500 |
| Sim9 | Fallback | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.811 | 0.043 | 0.811 | 500 |

The fixed sequence procedure with the specified sequence had global and disjunctive power close to 13% and conjunctive power of 9%. With the same pre-specified order for testing hypotheses, fallback procedure has superior power compared to fixed sequence procedure. This is because the fallback procedure can continue testing even if a non-significant outcome is encountered whereas the fixed sequence procedure has to stop when a hypothesis in the sequence is not rejected. Now we will consider a sequence where H_3 will be tested first followed by H_2 and H_1 because in our case, Dose3 has largest treatment effect followed by Dose2 and Dose1.

Select Sim 9 in the **Output Preview** and click the  icon. Now, select Simulation 9 in **Library**, click  and go to the the **Response Generation** tab. In **Test Sequence**

27 Multiple Comparison Procedures for Discrete Data

column in the table, specify 3 for Dose1, 2 for Dose2 and 1 for Dose3.

Variance: Un-pooled ▼

Generate Proportions Through DR Curve

Table of Proportions

| Arm | Response Rate | Test Sequence | Proportion of Alpha |
|---------|---------------|---------------|---------------------|
| Control | 0.35 | | |
| 1 | 0.4 | 3 | 0.333 |
| 2 | 0.45 | 2 | 0.333 |
| 3 | 0.55 | 1 | 0.333 |

Now click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional rows to the **Output Preview** labeled as Sim 10.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. of Pairwise Comps. | Sample Size | Rejection Region | Variance | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size |
|-------|-------------------------------|-----------|--------------------|--------------|------------------------|-------------|------------------|-----------|--------------|-------------------|-------------------|---------------------|
| Sim1 | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.807 | 0.035 | 0.807 | 500 |
| Sim2 | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.81 | 0.034 | 0.81 | 500 |
| Sim3 | Sidak | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.812 | 0.038 | 0.812 | 500 |
| Sim4 | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.811 | 0.09 | 0.811 | 500 |
| Sim5 | Hochberg's step up | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.812 | 0.097 | 0.812 | 500 |
| Sim6 | Hommel's step up | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.814 | 0.097 | 0.814 | 500 |
| Sim7 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.133 | 0.097 | 0.133 | 500 |
| Sim8 | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.892 | 0.097 | 0.892 | 500 |
| Sim9 | Fallback | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.811 | 0.043 | 0.811 | 500 |
| Sim10 | Fallback | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.811 | 0.084 | 0.811 | 500 |

Now the fixed sequence procedure with the pre-specified sequence (H_3, H_2, H_1) had global and disjunctive power of 89% and conjunctive power of 9.7%. The obtained power is very close to Sim 9. Therefore, specification of sequence in descending treatment effect does not make much difference in terms of power. The exact result of

the simulations may differ slightly, depending on the seed. Select Sim10 in the **Output Preview** and click the  icon to save it in **Library**.

27.8 Comparison of MC procedures

We have obtained the power (based on the simulations) for different MC procedures for the HIV example in the previous sections. Now the obvious question is which MC procedure to choose. To compare all the MC procedure, we will perform simulations for all the MC procedures under the following scenario.

- Treatment arms: placebo, dose1 (dose=0.3 mg), dose2 (dose=1 mg) and dose3 (dose=2 mg) with respective proportions as 0.35, 0.4, 0.45 and 0.55, respectively.
- Variance: Unpooled
- Proportion of Alpha: Equal (0.333)
- Type I Error: 0.025 (right-tailed)
- Number of Simulations:10000
- Total Sample Size:500
- Allocation ratio: 1 : 1 : 1 : 1

For comparability of simulation results, we have used similar seed for simulation under all MC procedures (we have used seed as 5643). Following output displays the powers under different MC procedures. Clean up the **Output Preview** area, select all the checkboxes corresponding to the procedures and hit **Simulate**.

| ID | Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. of Pairwise Comps. | Sample Size | Rejection Region | Variance | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size | |
|--------------------------|-------------------------------|---------------------|--------------------|--------------|------------------------|-------------|------------------|------------|--------------|-------------------|-------------------|---------------------|-----|
| <input type="checkbox"/> | Bonferroni | Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.819 | 0.039 | 0.819 | 500 |
| <input type="checkbox"/> | Sidak | Sidak | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.819 | 0.039 | 0.819 | 500 |
| <input type="checkbox"/> | Wtd. Bonferroni | Weighted Bonferroni | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.819 | 0.039 | 0.819 | 500 |
| <input type="checkbox"/> | Holm's Step down | Holm's step down | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.819 | 0.09 | 0.819 | 500 |
| <input type="checkbox"/> | Hochberg's Step up | Hochberg's step up | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.82 | 0.099 | 0.82 | 500 |
| <input type="checkbox"/> | Hommel's Step up | Hommel's step up | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.822 | 0.099 | 0.822 | 500 |
| <input type="checkbox"/> | Fixed Seq(1, 2, 3) | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.136 | 0.099 | 0.136 | 500 |
| <input type="checkbox"/> | Fallback (1, 2, 3) | Fallback | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.819 | 0.047 | 0.819 | 500 |
| <input type="checkbox"/> | Fixed Seq (3, 2, 1) | Fixed Sequence | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.895 | 0.099 | 0.895 | 500 |
| <input type="checkbox"/> | Fallback (3, 2, 1) | Fallback | 1-Sided | 0.025 | 0 | 3 | 500 | Right-Tail | Un-pooled | 0.819 | 0.085 | 0.819 | 500 |

Here we have used equal proportions for weighted Bonferroni and Fallback procedures. For the two fixed sequence testing procedures (fixed sequence and fallback) two sequences have been used - (H_1, H_2, H_3) and (H_3, H_2, H_1) . As expected, Bonferroni and weighted Bonferroni procedures provides similar powers. It appears that fixed sequence procedure with the pre-specified sequence (H_3, H_2, H_1) provides the power of 89.5% which is the maximum among all the procedures. However, fixed sequence procedure with the pre-specified sequence (H_1, H_2, H_3) provides power of 13.6%. Therefore, power in fixed sequence procedure is largely

27 *Multiple Comparison Procedures for Discrete Data*

dependent on the specification of sequence of testing and a mis-specification might result in huge drop in power.

All the remaining remaining procedures have almost equal global and disjunctive powers - about 82%. Now, in terms of conjunctive power, Hochberg's step-up and Hommel's step-up procedures have the highest conjunctive power of 9.9%. Therefore, we can choose either Hochberg's step-up or Hommel's step-up procedure for a prospective HIV study discussed in section [27.1](#).

28 *Multiple Endpoints-Gatekeeping Procedures for Discrete Data*

Clinical trials are often designed to assess benefits of a new treatment compared to a control treatment with respect to multiple clinical endpoints which are divided into hierarchically ordered families. Typically, the primary family of endpoints defines the overall outcome of the trial, provides the basis for regulatory claim and is included in the product label. The secondary families of endpoints play a supportive role and provide additional information for physicians, patients, payers and are useful for enhancing the product label. Gatekeeping procedures address multiplicity problems by explicitly taking into account the hierarchical structure of the multiple objectives. The term "gatekeeping" indicates the hierarchical decision structure where the higher ranked families serve as "gatekeepers" for the lower ranked family. Lower ranked families won't be tested if the higher ranked families have not passed requirements. Two types of gatekeeping procedures for discrete outcomes, parallel and serial, are described in this chapter. For more information about applications of gatekeeping procedures in a clinical trial setting and literature review on this topic, please refer to Dmitrienko and Tamhane (2007).

East uses simulations to assess the operating characteristics of different designs using gatekeeping procedures. For example, one could simulate the power for a variety of sample sizes in a simple batch procedure. It is important to note that when determining the sample size for a clinical trial with multiple co-primary endpoints, if the correlation among the endpoints is not taken into consideration, the sample size may be overestimated (Souza, et al 2010). East uses information about the correlation among the multiple endpoints in order to determine a more feasible sample size.

28.1 *MK-0974 (telcagepant) for Acute Migraine* Consider the randomized, placebo-controlled, double blind, parallel treatment clinical trial designed to compare two treatments for migraine, a common disease and leading cause of disability. Standard treatment includes the use of Triptans, which although generally well tolerated, have a vasoconstrictor effect, which can be problematic. This leaves a certain population of patients with underlying cardiovascular disease, uncontrolled hypertension or certain subtypes of migraine unable to access this treatment. In addition, for some patients this treatment has no or low beneficial effect and is associated with some undesirable side effects resulting in the discontinuation of the drug (Ho et al, 2008). In this study, multiple doses of the drug Telcagepant (300 mg, 150 mg), an antagonist of the CGRP receptor associated with migraine, and zolmitriptan (5mg) the standard treatment against migraine, are compared against a placebo. The five co-primary endpoints include pain freedom, pain relief, absence of photophobia (sensitivity to light), absence of phonophobia (sensitivity to sound), and absence of nausea two hours post treatment. Three co-secondary endpoints included more sustained measurements of pain freedom, pain relief, and total migraine freedom

28 Multiple Endpoints-Gatekeeping Procedures for Discrete Data

for up to a 24 hour period. The study employed a full analysis set where the multiplicity of endpoints was addressed using a step-down closed testing procedure. Due to the negative aspects of zolmitriptan, investigators were primarily interested in determining the efficacy of Telcagepant for the acute treatment of migraine with the hope of an alternative treatment with fewer associated side effects. This study will be used to illustrate the two gatekeeping procedures East provides for multiple discrete endpoints.

28.2 Serial Gatekeeping Design - Simulation for Discrete Outcomes

Serial gatekeeping procedures were studied by Maurer, Hothorn and Lehmacher (1995), Bauer et al. (1998) and Westfall and Krishen (2001). Serial gatekeepers are encountered in trials where endpoints are usually ordered from most important to least important. Suppose that a trial is declared successful only if the treatment effect is demonstrated on both primary and secondary endpoints. If endpoints in the primary trial are successful, it is only then of interest to assess the secondary endpoints. Correlation coefficients between the endpoints are bounded and East computes the valid range of acceptable values. As the number of endpoints increases, the restriction imposed on the valid range of correlation values is also greater. Therefore for illustration purpose, the above trial is simplified to consider three primary endpoints, pain freedom (PF), absence of phonophobia (phono) and absence of photophobia (photo) at two hours post treatment. Only one endpoint from the secondary family, sustained pain freedom (SPF), will be included in the example. Additionally, where the original trial studied multiple doses and treatments, this example will use only two groups to focus the comparison on the higher dose of Telcagepant of 300mg, and placebo. The example includes correlation values intended to represent zero, mild and moderate correlation accordingly, to examine its effect on power.

The efficacy, or response rate, of the endpoints for subjects in the treatment group and placebo group and a sample correlation matrix follows:

| | Response Telcagepant 300mg | Response Placebo |
|-------|----------------------------|------------------|
| PF | 0.269 | 0.096 |
| phono | 0.578 | 0.368 |
| photo | 0.51 | 0.289 |
| SPF | 0.202 | 0.05 |

| | ρ_{12} | ρ_{13} | ρ_{23} | ρ_{14} | ρ_{24} | ρ_{34} |
|--------|-------------|-------------|-------------|-------------|-------------|-------------|
| Sim 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sim 2 | 0 | 0 | 0.3 | 0 | 0 | 0 |
| Sim 3 | 0 | 0 | 0.5 | 0 | 0 | 0 |
| Sim 4 | 0 | 0 | 0.8 | 0 | 0 | 0 |
| Sim 5 | 0.3 | 0.3 | 0.3 | 0 | 0 | 0 |
| Sim 6 | 0.3 | 0.3 | 0.5 | 0 | 0 | 0 |
| Sim 7 | 0.3 | 0.3 | 0.8 | 0 | 0 | 0 |
| Sim 8 | 0 | 0 | 0.3 | 0.3 | 0 | 0 |
| Sim 9 | 0 | 0 | 0.3 | 0.5 | 0 | 0 |
| Sim 10 | 0 | 0 | 0.3 | 0.7 | 0 | 0 |
| Sim 11 | 0 | 0 | 0.8 | 0.3 | 0 | 0 |
| Sim 12 | 0 | 0 | 0.8 | 0.5 | 0 | 0 |
| Sim 13 | 0 | 0 | 0.8 | 0.7 | 0 | 0 |
| Sim 14 | 0 | 0 | 0.8 | 0.7 | 0 | 0 |

To construct the above simulations, in the **Design** tab on the **Discrete** group, click **Two Samples** and select **Multiple Comparisons-Multiple Endpoints**

28 Multiple Endpoints-Gatekeeping Procedures for Discrete Data

Design: Discrete Endpoint: Two-Sample Test - Multiple Comparisons Design - Multiple Endpoints

Number of Endpoints: 2 Include Options

Test Parameters | Response Generation | Simulation Controls

Test Type: 1-Sided
 Type I Error (α): 0.025
 Sample Size (n): 100

Common Rejection Region: Right-Tail

Endpoint Information

| EndPoint ID | Family Rank | Rejection Region |
|-------------|-------------|------------------|
| Endpoint1 | 1 | Right-Tail |
| Endpoint2 | 2 | Right-Tail |

Gatekeeping Procedure
 Parallel Serial

Test for Gatekeeper Families

Bonferroni
 Truncated Holm (Truncation Constant: 0)
 Truncated Hochberg (Truncation Constant: 0)

Test for Last Family

Bonferroni
 Sidak
 Weighted Bonferroni
 Holm's step down
 Hochberg's step up
 Hommel's step up
 Fixed Sequence
 Fallback

Simulate

At the top of this input window, the user must specify the total number of endpoints in the trial. Other input parameters such as Test Type, Type I Error (α), Sample Size (n), and whether or not a Common Rejection Region is to be used for the endpoints. If a different rejection region is desired for different endpoints, this information should be specified in the **Endpoint Information** box. Here the user can change the label, select the family rank for each endpoint and choose the rejection region (either right or left tailed).

As discussed above there are typically two types of gatekeeping procedures - serial and parallel. Parallel gatekeeping requires the rejection of at least one hypothesis test - that is only one of the families of endpoints must be significant, no matter the rank. Serial gatekeeping uses the fact that the families are hierarchically ordered, and subsequent families are only tested if the previously ranked families are significant. Once the **Gatekeeping Procedure** is selected, the user must then select the multiple comparison procedure which will be used to test the last family of endpoints. These tests are discussed in Chapter 27. If Parallel Gatekeeping is selected, the user must also specify a test for Gatekeeper Families, specifically Bonferonni, Truncated Holm or Truncated Hochberg, and is discussed more in the Parallel example which follows. The type I error specified on this screen is the nominal level of the family-wise error rate, which is defined as the probability of falsely declaring the efficacy of the new treatment compared to control with respect to any endpoint.

For the migraine example, **PF**, **phono**, and **photo** form the primary family, and **SPF** is the only outcome in the secondary family. Suppose that we would like to see the power for a sample size of 200 at a nominal type I error rate 0.025 using Bonferroni test for the secondary family. The input window will look as follows:

Design: Discrete Endpoint: Two-Sample Test - Multiple Comparisons Design - Multiple Endpoints

Number of Endpoints: Include Options

Test Parameters
Response Generation
Simulation Controls

Test Type:

Type I Error (α):

Sample Size (n):

Common Rejection Region

| EndPoint ID | Family Rank | Rejection Region |
|-------------|-------------|------------------|
| PF | 1 | Right-Tail |
| phono | 1 | Right-Tail |
| photo | 1 | Right-Tail |
| SPF | 2 | Right-Tail |

Gatekeeping Procedure

Parallel Serial

Test for Last Family

Bonferroni Hochberg's step up

Sidak Hommel's step up

Weighted Bonferroni Fixed Sequence

Holm's step down Fallback

Simulate

In addition to the **Test Parameters** tab, there is a tab labeled **Response Generation**. This is where the user specifies the underlying joint distribution among the multiple endpoints for the control arm and for the treatment arm. This is assumed to be multivariate binary with a specified correlation matrix. For the first simulation, the

28 Multiple Endpoints-Gatekeeping Procedures for Discrete Data

Common Correlation box can be checked with default value of 0.

Design: Discrete Endpoint: Two-Sample Test - Multiple Comparisons Design - Multiple Endpoints

Number of Endpoints: Include Options

Variance:

Endpoint Information:

| EndPoint ID | Response Rate Control | Response Rate Treatment |
|-------------|-----------------------|-------------------------|
| PF | 0.096 | 0.269 |
| phono | 0.368 | 0.578 |
| photo | 0.289 | 0.51 |
| SPF | 0.05 | 0.202 |

Correlation Matrix:

Common Correlation:

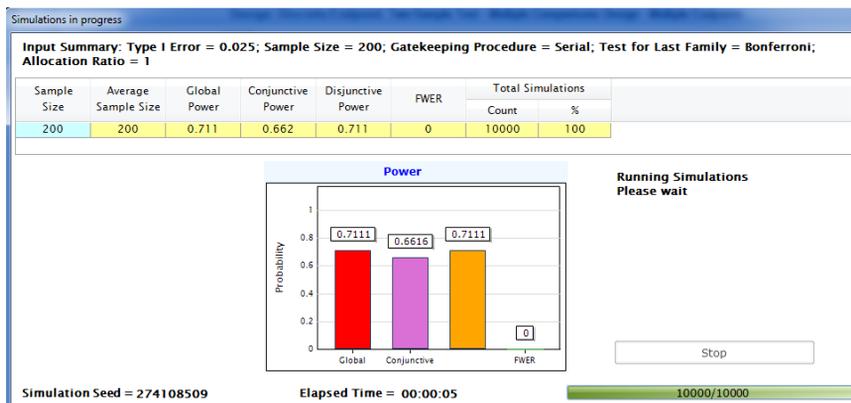
| PF | phono | photo | SPF |
|----|-------|-------|-----|
| 1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 |
| 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 1 |

Valid Range: [-0.0748, 0.3006]

Simulate

The number of simulations to be performed and other simulation parameters can be specified in the Simulation Controls window. By default, 10000 simulations will be performed. The summary statistics for each simulated trial and subject-level data can be saved by checking the appropriate boxes in the Output Options area. Once all design parameters are specified, click the **Simulate** button at the bottom right of the screen. Preliminary output is displayed in the output preview area and all results

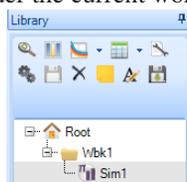
displayed in the yellow cells are summary outputs generated from simulations.



Output Preview

| ID | Test for Last Family | Test Type | Design Type | Specified α | Overall FWER | Global Power | Conjunctive Power | Disjunctive Power | No. of Endpoints | No. of Families | Sample Size | Rejection Region | Variance | Conjunctive Power First Family |
|------|----------------------|-----------|-------------|--------------------|--------------|--------------|-------------------|-------------------|------------------|-----------------|-------------|------------------|----------|--------------------------------|
| Sim1 | Bonferroni | 1-Sided | Serial | 0.025 | 0 | 0.701 | 0.651 | 0.701 | 4 | 2 | 200 | Right-Tail | Pooled | 0.701 |

To view the detailed output, first save the simulation into a workbook in the library by selecting the simulation in the Output Preview window and clicking . A simulation node will appear in the library under the current workbook.



Double click the simulation node **Sim1** in the Library to see the detailed output which summarizes all the main input parameters, including the multiple comparison procedure used for the last family of endpoints, the nominal type I error level, total sample size, mean values for each endpoint in the control arm and that in the experimental arm etc. It also displays a comprehensive list of different types of power: These different types of power are defined as follows:

28 Multiple Endpoints-Gatekeeping Procedures for Discrete Data

Overall Power and FWER:

Global: probability of declaring significance on any of the endpoints

Conjunctive: probability of declaring significance on all of the endpoints for which the treatment arm is truly better than the control arm

Disjunctive: probability of declaring significance on any of the endpoints for which the treatment arm is truly better than the control arm

FWER: probability of making at least one type I error among all the endpoints

Power and FWER for Individual Gatekeeper Family except the Last Family:

Conjunctive Power: probability of declaring significance on all of the endpoints in the particular gatekeeper family for which the treatment arm is truly better than the control arm

FWER: probability of making at least one type I error when testing the endpoints in the particular gatekeeper family

Power and FWER for the Last Family:

Conjunctive Power: probability of declaring significance on all of the endpoints in the last family for which the treatment arm is truly better than the control arm

Disjunctive Power: probability of declaring significance on any of the endpoints in the last family for which the treatment arm is truly better than the control arm

FWER: probability of making at least one type I error when testing the endpoints in the last family

Marginal Power: probability of declaring significance on the particular endpoint

Simulation: Discrete Endpoint: Two-Sample Test - Multiple Comparisons Design - Multiple Endpoints - Serial Gatekeeping

| Test Parameters | |
|-----------------------------------|------------|
| Simulation ID | Sim1 |
| Test Type | 1-Sided |
| Rejection Region | Right-Tail |
| Nominal Type-1 Error (α) | 0.025 |
| Sample Size (n) | 200 |
| Variance | Pooled |
| Multiple Endpoint Parameters | |
| Total Number of Endpoints | 4 |
| Total Number of Families | 2 |
| Gatekeeping Procedure | Serial |
| Test for Last Family | Bonferroni |

Overall Powers

| | |
|---|-------|
| Global (Reject any null hypothesis) | 0.701 |
| Conjunctive (Reject all false null hypotheses) | 0.651 |
| Disjunctive (Reject at least one false null hypothesis) | 0.701 |
| FWER (Reject any true null hypothesis) | 0 |

Output for Gatekeeper Families

| Family Rank | Conjunctive Power | FWER |
|-------------|-------------------|------|
| 1 | 0.701 | 0 |

Output for Last Family

| Family Rank | Conjunctive Power | Disjunctive Power | FWER |
|-------------|-------------------|-------------------|------|
| 2 | 0.651 | 0.651 | 0 |

Marginal Powers for Endpoints

| Endpoint ID | Marginal Power |
|-------------|----------------|
| PF | 0.701 |
| phono | 0.701 |
| photo | 0.701 |
| SPF | 0.651 |

Test Parameters

| Endpoint ID | Family Rank | Rejection Region | Null Hypothesis | Alternate Hypothesis |
|-------------|-------------|------------------|------------------------|----------------------|
| PF | 1 | Right-Tail | $\pi_{1-\pi_c} \leq 0$ | $\pi_{1-\pi_c} > 0$ |
| phono | 1 | Right-Tail | $\pi_{1-\pi_c} \leq 0$ | $\pi_{1-\pi_c} > 0$ |
| photo | 1 | Right-Tail | $\pi_{1-\pi_c} \leq 0$ | $\pi_{1-\pi_c} > 0$ |
| SPF | 2 | Right-Tail | $\pi_{1-\pi_c} \leq 0$ | $\pi_{1-\pi_c} > 0$ |

Response Parameters

| Endpoint ID | Response Rate Control | Response Rate Treatment |
|-------------|-----------------------|-------------------------|
| PF | 0.096 | 0.269 |
| phono | 0.368 | 0.578 |
| photo | 0.289 | 0.51 |
| SPF | 0.05 | 0.202 |

Correlation Matrix

| | PF | phono | photo | SPF |
|-------|----|-------|-------|-----|
| PF | 1 | 0 | 0 | 0 |
| phono | 0 | 1 | 0 | 0 |
| photo | 0 | 0 | 1 | 0 |
| SPF | 0 | 0 | 0 | 1 |

Simulation Seed and Elapsed Time

| | |
|------------------------------|-----------|
| Starting Seed: | 274336257 |
| Total Number of Simulations: | 10000 |
| Elapsed Time: | 00:00:06 |

Summary

This serial gatekeeping procedure uses intersection-union test for the gatekeeping familie(s) and Bonferroni test for the last family of endpoints. The conjunctive power of 0.651 is achieved in detecting all the endpoints for which the treatment arm is truly better than the control arm for a total sample size of 100. The disjunctive power of 0.701 is achieved in detecting at least one endpoint for which the treatment arm is truly better than the control arm for a total sample size of 200.

For the migraine example, the conjunctive power, which characterizes the power for the study, is 0.701% for a total sample size of 200. Using Bonferroni test for the last family, the design has 0.651% probability (disjunctive power for the last family) to

28 Multiple Endpoints-Gatekeeping Procedures for Discrete Data

detect the benefit of Telcagepant 300mg with respect to at least one secondary endpoints. It has 0.651% chance (conjunctive power for the last family) to declare the benefit of Telcagepant 300 mg with respect to both of the secondary endpoints. For a sample size of 200 this relatively low power is typically undesirable. One can find the sample size to achieve a target power by simulating multiple designs in a batch mode. For example, the simulation of a batch of designs for a range of sample size 200 to 300 in steps of 20 is shown by the following.

The screenshot shows a software interface for configuring a simulation. At the top, 'Number of Endpoints' is set to 4. The 'Test Parameters' tab is active, showing 'Test Type' as 1-Sided, 'Type I Error (α)' as 0.025, and 'Sample Size (n)' as 200:300:25. The 'Simulation Controls' tab is also visible, showing 'Gatekeeping Procedure' set to Serial and 'Test for Last Family' options including Bonferroni, Sidak, Weighted Bonferroni, Hochberg's step up, Hommel's step up, Fixed Sequence, Holm's step down, and Fallback. A 'Common Rejection Region' is set to Right-Tail. Below this is an 'Endpoint Information' table:

| EndPoint ID | Family Rank | Rejection Region |
|-------------|-------------|------------------|
| PF | 1 | Right-Tail |
| phono | 1 | Right-Tail |
| photo | 1 | Right-Tail |
| SPF | 2 | Right-Tail |

A 'Simulate' button is located at the bottom right of the interface.

Multiple designs can be viewed side by side for easy comparison by selecting the simulations and clicking the  in the output preview area:

| | Wbk1_Sim1 | Wbk1_Sim2 | Wbk1_Sim3 | Wbk1_Sim4 | Wbk1_Sim5 |
|-------------------------------------|------------|------------|------------|------------|------------|
| Mnemonic | PN-25-ME | PN-25-ME | PN-25-ME | PN-25-ME | PN-25-ME |
| Test Parameters | | | | | |
| Test for Last Family | Bonferroni | Bonferroni | Bonferroni | Bonferroni | Bonferroni |
| Test Type | 1-Sided | 1-Sided | 1-Sided | 1-Sided | 1-Sided |
| Design Type | Serial | Serial | Serial | Serial | Serial |
| Specified α | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 |
| No. of Endpoints | 4 | 4 | 4 | 4 | 4 |
| No. of Families | 2 | 2 | 2 | 2 | 2 |
| Rejection Region | Right-Tail | Right-Tail | Right-Tail | Right-Tail | Right-Tail |
| Variance | Pooled | Pooled | Pooled | Pooled | Pooled |
| MCP Results | | | | | |
| Overall FWER | 0 | 0 | 0 | 0 | 0 |
| Global Power | 0.697 | 0.761 | 0.827 | 0.883 | 0.918 |
| Conjunctive Power | 0.645 | 0.723 | 0.8 | 0.866 | 0.907 |
| Disjunctive Power | 0.697 | 0.761 | 0.827 | 0.883 | 0.918 |
| Conjunctive Power First Family | 0.697 | 0.761 | 0.827 | 0.883 | 0.918 |
| Sample Size | | | | | |
| Maximum | 200 | 225 | 250 | 275 | 300 |
| Other Parameters | | | | | |
| Starting Seed | 346835258 | 346841617 | 346845917 | 346849384 | 346853656 |
| Simulation Results (Overall) | | | | | |
| Average Sample Size | 200 | 225 | 250 | 275 | 300 |

For this example, to obtain a conjunctive power between 80% and 90% the study

would need to be constructed with somewhere between 250 and 300 subjects. For the remainder of this example, we will use sample size of 250 subjects under the correlation assumptions in the above table.

28.3 *Parallel Gatekeeping Design - Simulation for Discrete Outcomes*

A common concern in clinical trials with multiple primary endpoints, is whether or not statistical significance should be achieved on all endpoints. As the number of endpoints increases, this generally becomes more difficult. Parallel gatekeeping procedures are often used in clinical trials with multiple primary objectives where each individual objective can characterize a successful overall trial outcome. In other words, the trial can be declared to be successful if at least one primary objective is met.

Again, consider the same randomized, placebo-controlled, double blind, parallel treatment clinical trial designed to compare two treatments for migraine presented in the serial gatekeeping example. For the purpose of this example the trial is again simplified to study only three primary family endpoints, pain freedom (PF), absence of phonophobia (phono) and absence of photophobia (photo) at two hours post treatment. The singular endpoint in the secondary family is sustained pain freedom (SPF), and will be included in the example where, using East, power estimates will be computed via simulation. The example correlation values are intended to represent a common and moderate association among the endpoints. In general, serial gatekeeping designs require a larger sample size than parallel designs, therefore this example will use a total sample size of 125, at one-sided significance level of $\alpha = 0.025$.

The efficacy, or response rate, of the endpoints for subjects in the treatment group and placebo group and a sample correlation matrix are as follows:

| | Response Telcagepant 300mg | Response Placebo |
|-------|----------------------------|------------------|
| PF | 0.269 | 0.096 |
| phono | 0.578 | 0.368 |
| photo | 0.51 | 0.289 |
| SPF | 0.202 | 0.05 |

28 Multiple Endpoints-Gatekeeping Procedures for Discrete Data

| | ρ_{12} | ρ_{13} | ρ_{23} | ρ_{14} | ρ_{24} | ρ_{34} |
|-------|-------------|-------------|-------------|-------------|-------------|-------------|
| Sim 1 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 |
| Sim 2 | 0 | 0 | 0.8 | 0.3 | 0.0 | 0.0 |
| Sim 3 | 0.3 | 0.3 | 0.8 | 0.3 | 0 | 0 |

We now construct a new set of simulations to assess the operating characteristics of the study using a **Parallel Gatekeeping** design for the above response generation information. In the **Design** tab on the **Discrete** group, click **Two Samples** and select **Multiple Comparisons-Multiple Endpoints**

Design: Discrete Endpoint: Two-Sample Test - Multiple Comparisons Design - Multiple Endpoints

Number of Endpoints: 2 Include Options

Test Parameters | Response Generation | **Simulation Controls**

Test Type: 1-Sided

Type I Error (α): 0.025

Sample Size (n): 100

Gatekeeping Procedure: Parallel Serial

Test for Gatekeeper Families: Bonferroni Truncated Holm Truncated Hochberg

Truncation Constant: 0

Test for Last Family: Bonferroni Sidak Weighted Bonferroni Holm's step down Hochberg's step up Hommel's step up Fixed Sequence Fallback

Common Rejection Region: Right-Tail

| EndPoint ID | Family Rank | Rejection Region |
|-------------|-------------|------------------|
| Endpoint1 | 1 | Right-Tail |
| Endpoint2 | 2 | Right-Tail |

Simulate

In the **Gatekeeping Procedure** box, keep the default of **Parallel** and **Bonferroni** for the **Test for Gatekeeper Families**. For the **Test for Last Family**, also ensure that **Bonferroni** is selected as the multiple testing procedure. In the **Endpoint Information** box, specify which family each specific endpoint belongs to using the

column with the label **Family Rank**.

In the **Response Generation** window the **Variance** can be specified to be either Pooled or Un-pooled. In the **Endpoint Information** box, the Response Rates for treatment and control for each endpoint are specified. If the endpoints share a common correlation, select the **Common Correlation** checkbox and enter the correlation value to the right. East will only allow a value within the **Valid Range**. Otherwise input the specific correlation for each pair of endpoints in the **Correlation Matrix**.

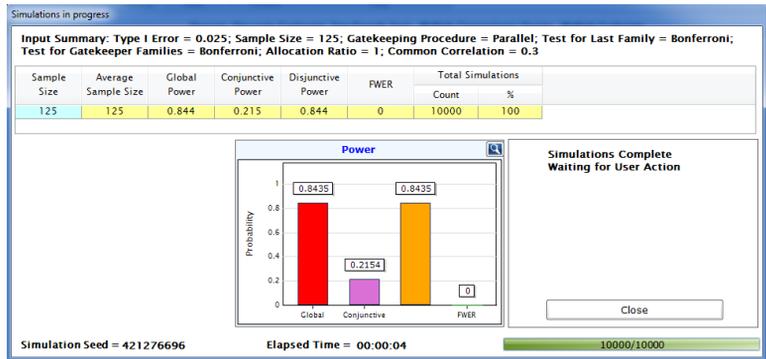
| EndPoint ID | Response Rate Control | Response Rate Treatment |
|-------------|-----------------------|-------------------------|
| PF | 0.096 | 0.269 |
| phono | 0.368 | 0.578 |
| photo | 0.289 | 0.51 |
| SPF | 0.05 | 0.202 |

| | PF | phono | photo | SPF |
|-------|-----|-------|-------|-----|
| PF | 1 | 0.3 | 0.3 | 0.3 |
| phono | 0.3 | 1 | 0.3 | 0.3 |
| photo | 0.3 | 0.3 | 1 | 0.3 |
| SPF | 0.3 | 0.3 | 0.3 | 1 |

In the **Simulation Controls** window, the user can specify the total number of simulations, refresh frequency, and random number seed. Simulation data can be saved for more advanced analyses. After all the input parameter values have been specified, click the **Simulate** button on the bottom right of the window to begin the simulation.

28 Multiple Endpoints-Gatekeeping Procedures for Discrete Data

The progress window will report how many simulations have been completed.



When complete, close the progress report screen and the preliminary simulation summary will be displayed in the output preview window. Here, one can see the overall power summary.

| | |
|-------------------------------------|------------|
| Mnemonic | PN-2S-ME |
| Test Parameters | |
| Test for Last Family | Bonferroni |
| Test for Gatekeeper Families | Bonferroni |
| Test Type | 1-Sided |
| Design Type | Parallel |
| Specified α | 0.025 |
| No. of Endpoints | 4 |
| No. of Families | 2 |
| Rejection Region | Right-Tail |
| Variance | Pooled |
| MCP Results | |
| Overall FWER | 0 |
| Global Power | 0.844 |
| Conjunctive Power | 0.215 |
| Disjunctive Power | 0.844 |
| Conjunctive Power First Family | 0.241 |
| Disjunctive Power First Family | 0.844 |
| Sample Size | |
| Maximum | 125 |
| Other Parameters | |
| Starting Seed | 421276696 |
| Simulation Results (Overall) | |
| Average Sample Size | 125 |

To see the detailed output, save the simulation in the current workbook by clicking the  icon. A simulation node will be appended to the corresponding workbook in the library. Double click the simulation node in the library to display the detailed outputs.

Simulation: Discrete Endpoint: Two-Sample Test - Multiple Comparisons Design - Multiple Endpoints - Parallel Gatekeeping

| Test Parameters | |
|-----------------------------------|------------|
| Simulation ID | Sim2 |
| Test Type | 1-Sided |
| Rejection Region | Right-Tail |
| Nominal Type-1 Error (α) | 0.025 |
| Sample Size (n) | 125 |
| Variance | Pooled |
| Multiple Endpoint Parameters | |
| Total Number of Endpoints | 4 |
| Total Number of Families | 2 |
| Gatekeeping Procedure | Parallel |
| Test for Gatekeeper Families | Bonferroni |
| Common Correlation | 0.3 |
| Test for Last Family | Bonferroni |

Overall Powers

| | |
|---|-------|
| Global (Reject any null hypothesis) | 0.844 |
| Conjunctive (Reject all false null hypotheses) | 0.215 |
| Disjunctive (Reject at least one false null hypothesis) | 0.844 |
| FWER (Reject any true null hypothesis) | 0 |

Output for Gatekeeper Families

| Family Rank | Conjunctive Power | Disjunctive Power | FWER |
|-------------|-------------------|-------------------|------|
| 1 | 0.241 | 0.844 | 0 |

Output for Last Family

| Family Rank | Conjunctive Power | Disjunctive Power | FWER |
|-------------|-------------------|-------------------|------|
| 2 | 0.603 | 0.603 | 0 |

Marginal Powers for Endpoints

| Endpoint ID | Marginal Power |
|-------------|----------------|
| PF | 0.552 |
| phono | 0.519 |
| photo | 0.576 |
| SPF | 0.603 |

Test Parameters

| Endpoint ID | Family Rank | Rejection Region | Null Hypothesis | Alternate Hypothesis |
|-------------|-------------|------------------|------------------------|----------------------|
| PF | 1 | Right-Tail | $\pi_{1-\pi_c} \leq 0$ | $\pi_{1-\pi_c} > 0$ |
| phono | 1 | Right-Tail | $\pi_{1-\pi_c} \leq 0$ | $\pi_{1-\pi_c} > 0$ |
| photo | 1 | Right-Tail | $\pi_{1-\pi_c} \leq 0$ | $\pi_{1-\pi_c} > 0$ |
| SPF | 2 | Right-Tail | $\pi_{1-\pi_c} \leq 0$ | $\pi_{1-\pi_c} > 0$ |

Response Parameters

| Endpoint ID | Response Rate Control | Response Rate Treatment |
|-------------|-----------------------|-------------------------|
| PF | 0.096 | 0.269 |
| phono | 0.368 | 0.578 |
| photo | 0.289 | 0.51 |
| SPF | 0.05 | 0.202 |

Correlation Matrix

| | PF | phono | photo | SPF |
|-------|-----|-------|-------|-----|
| PF | 1 | 0.3 | 0.3 | 0.3 |
| phono | 0.3 | 1 | 0.3 | 0.3 |
| photo | 0.3 | 0.3 | 1 | 0.3 |
| SPF | 0.3 | 0.3 | 0.3 | 1 |

Simulation Seed and Elapsed Time

Starting Seed: 421276696
 Total Number of Simulations: 10000
 Elapsed Time: 00:00:04

Summary

This parallel gatekeeping procedure uses Bonferroni test for gatekeeping familie(s) and Bonferroni test for the last family. The conjunctive power of 0.215 is achieved in detecting all endpoints for which the treatment arm is truly better than the control arm for a total sample size of 125. The disjunctive power of 0.844 is achieved in detecting at least one endpoint for which the treatment arm is truly better than the control arm for a total sample size of 125.

28 Multiple Endpoints-Gatekeeping Procedures for Discrete Data

As with serial gatekeeping, East provides following types of power:

Overall Power and FWER:

Global: probability of declaring significance on any of the endpoints.

Conjunctive: probability of declaring significance on all of the endpoints for which the treatment arm is truly better than the control arm.

Disjunctive: probability of declaring significance on any of the endpoints for which the treatment arm is truly better than the control arm.

FWER: probability of making at least one type I error among all the endpoints.

Power and FWER for Individual Gatekeeper Families except the Last Family:

Conjunctive Power: probability of declaring significance on all of the endpoints in the particular gatekeeper family for which the treatment arm is truly better than the control arm.

Disjunctive Power: probability of declaring significance on any of the endpoints in the particular gatekeeper family for which the treatment arm is truly better than the control arm.

FWER: probability of making at least one type I error when testing the endpoints in the particular gatekeeper family.

Power and FWER for the Last Gatekeeper Family:

Conjunctive Power: probability of declaring significance on all of the endpoints in the last family for which the treatment arm is truly better than the control arm.

Disjunctive Power: probability of declaring significance on any of the endpoints in the last family for which the treatment arm is truly better than the control arm.

FWER: probability of making at least one type I error when testing the endpoints in the last family.

Marginal Power: probability of declaring significance on the particular endpoint.

For the migraine example under the lower common correlation assumption, we see that the gatekeeping procedure using the Bonferroni test for both the primary family and the secondary family provides 84.4% power to detect the difference in at least one of the three primary measures of migraine relief. It only provides 24.1% power to detect the differences in all types of relief. The marginal power table displays the probabilities of declaring significance on the particular endpoint after multiplicity adjustment. For example, the power to detect sustained pain relief beyond 2 hours for a dose of 300 mg of telecapant is 60.3

To assess the robustness of this procedure with respect to the correlation among the

Table 28.1: Power Comparisons under Different Correlation Assumptions

| Correlation | Primary Family | | Secondary Family | | Overall Power | |
|-------------|----------------|-----------|------------------|-----------|---------------|-----------|
| | Disjunct. | Conjunct. | Disjunct. | Conjunct. | Disjunct. | Conjunct. |
| Sim 1 | 0.839 | 0.242 | 0.599 | 0.99 | 0.839 | 0.218 |
| Sim 2 | 0.838 | 0.244 | 0.579 | 0.579 | 0.838 | 0.202 |
| Sim 3 | 0.787 | 0.286 | 0.554 | 0.554 | 0.787 | 0.234 |

different endpoints, the simulation can be run again with different combinations of correlations. Right click on the simulation node in the Library and select **Edit Simulation** from the dropdown list. Next click on the **Response Generation** tab, update the correlation matrix, and click **Simulate**. This can be repeated for all desired correlation combinations and be compared in an output summary.

| | Migraine:Sim1 | Migraine:Sim2 | Migraine:Sim3 |
|-------------------------------------|---------------|---------------|---------------|
| Mnemonic | PN-25-ME | PN-25-ME | PN-25-ME |
| Test Parameters | | | |
| Test for Last Family | Bonferroni | Bonferroni | Bonferroni |
| Test for Gatekeeper Families | Bonferroni | Bonferroni | Bonferroni |
| Test Type | 1-Sided | 1-Sided | 1-Sided |
| Design Type | Parallel | Parallel | Parallel |
| Specified α | 0.025 | 0.025 | 0.025 |
| No. of Endpoints | 4 | 4 | 4 |
| No. of Families | 2 | 2 | 2 |
| Rejection Region | Right-Tail | Right-Tail | Right-Tail |
| Variance | Pooled | Pooled | Pooled |
| MCP Results | | | |
| Overall FWER | 0 | 0 | 0 |
| Global Power | 0.839 | 0.838 | 0.787 |
| Conjunctive Power | 0.218 | 0.202 | 0.234 |
| Disjunctive Power | 0.839 | 0.838 | 0.787 |
| Conjunctive Power First Family | 0.242 | 0.244 | 0.286 |
| Disjunctive Power First Family | 0.839 | 0.838 | 0.787 |
| Sample Size | | | |
| Maximum | 125 | 125 | 125 |
| Other Parameters | | | |
| Starting Seed | 509303891 | 509407118 | 509697192 |
| Simulation Results (Overall) | | | |
| Average Sample Size | 125 | 125 | 125 |

The following table summarizes the power comparisons under different correlation assumptions. Note that the disjunctive power decreases as the correlation increases and conjunctive power increases as the correlation increases.

There are three available parallel gatekeeping methods: Bonferroni, Truncated Holm and Truncated Hochberg. The multiple comparison procedures applied to the gatekeeper families need to satisfy the so-called separable condition. A multiple comparison procedure is separable if the type I error rate under partial null configuration is strictly less than the nominal level α . Bonferroni is a separable

28 Multiple Endpoints-Gatekeeping Procedures for Discrete Data

Table 28.2: Impact of Truncation Constant on Power in the Truncated Holm Procedure

| Truncation Constant | Primary Family | | Secondary Family | | Overall Power | |
|---------------------|----------------|-----------|------------------|-----------|---------------|-----------|
| | Conjunct. | Disjunct. | Conjunct. | Disjunct. | Conjunct. | Disjunct. |
| 0 | 0.234 | 0.84 | 0.59 | 0.59 | 0.21 | 0.84 |
| 0.25 | 0.28 | 0.833 | 0.569 | 0.569 | 0.248 | 0.833 |
| 0.5 | 0.315 | 0.836 | 0.542 | 0.542 | 0.275 | 0.836 |
| 0.8 | 0.383 | 0.838 | 0.488 | 0.488 | 0.334 | 0.838 |

procedure, however, the regular Holm and Hochberg procedure are not separable and can't be applied directly to the gatekeeper families. The truncated versions obtained by taking the convex combinations of the critical constants for the regular Holm/Hochberg procedure and Bonferroni procedure are separable and more powerful than Bonferroni test. The truncation constant leverages the degree of conservativeness. The larger value of the truncation constant results in more powerful procedure. If the truncation constant is set to be 1, it reduces to the regular Holm or Hochberg test.

To see this, simulate the design using the truncated Holm procedure for the primary family and Bonferroni test for the second family for the migraine example with common correlation 0.3. The table below compares the conjunctive power and disjunctive power for each family and the overall ones for different truncation parameter values. As the value of the truncation parameter increases, the conjunctive power for the primary family increases and the disjunctive power remain unchanged. Both the conjunctive power and disjunctive power for the secondary family decrease as we increase the truncation parameter. The overall conjunctive power also increases but the overall disjunctive power remains the same with the increase of truncation parameter.

The next table shows the marginal powers of this design for different truncation parameter values. The marginal powers for the two endpoints in the primary family increase. On the other hand, the marginal powers for the endpoint in the secondary family decrease.

The last two tables display the operating characteristics for the Hochberg test with different truncation constant values. Note that both the conjunctive and disjunctive powers for the primary family increase as the truncation parameter increases. However, the power for the secondary family decreases with the larger truncation

Table 28.3: Impact of Truncation Constant on Marginal Power in the Truncated Holm Procedure

| Truncation Constant | Primary Family | | | Secondary Family |
|---------------------|----------------|-------|-------|------------------|
| | PF | Phono | Photo | SPF |
| 0 | 0.54 | 0.512 | 0.568 | 0.59 |
| 0.25 | 0.582 | 0.512 | 0.58 | 0.569 |
| 0.5 | 0.591 | 0.541 | 0.596 | 0.542 |
| 0.8 | 0.625 | 0.568 | 0.631 | 0.488 |

Table 28.4: Impact of Truncation Constant on Power in the Truncated Hochberg Procedure

| Truncation Constant | Primary Family | | Secondary Family | | Overall Power | |
|---------------------|----------------|-----------|------------------|-----------|---------------|-----------|
| | Conjunct. | Disjunct. | Conjunct. | Disjunct. | Conjunct. | Disjunct. |
| 0 | 0.234 | 0.844 | 0.595 | 0.595 | 0.208 | 0.844 |
| 0.25 | 0.303 | 0.838 | 0.578 | 0.578 | 0.268 | 0.838 |
| 0.5 | 0.322 | 0.841 | 0.544 | 0.544 | 0.281 | 0.841 |
| 0.8 | 0.407 | 0.847 | 0.494 | 0.494 | 0.351 | 0.847 |

parameter value. The marginal powers for the primary family and for the secondary family behave similarly. The overall conjunctive and disjunctive powers also increase as we increase the truncation parameter.

28 Multiple Endpoints-Gatekeeping Procedures for Discrete Data

Table 28.5: Impact of Truncation Constant in Truncated Hochberg Procedure on Marginal Power

| Truncation Constant | Primary Family | | | Secondary Family |
|---------------------|----------------|-------|-------|------------------|
| | PF | Photo | Phono | SPF |
| 0 | 0.552 | 0.52 | 0.564 | 0.595 |
| 0.25 | 0.595 | 0.529 | 0.603 | 0.578 |
| 0.5 | 0.603 | 0.54 | 0.598 | 0.544 |
| 0.8 | 0.642 | 0.592 | 0.647 | 0.494 |

29 Two-Stage Multi-arm Designs using p -value combination

29.1 Introduction

In the drug development process, identification of promising therapies and inference on selected treatments are usually performed in two or more stages. The procedure we will be discussing here is an adaptive two-stage design that can be used for the situation of multiple treatments to be compared with a control. This will allow integration of both the stages within a single confirmatory trial controlling the multiple level type-I error. After the interim analysis in the first stage, the trial may be terminated early or continued with a second stage, where the set of treatments may be reduced due to lack of efficacy or presence of safety problems with some of the treatments. This procedure in East is highly flexible with respect to stopping rules and selection criteria and also allows re-estimation of the sample size for the second stage. Simulations show that the method may be substantially more powerful than classical one-stage multiple treatment designs with the same total sample size because second stage sample size is focused on evaluating only the promising treatments identified in the first stage. This procedure is available for continuous as well discrete endpoint studies. The current chapter deals with the discrete endpoint studies only; continuous endpoint studies are handled similarly.

29.2 Study Design

29.2.1 Introduction to the Study

29.2.2 Methodology

29.2.3 Study Design Inputs

29.2.4 Simulating under Different Alternatives

This section will explore different design options available in East with the help of an example.

29.2.1 Introduction to the Study

A new chemical entity (NCE) is being developed for the treatment of reward deficiency syndrome, specifically alcohol dependence and binge eating disorder. Compared with other orally available treatments, NCE was designed to exhibit enhanced oral bioavailability, thereby providing improved efficacy for the treatment of alcohol dependence.

- Primary Objective: To evaluate the safety and efficacy of NCE compared with placebo when administered daily for 12 weeks to adults with alcohol dependence.
- Secondary Objective: To determine the optimal dose or doses of NCE.

The **primary endpoint** is defined as the percent of subjects abstinent from heavy drinking during Weeks 5 through 12 of treatment based on self-report of drinking activity. A heavy drinking day is defined as 4 or more standard alcoholic drinks in 1 day for females and 5 or more standard alcoholic drinks in 1 day for males. The endpoint is based on the patient-reported number of standard alcoholic drinks per day, transformed into a binary outcome measure, abstinence from heavy drinking.

29 Two-Stage Multi-arm Designs using p-value combination

29.2.2 Methodology

This is a multicenter, randomized, double-blind, placebo-controlled study conducted in two parts using a 2-stage adaptive design. In Stage 1, approximately 400 eligible subjects will be randomized equally among four treatment arms (NCE [doses: 1, 2.5, or 10 mg]) and matching placebo. After all subjects in Stage 1 have completed the 12-week treatment period or discontinued earlier, an interim analysis will be conducted to

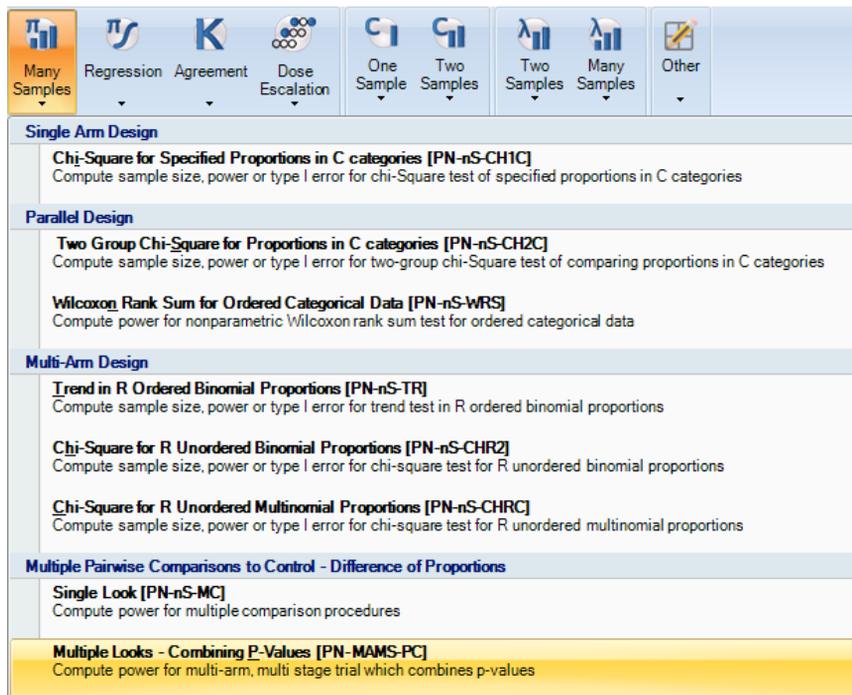
1. compare the proportion of subjects in each dose group who have achieved abstinence from heavy drinking during Weeks 5 through 12,
2. to assess safety within each dose group and
3. drop the less efficient doses.

Based on the interim analysis, Stage 2 of the study will either continue with additional subjects enrolling into 2 or 3 arms (placebo and 1 or 2 favorable, active doses) or the study will be halted completely if unacceptable toxicity has been observed.

In this example, we will have the following workflow to cover different options available in East:

1. Start with four arms (3 doses + Placebo)
2. Evaluate the three doses at the interim analysis and based on the **Treatment Selection Rules** carry forward one or two of the doses to the next stage
3. While we select the doses, also increase the sample size of the trial by using **Sample Size Re-estimation (SSR)** tool to improve conditional power if necessary
In a real trial, both the above actions (early stopping as well as sample size re-estimation) will be performed after observing the interim data.
4. See the final design output in terms of different powers, probabilities of selecting particular dose combinations
5. See the early stopping boundaries for efficacy and futility on adjusted p-value scale
6. Monitor the actual trial using the **Interim Monitoring** tool in East.

Start East. Click Design tab, then click **Many Samples** in the **Discrete** category, and then click **Multiple Looks- Combining p-values** test.



This will bring up the input window of the design with some default values. Enter the inputs as discussed below.

29.2.3 Study Design Inputs

Let us assume that three doses of the treatment *1mg*, *2.5mg*, *10mg* are compared with the Placebo arm. Preliminary sample size estimates are provided to achieve an overall study power of at least 80% at an overall, adequately adjusted 1-sided type-1 or alpha level of 2.5%, after taking into account all interim and final hypothesis tests. Note that we always use 1-sided alpha since dose-selection rules are usually 1-sided.

In Stage 1, 400 subjects are initially planned for enrollment (4 arms with 100 subjects each). Following an interim analysis conducted after all subjects in Stage 1 have completed 12 weeks of treatment or discontinued earlier, an additional 200 subjects will be enrolled into 2 doses for Stage 2 (placebo and one active dose). So we start with the total of 400+200 = 600 subjects.

The multiplicity adjustment methods available in East to compute the adjusted p-value

29 Two-Stage Multi-arm Designs using p-value combination

(p-value corresponding to global NULL) are *Bonferroni*, *Sidak*, *Simes*. For discrete endpoint test, *Dunnnett Single Step* is not available since we will be using Z-statistic. Let us use the *Bonferroni* method for this example. The p-values obtained from both the stages can be combined by using the “Inverse Normal” method. In the “Inverse Normal” method, East first computes the weights as follows:

$$w^{(1)} = \sqrt{\frac{n^{(1)}}{n}} \tag{29.1}$$

And

$$w^{(2)} = \sqrt{\frac{n^{(2)}}{n}} \tag{29.2}$$

where $n^{(1)}$ and $n^{(2)}$ are the total sample sizes corresponding to Stage 1 and stage 2 respectively and n is the total sample size.

EAST displays these weights by default but these values are editable and user can specify any other weights as long as

$$w^{(1)^2} + w^{(2)^2} = 1 \tag{29.3}$$

Final p-value is given by

$$p = 1 - \Phi\left(w^{(1)}\Phi^{-1}(1 - p^{(1)}) + w^{(2)}\Phi^{-1}(1 - p^{(2)})\right) \tag{29.4}$$

The weights specified on this tab will be used for p-value computation. $w^{(1)}$ will be used for data before interim look and $w^{(2)}$ will be used for data after interim look. Thus, according to the samples sizes planned for the two stages in this example, the weights are calculated as $\sqrt{(400/600)}$ and $\sqrt{(200/600)}$. **Note :** These weights are updated by East once we specify the first look position as 400/600 in the **Boundary** tab. So leave these as default values for now. Set the *Number of Arms* as 4 and enter the rest of the inputs as shown below:

Number of Arms: 4 Number of Looks: 2

Test Parameters Boundary Response Generation Treatment Selection Simulation Controls

Test Type: 1-Sided Multiplicity Adjustment: Bonferroni Specify Variance
 Rejection Region: Right-Tail p-value Combination: Inverse Normal Pooled Estimate
 Type I Error (α): 0.025 Weight1: 0.707 Unpooled Estimate
 Total Sample Size (n): 600 Weight2: 0.707

We can certainly have early stopping boundaries for efficacy and/or futility. But generally, in designs like this, the objective is to select the best dose(s) and not stop

early. So for now, select the **Boundary** tab and set both the boundary families to “None”. Also, set the timing of the interim analysis as 0.667 which will be after observing the data on 400 subjects out of 600. Enter 400/600 as shown below. Notice the updated weights on the bf Test Parameters tab.

Number of Arms: 4 Number of Looks: 2

Test Parameters Boundary Response Generation Treatment Selection Simulation Controls

Efficacy
 Boundary Family: None

Futility
 Boundary Family: None

Info. Fraction at Interim Look: 400/600

| Look # | Info. Fraction |
|--------|----------------|
| 1 | 0.667 |
| 2 | 1.000 |

The next tab is **Response Generation** which is used to specify the true underlying proportion of response on the individual dose groups and the initial allocation from which to generate the simulated data.

Test Parameters Boundary Response Generation

Generate Response Proportions Through DR Curve

| Treatment Label | Response Proportion | Initial Allocation (n _i /n _c) |
|-----------------|---------------------|--|
| Control | 0.1 | 1.000 |
| 1 mg | 0.14 | 1.000 |
| 2.5 mg | 0.18 | 1.000 |
| 10 mg | 0.22 | 1.000 |

Before we update the **Treatment Selection** tab, go to the **Simulation Control Parameters** tab where we can specify the number of simulations to run, the random number seed and also to save the intermediate simulation data. For now, enter the

29 Two-Stage Multi-arm Designs using p -value combination

inputs as shown below and keep all other inputs as default.

Click on the **Treatment Selection** tab. This tab is to select the scale to compute the treatment-wise effects. For selecting treatments for the second stage, the treatment effect scale will be required, but the control treatment will not be considered for selection. It will always be there in the second stage. The list under **Treatment Effect Scale** allows you to set the selection rules on different scales. Select **Estimated δ** from this list. It means that all the selection rules we specify on this tab will be in terms of the estimated value of treatment effect, δ , i.e., difference from placebo.

Here is a list of all available treatment effect scales:

Estimated Proportion, Estimated δ , Test Statistic, Conditional Power, Isotonic Proportion, Isotonic δ .

For more details on these scales, refer to the Appendix **K** chapter on this method.

The next step is to set the treatment selection rules for the second stage.

Select Best r Treatments: The best treatment is defined as the treatment having the highest or lowest mean effect. The decision is based on the rejection region. If it is “Right-Tail” then the highest should be taken as best. If it is “Left-Tail” then the lowest is taken as best. Note that the rejection region does not affect the choice of treatment based on conditional power.

Select treatments within ϵ of Best Treatment: Suppose the treatment effect scale is **Estimated δ** . If the best treatment has a treatment effect of δ_b and ϵ is specified

as 0.1 then all the treatments which have a δ as $\delta_b - 0.1$ or more are chosen for Stage 2.

Select treatments greater than threshold ζ : The treatments which have the treatment effect scale greater or less than the threshold (ζ) specified by the user according to the rejection region. But if the treatment effect scale is chosen as the conditional power then it will be greater than all the time.

Use R for Treatment Selection: If you wish to define any customized treatment selection rules, it can be done by writing an R function for those rules to be used within East. This is possible due to the R Integration feature in East. Refer to the appendix chapter on **R Functions** for more details on syntax and use of this feature. A template file for defining treatment selection rules is also available in the subfolder **RSamples** under your East installation directory.

| Tasks | File Name | Function Name |
|--------------------------|-----------|---------------|
| Treatment Selection | | |
| Initialize R Environment | | |

For more details on using R to define Treatment selection rules, refer to section [O.10](#).

For this example, select the first rule **Select Best r treatments** and set $r = 1$ which indicates that East will select the best dose for Stage 2 out the three doses. We will leave the default allocation ratio selections to yield equal allocation between the

29 Two-Stage Multi-arm Designs using p -value combination

control and selected best dose in Stage 2.

Test Parameters Boundary Response Generation **Treatment Selection** Simulation Controls

Use R For Treatment Selection

Treatment Effect Scale: Estimated δ

Treatment Selection Rules

Select Best r Treatments $r =$ Select Treatments within ϵ of Best Treatment $\epsilon =$

Select Treatments Greater than Threshold ζ $\zeta =$

Allocation Ratio after Selection

| Arm | Allocation Ratio ($n_{\cdot(i)}/n_{\cdot c}$) |
|---------|--|
| Control | 1.000 |
| Best | 1.000 |

Click the **Simulate** button to run the simulations. When the simulations are over, a row gets added in the **Output Preview** area. Save this row to the **Library** by clicking the  icon in the toolbar. Rename this scenario as **Best1**. Double click it to see the

detailed output.

⊖ Overall Powers

| | |
|---|-------|
| Global (Reject any Hi) | 0.808 |
| Conjunctive (Reject all Hi where $\pi_i > \pi_0$) | 0 |
| Disjunctive (Reject at least one Hi where $\pi_i > \pi_0$) | 0.808 |
| FWER (Reject any Hi where $\pi_i \leq \pi_0$) | 0 |

⊖ Lookwise Summary

| Look # | Average Total Sample Size (n) | Trial Termination | |
|---------|-------------------------------|-------------------|---------|
| | | Efficacy | |
| | | Count | % |
| 1 | 400 | 0 | 0.000% |
| 2 | 600 | 8083 | 80.830% |
| OverAll | | 8083 | 80.830% |

⊖ Detailed Efficacy Outcomes for all 10000 Simulations

| Efficacy Outcome | Average Sample Size | Count (n) | %(n/10000) |
|------------------|---------------------|-----------|------------|
| 1mg only | 600 | 167 | 1.670% |
| 2.5mg only | 600 | 1593 | 15.930% |
| 10mg only | 600 | 6323 | 63.230% |
| None | 600 | 1917 | 19.170% |
| Total | | 10000 | 100.000% |

⊖ Marginal Probabilities of Selection and Efficacy

| Label | Average Sample Size | Marginal Probability of Selection | | Marginal Probability of Efficacy | |
|-------|---------------------|-----------------------------------|---------|----------------------------------|---------|
| | | Count | % | Count | % |
| 1mg | 103.82 | 382 | 3.820% | 167 | 1.670% |
| 2.5mg | 122.98 | 2298 | 22.980% | 1593 | 15.930% |
| 10mg | 173.2 | 7320 | 73.200% | 6323 | 63.230% |

The first table in the detailed output shows the overall power including global power, conjunctive power, disjunctive power and FWER. The definitions for different powers are as follows:

- Global Power: probability of demonstrating statistical significance on one or more treatment groups
- Conjunctive Power: probability of demonstrating statistical significance on all treatment groups which are truly effective
- Disjunctive Power: probability of demonstrating statistical significance on at least one treatment group which is truly effective
- FWER: probability of incorrectly demonstrating statistical significance on at

29 Two-Stage Multi-arm Designs using p -value combination

least one treatment group which is truly ineffective

For our example, there is 0.8 global power, i.e., the probability of this design to reject any null hypothesis, where the set of null hypothesis are the TRUE proportion of responders at each dose equals that of control. Also shown are conjunctive and disjunctive power, as well as Family Wise Error Rate (FWER).

The **Lookwise Summary** table summarizes the number of simulated trials that ended with a conclusion of efficacy, i.e., rejected any null hypothesis, at each look. In this example, no simulated trial stopped at the interim analysis with an efficacy conclusion since there were no stopping boundaries, but 8083 simulations yielded an efficacy conclusion via the selected dose after Stage 2. This is consistent with the global power.

The next table **Detailed Efficacy Outcomes for all 10000 Simulations**, summarizes the number of simulations for which each dose was selected for Stage 2 and yielded an efficacy conclusion. For example, the dose 10mg was observed to be efficacious in 63% of simulated trials whereas none of the three doses were efficacious in 19% of trials.

The last output table **Marginal Probabilities of Selection and Efficacy**, summarizes the number and percent of simulations in which each dose was selected for Stage 2, regardless of whether it was found significant at end of Stage 2 or not, as well as the number and percent of simulations in which each dose was selected and found significant. Average sample size is also shown. Note that since this design only selected the single best dose, this table gives almost the same information as the above one.

Selecting multiple doses (arms) for Stage 2 would be of more effective than selecting just the best one.



Click the  button on the bottom left corner of the screen. This will take us back to the input window of the last simulation scenario. Go to **Treatment Selection** tab and set $r = 2$. It means that we are interested in carrying forward the two best doses out of the three. Run the simulations by keeping the sample size fixed as 600. The simulated power drops to approximately 73%. Note that the loss of power for this 2-best-doses-choice scenario in comparison to the previous example which chose only the best dose. This is because of the smaller sample sizes per dose in stage 2 for this 2-best-doses scenario since the sample size is split in Stage 2 among 2 doses and control instead of between only 1 dose and control in the best dose scenario.

Now go to **Test Parameters** tab and change the sample size to 700 assuming that each of the two doses and Placebo will get 100 subjects in Stage 2. Accordingly, update the look position on **Boundaries** tab to 400/700 as well. Click the **Simulate** button to run the simulations. When the simulations are over, a row gets added in the **Output**

Preview area. Save this row to the **Library** by clicking the  icon in the toolbar. Rename this scenario as **Best2**. Double click it to see the detailed output.

⊖ Overall Powers

| | |
|--|-------|
| Global (Reject any H_i) | 0.809 |
| Conjunctive (Reject all H_i where $\pi_i > \pi_0$) | 0 |
| Disjunctive (Reject at least one H_i where $\pi_i > \pi_0$) | 0.809 |
| FWER (Reject any H_i where $\pi_i \leq \pi_0$) | 0 |

⊖ Lookwise Summary

| Look # | Average Total Sample Size (n) | Trial Termination | |
|---------|-------------------------------|-------------------|---------|
| | | Efficacy | |
| | | Count | % |
| 1 | 400 | 0 | 0.000% |
| 2 | 700 | 8088 | 80.880% |
| OverAll | | 8088 | 80.880% |

⊖ Detailed Efficacy Outcomes for all 10000 Simulations

| Efficacy Outcome | Average Sample Size | Count (n) | %(n/10000) |
|-------------------|---------------------|-----------|------------|
| 1mg only | 700 | 10 | 0.100% |
| 2.5mg only | 700 | 250 | 2.500% |
| 10mg only | 700 | 3067 | 30.670% |
| 1mg & 2.5mg only | 700 | 130 | 1.300% |
| 1mg & 10mg only | 700 | 555 | 5.550% |
| 2.5mg & 10mg only | 700 | 4076 | 40.760% |
| None | 700 | 1912 | 19.120% |
| Total | | 10000 | 100.000% |

⊖ Marginal Probabilities of Selection and Efficacy

| Label | Average Sample Size | Marginal Probability of Selection | | Marginal Probability of Efficacy | |
|-------|---------------------|-----------------------------------|---------|----------------------------------|---------|
| | | Count | % | Count | % |
| 1mg | 124.66 | 2466 | 24.660% | 695 | 6.950% |
| 2.5mg | 179.93 | 7993 | 79.930% | 4456 | 44.560% |
| 10mg | 195.41 | 9541 | 95.410% | 7698 | 76.980% |

The interpretation of first two tables is same as described above. It restores the power to 80% and also gives us the design details when two of the three doses were selected.

29 Two-Stage Multi-arm Designs using p -value combination

The table **Detailed Efficacy Outcomes for all 10000 Simulations** summarizes the number of simulations for which each individual dose group or pairs of doses were selected for Stage 2 and yielded an efficacy conclusion. For example, the pair (2.5mg, 10mg only) was observed to be efficacious in 41% of the trials (4076/10000).

The next table **Marginal Probabilities of Selection and Efficacy**, summarizes the number and percent of simulations in which each dose was selected for Stage 2, regardless of whether it was found significant at end of Stage 2 or not, as well as the number and percent of simulations in which each dose was selected and found significant. Average sample size is also shown. It tells us how frequently the dose (either alone or with some other dose) was selected and efficacious. For example, dose 1mg was selected in approximately 25% trials and was efficacious in approximately 7% trials (which is the sum of 10, 130 and 555 simulations from previous table.)

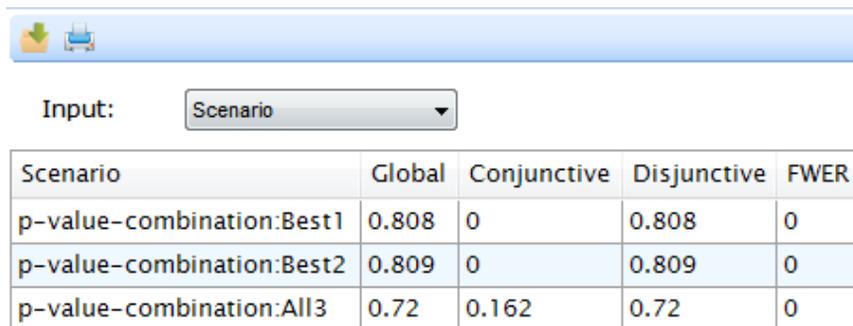
The advantage of 2-stage “treatment selection design” or “drop-the-loser” design is that it allows to drop the less performing/futile arms based on the interim data and still preserves the type-1 error as well as achieve the desired power.

In the **Best1** scenario, we dropped two doses ($r = 1$) and in the **Best2** scenario, we dropped one dose ($r = 2$). Suppose, we had decided to proceed to stage 2 without dropping any doses. In this case, Power would have dropped significantly. To verify this in East, run the above scenario with $r = 3$ and save it to **Library**. Rename this scenario as **All3**. Double click it to see the detailed output. We can observe that the power drops from 80% to 72%.

The three scenarios created so far can be compared in the tabular manner as well.

Select the three nodes in the **Library**, click the  icon in the toolbar and select

“Power” from the dropdown. A table as shown below will be created by East.



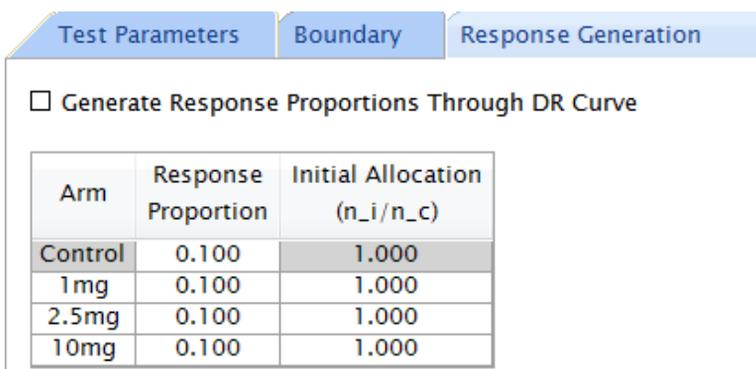
Input: Scenario

| Scenario | Global | Conjunctive | Disjunctive | FWER |
|---------------------------|--------|-------------|-------------|------|
| p-value-combination:Best1 | 0.808 | 0 | 0.808 | 0 |
| p-value-combination:Best2 | 0.809 | 0 | 0.809 | 0 |
| p-value-combination:All3 | 0.72 | 0.162 | 0.72 | 0 |

29.2.4 Simulating under Different Alternatives

Since this is a simulation based design, we can perform sensitivity analyses by changing some of the inputs and observing effects on the overall power and other output. Let us first make sure that this design preserves the total type1 error. It can be done by running the simulations under “Null” hypothesis.

Click the  button on the bottom left corner of the screen. Go to **Response Generation** tab and enter the inputs as shown below:



Generate Response Proportions Through DR Curve

| Arm | Response Proportion | Initial Allocation (n _i /n _c) |
|---------|---------------------|--|
| Control | 0.100 | 1.000 |
| 1mg | 0.100 | 1.000 |
| 2.5mg | 0.100 | 1.000 |
| 10mg | 0.100 | 1.000 |

Also set $r = 2$ in the **Treatment Selection** tab. Run the simulations and go to the

29 Two-Stage Multi-arm Designs using p-value combination

detailed output by saving the row from **Output Preview** to the **Library**. Notice the global power and simulated FWER is less than design type I error which means the overall type I error is preserved.

29.3 Sample Size Re-estimation

As we have seen above, the desired power of 80% is achieved with the sample size of 700 if the initial assumptions ($\pi_c = 0.1, \pi_{1mg} = 0.14, \pi_{2.5mg} = 0.18, \pi_{10mg} = 0.22$) hold true. But if they do not, then the original sample size of 700 may be insufficient to achieve 80% power. The adaptive sample size re-estimation is suited to this purpose. In this approach we start out with a sample size of 700 subjects, but take an interim look after data are available on 400 subjects. The purpose of the interim look is not to stop the trial early but rather to examine the interim data and continue enrolling past the planned 700 subjects if the interim results are promising enough to warrant the additional investment of sample size. This strategy has the advantage that the sample size is finalized only after a thorough examination of data from the actual study rather than through making a large up-front sample size commitment before any data are available. Furthermore, if the sample size may only be increased but never decreased from the originally planned 700 subjects, there is no loss of efficiency due to overruns. Suppose the proportions of response on the four arms are as shown below. Update the **Response Generation** tab accordingly and also set the seed as 100 in the **Simulation Controls** tab.



Generate Response Proportions Through DR Curve

| Arm | Response Proportion | Initial Allocation (n _i /n _c) |
|---------|---------------------|--|
| Control | 0.100 | 1.000 |
| 1mg | 0.130 | 1.000 |
| 2.5mg | 0.170 | 1.000 |
| 10mg | 0.200 | 1.000 |

Run 10000 simulations and save the simulation row to the **Library** by clicking the



icon in the toolbar.

Overall Powers

| | |
|---|------|
| Global (Reject any Hi) | 0.67 |
| Conjunctive (Reject all Hi where $\pi_i > \pi_0$) | 0 |
| Disjunctive (Reject at least one Hi where $\pi_i > \pi_0$) | 0.67 |
| FWER (Reject any Hi where $\pi_i \leq \pi_0$) | 0 |

Lookwise Summary

| Look # | Average Total Sample Size (n) | Trial Termination | |
|---------|-------------------------------|-------------------|---------|
| | | Efficacy | |
| | | Count | % |
| 1 | 400 | 0 | 0.000% |
| 2 | 700 | 6703 | 67.030% |
| OverAll | | 6703 | 67.030% |

Notice that the global power has dropped from 80% to 67%. Let us re-estimate the sample size to achieve the desired power. Add the **Sample Size Re-estimation** tab by



clicking the button . A new tab is added as shown below.

Test Parameters | Boundary | Response Generation | Treatment Selection | **Sample Size Re-estimation**

SSR At: Look # 1

Max. Sample Size if Adapt (multiplier; total #): 1 700

Target CP for Re-estimating Sample Size: 0.9

Promising Zone Scale: (Based on Current Winner) Cond. Power CP

Promising Zone: Min. CP: 0.3 Max. CP: 0.9

SSR Function in Promising Zone: Continuous

SSR At: For a *K*-look group sequential design, one can decide the time at which conditions for adaptations are to be checked and actual adaptation is to be carried out. This can be done either at some intermediate look or after some specified information fraction. The possible values of this parameter depend

29 Two-Stage Multi-arm Designs using p -value combination

upon the user choice. The default choice for this design is always the **Look #**. and is fixed to 1 since it is always a 2-look design.

Target CP for Re-estimating Sample Size: The primary driver for increasing the sample size at the interim look is the desired (or target) conditional power or probability of obtaining a positive outcome at the end of the trial, given the data already observed. For this example we have set the conditional power at the end of the trial to be 80%. East then computes the sample size that would be required to achieve this desired conditional power.

Maximum Sample Size if Adapt (multiplier; total): As just stated, a new sample size is computed at the interim analysis on the basis of the observed data so as to achieve some target conditional power. However the sample size so obtained will be overruled unless it falls between pre-specified minimum and maximum values. For this example, the range of allowable sample sizes is [700, 1400]. If the newly computed sample size falls outside this range, it will be reset to the appropriate boundary of the range. For example, if the sample size needed to achieve the desired 80% conditional power is less than 700, the new sample size will be reset to 700. In other words we will not decrease the sample size from what was specified initially. On the other hand, the upper bound of 1400 subjects demonstrates that the sponsor is prepared to increase the sample size up to double the initial investment in order to achieve the desired 80% conditional power. But if 80% conditional power requires more than 1400 subjects, the sample size will be reset to 1400, the maximum allowed.

Promising Zone Scale: One can define the promising zone as an interval based on conditional power, test statistic, or estimated δ . The input fields change according to this choice. The decision of altering the sample size is taken based on whether the interim value of conditional power / test statistic / δ lies in this interval or not. Let us keep the default scale which is Conditional Power.

Promising Zone: Minimum/Maximum Conditional Power (CP): The sample size will only be altered if the estimate of CP at the interim analysis lies in a pre-specified range, referred to as the “Promising Zone”. Here the promising zone is 0.30 – 0.80. The idea is to invest in the trial in stages. Prior to the interim analysis the sponsor is only committed to a sample size of 700 subjects. If, however, the results at the interim analysis appear reasonably promising, the sponsor would be willing to make a larger investment in the trial and thereby improve the chances of success. Here we have somewhat arbitrarily set the lower bound for a promising interim outcome to be $CP = 0.30$. An estimate

$CP < 0.30$ at the interim analysis is not considered promising enough to warrant a sample size increase. It might sometimes be desirable to also specify an upper bound beyond which no sample size change will be made. Here we have set that upper bound of the promising zone at $CP = 0.80$. In effect we have partitioned the range of possible values for conditional power at the interim analysis into three zones; *unfavorable* ($CP < 0.3$), *promising* ($0.3 \leq CP < 0.8$), and *favorable* ($CP \geq 0.8$). Sample size adaptations are made only if the interim CP falls in the promising zone at the interim analysis. The promising zone defined on the Test Statistic scale or the Estimated δ scale works similarly.

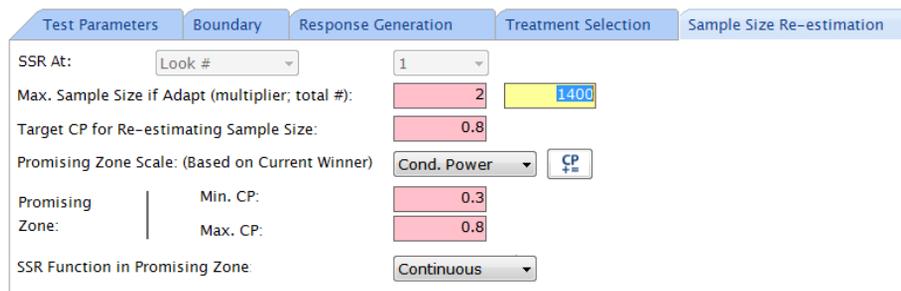
SSR Function in Promising Zone: The behavior in the promising zone can either be defined by a continuous function or a step function. The default is continuous where East accepts the two quantities - (Multiplier, Target CP) and re-estimates the sample size depending upon the interim value of CP/test statistic/effect size. The SSR function can be defined as a step-function as well. This can be done with a single piece or with multiple pieces. For each piece, define the step function in terms of:

- the interval of CP/test statistic/effect size. This depends upon the choice of promising zone scale.
- the value of re-estimated sample size in that interval.
- for single piece, just the total re-estimated sample size is required as an input.

If the interim value of CP/ test statistic/effect size lies in the promising zone then the re-estimation will be done using this step function.

Let us set the inputs on **Sample Size Re-estimation** tab as shown below:

29 Two-Stage Multi-arm Designs using p -value combination



| Test Parameters | | Boundary | Response Generation | Treatment Selection | Sample Size Re-estimation |
|--|-------------|----------|---------------------|---------------------|---------------------------|
| SSR At: | Look # | | 1 | | |
| Max. Sample Size if Adapt (multiplier; total #): | 2 | | | | 1400 |
| Target CP for Re-estimating Sample Size: | 0.8 | | | | |
| Promising Zone Scale: (Based on Current Winner) | Cond. Power | | | | CP |
| Promising Zone: | Min. CP: | 0.3 | | | |
| | Max. CP: | 0.8 | | | |
| SSR Function in Promising Zone: | | | Continuous | | |

Run 10000 simulations and see the Details. Just for the comparison purpose, re-run the simulations but this time, set the multiplier in the **Sample Size Re-estimation** tab to 1 which means we are not interested in sample size re-estimation. Both the scenarios can also be run by entering two values 1, 2 in the cell for Multiplier.

With Sample Size Re-estimation

Overall Powers

| | |
|---|-------|
| Global (Reject any H ₀) | 0.747 |
| Conjunctive (Reject all H ₀ where m _i > π ₀) | 0 |
| Disjunctive (Reject at least one H ₀ where m _i > π ₀) | 0.747 |
| FWER (Reject any H ₀ where m _i ≤ π ₀) | 0 |

Zone-wise Averages

| Zone | Global Power | | Conjunctive Power | | Disjunctive Power | | FWER | | Not Rejecting Any H ₀ | | Total Simulations | | Average Sample Size |
|-------------|--------------|---------|-------------------|--------|-------------------|---------|-------|--------|----------------------------------|---------|-------------------|----------|---------------------|
| | Count | Row % | Count | Row % | Count | Row % | Count | Row % | Count | Row % | Count | Column % | |
| Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| Unfavorable | 56 | 5.534% | 0 | 0.000% | 56 | 5.534% | 0 | 0.000% | 956 | 94.466% | 1012 | 10.120% | 700 |
| Promising | 1325 | 84.773% | 0 | 0.000% | 1325 | 84.773% | 0 | 0.000% | 238 | 15.227% | 1563 | 15.630% | 1375.652 |
| Favorable | 6086 | 81.966% | 0 | 0.000% | 6086 | 81.966% | 0 | 0.000% | 1339 | 18.034% | 7425 | 74.250% | 700 |
| Efficacy | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| All Trials | 7467 | 74.670% | 0 | 0.000% | 7467 | 74.670% | 0 | 0.000% | 2533 | 25.330% | 10000 | 100.000% | 805.604 |

Promising Zone defined as 0.3 ≤ CP < 0.8

Without Sample Size Re-estimation

Overall Powers

| | |
|---|------|
| Global (Reject any H ₀) | 0.67 |
| Conjunctive (Reject all H ₀ where m _i > π ₀) | 0 |
| Disjunctive (Reject at least one H ₀ where m _i > π ₀) | 0.67 |
| FWER (Reject any H ₀ where m _i ≤ π ₀) | 0 |

Zone-wise Averages

| Zone | Global Power | | Conjunctive Power | | Disjunctive Power | | FWER | | Not Rejecting Any H ₀ | | Total Simulations | | Average Sample Size |
|-------------|--------------|---------|-------------------|--------|-------------------|---------|-------|--------|----------------------------------|---------|-------------------|----------|---------------------|
| | Count | Row % | Count | Row % | Count | Row % | Count | Row % | Count | Row % | Count | Column % | |
| Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| Unfavorable | 58 | 5.841% | 0 | 0.000% | 58 | 5.841% | 0 | 0.000% | 935 | 94.159% | 993 | 9.930% | 700 |
| Promising | 501 | 32.852% | 0 | 0.000% | 501 | 32.852% | 0 | 0.000% | 1024 | 67.148% | 1525 | 15.250% | 700 |
| Favorable | 6144 | 82.117% | 0 | 0.000% | 6144 | 82.117% | 0 | 0.000% | 1338 | 17.883% | 7482 | 74.820% | 700 |
| Efficacy | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| All Trials | 6703 | 67.030% | 0 | 0.000% | 6703 | 67.030% | 0 | 0.000% | 3297 | 32.970% | 10000 | 100.000% | 700 |

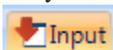
Promising Zone defined as 0.3 ≤ CP < 0.8

We observe from the table the power of adaptive implementation is approximately 75%

which is almost 8% improvement over the non-adaptive design. This increase in power has come at an average cost of 805-700 = 105 additional subjects. Next we observe from the **Zone-wise Averages** table that 1563 of 10000 trials (16%) underwent sample size re-estimation and of those 1563 trials, 84% were able to reject the Global null hypothesis. The average sample size, conditional on adaptation is 1376.

29.4 Adding Early Stopping Boundaries

One can also incorporate stopping boundaries to stop at the interim early for efficacy or futility. The efficacy boundary can be defined based on **Adjusted p-value** scale whereas futility boundary can be on **Adjusted p-value** or δ scale.



Click the  button on the bottom left corner of the screen. This will take you back to the input window of the last simulation scenario. Go to **Boundary** tab and set Efficacy and Futility boundaries to “Adjusted p-value”. These boundaries are for

29 Two-Stage Multi-arm Designs using p-value combination

early stopping at look1. As the note on this tab says:

- If any one adjusted p-value is \leq efficacy p-value boundary then stop the trial for efficacy
- If only all the adjusted p-values are $>$ futility p-value then stop the trial for futility. Else carry forward all the treatments to the next step of treatment selection.

Stopping early for efficacy or futility is step which is carried out before treatment selection rules are applied. The simulation output has the same explanation as above except the **Lookwise Summary** table may have some trials stopped at the first look due to efficacy or futility.

29.5 Monitoring this trial

Select the simulation node with **SSR** implementation and click the **IM** icon. It will invoke the **Interim Monitoring** dashboard. Click the **Enter Interim Data** icon to open the **Test Statistic Calculator**. Enter the data as shown below:

| Arm | Sample Size | Proportion | Test Statistic | Raw p-values |
|---------|-------------|------------|----------------|--------------|
| Control | 100 | 0.09 | NA | NA |
| 1mg | 100 | 0.15 | 1.311 | 0.095 |
| 2.5mg | 100 | 0.17 | 1.694 | 0.045 |
| 10mg | 100 | 0.18 | 1.879 | 0.030 |

Click **Recalc** to calculate the test statistic as well as the raw p-values. Notice that the p-value for *1mg* is 0.095 which is greater than 0.025. We will drop this dose in the second stage. On clicking **OK**, it updates the dashboard.

| Look # | Number of Arms | Adjusted p-value for Best Treatment | Prespecified Weights | Prespecified Efficacy Boundary (p-value) |
|--------|----------------|-------------------------------------|----------------------|--|
| 1 | 4 | 0.09 | 0.756 | NA |
| 2 | | | 0.655 | 0.025 |

Click the "Edit Interim Data" button to edit the Look # 1 data.

Stopping Boundaries

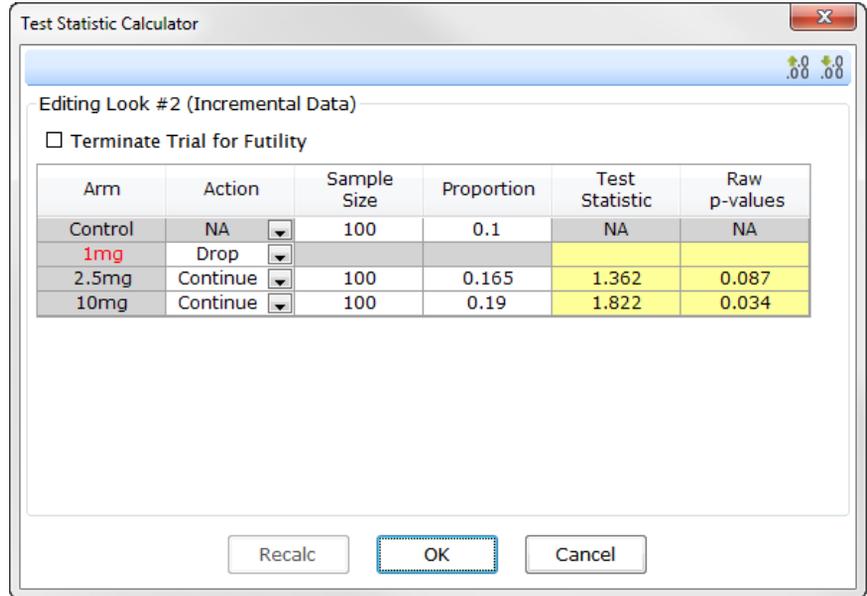
The plot shows a grid with x-axis from 0 to 3 and y-axis from 0 to 0.1. A green dot is at (1, 0.09) and a blue dot is at (2, 0.025).

| Look # | Efficacy (p-value) |
|--------|--------------------|
| 1 | NA |
| 2 | 0.025 |

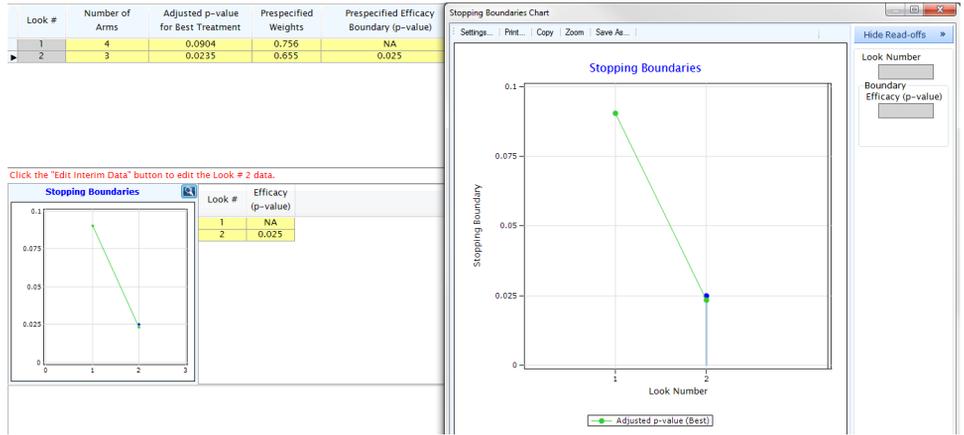
Open the test statistic calculator for the second look and enter the following information and also drop the dose 1mg. Click **Recalc** to calculate the test statistic as

29 Two-Stage Multi-arm Designs using p-value combination

well as the raw p-values.



On clicking **OK**, it updates the dashboard. Observe that the adjusted p-value for *10mg* crosses the efficacy boundary. It can also be observed in the **Stopping Boundaries** chart.



30 *Binomial Superiority Regression*

30.1 *Logistic Regression with Single Normal Covariate*

Logistic regression is widely used for modeling the probability of a binary response in the presence of covariates. In this section we will show how East may be used to design clinical trials with binomial endpoints, while adjusting for the effects of covariates through the logistic regression model. The sample size calculations for the logistic regression models discussed here and implemented in East are based on the methods of Hsieh et al., 1997. We note, however, that these methods are limited to continuous covariates only. When the covariate is normal, the log odds value β_1 is zero if and only if the group means between the two response categories are the same assuming equal variances.

Suppose in a logistic regression model, Y is a binary response variable and X_1 is a covariate related to Y . The model is given by

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 \tag{30.1}$$

where $P = P(Y = 1)$. The null hypothesis that the coefficient of the covariate β_1 is zero is tested against the two sided alternative hypothesis that β_1 is not equal to zero. The slope coefficient β_1 is the change in log odds for every one unit increase in X_1 . The sample size required for a two sided test with type-I error rate of α to have a power $1 - \beta$ is

$$n = \frac{(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2}{P_1(1 - P_1)\beta^{*2}} \tag{30.2}$$

Where β^* is the effect size to be tested, P_1 is the event rate at the mean of X and Z_u is the upper u -th percentile of the standard normal distribution.

30.1.1 *Trial Design*

We use a Department of Veterans Affairs Cooperative Study entitled 'A Psychophysiological Study of Chronic Post-Traumatic Stress Disorder' to illustrate the preceding sample size calculation for logistic regression with continuous covariates. The study developed and validated a logistic regression model to explore the use of certain psychophysiological measurements for the prognosis of combat-related

post-traumatic stress disorder (PTSD). In the study, patients' four psychophysiological measurements-heart rate, blood pressures, EMG and skin conductance- were recorded while patients were exposed to video tapes containing combat and neutral scenes. Among the psychophysiological variables, the difference of the heart rates obtained while viewing the combat and the neutral tapes (DCNHR) is considered a good predictor of the diagnosis of PTSD. The prevalence rate of PTSD among the Vietnam veterans was assumed to be 20 per cent. Therefore, we assumed a four to one sample size ratio for the non-PTSD versus PTSD groups. The effect size of DCNHR is approximately 0.3 which is the difference of the group means divided by the standard deviation. We would like to determine the sample size to achieve 90% power based on a two-sided test at significance level 0.05 (Hsieh et. al.,1998).

Start East. Click **Design** tab, then click **Regression** in the **Discrete** group, and then click **Logistic Regression - Odds Ratio**.

The input dialog box, with default input values will appear in the upper pane of this window. Enter 0.2 in **Proportion Success at $X = \mu, (P_0)$** and 1.349 in **Odds Ratio $P_1(1 - P_0)/P_0(1 - P_1)$** field.

Enter the rest of the inputs as shown below and click **Compute**.

Test Parameters

| | | | |
|----------------------------|------------|---|-------|
| Test Type: | 2-Sided ▾ | Proportion Successes at $X = \mu, (P_0)$: | 0.2 |
| Type I Error (α): | 0.05 ○ | Proportion Successes at $X = \mu + \sigma, (P_1)$: | 0.252 |
| Power: | 0.9 ○ | Odds Ratio[$P_1(1 - P_0)/[P_0(1 - P_1)]$]: | 1.349 |
| Sample Size (n): | Computed ⊕ | <input type="checkbox"/> Other Covariates | |

The design output will be displayed in the Output Preview, with the computed sample size highlighted in yellow. A total of 733 subjects must be enrolled in order to achieve 90% power under the alternative hypothesis. Besides sample size, one can also compute the power and the level of significance for this design.

| | ID | Test Type | Specified α | Power | Sample Size | Δ | Prop P0 | Prop P1 | Odds Ratio |
|---|------|-----------|--------------------|-------|-------------|----------|---------|---------|------------|
| π | Des1 | 2-Sided | 0.05 | 0.9 | 733 | 0.299 | 0.2 | 0.252 | 1.349 |

You can select this design by clicking anywhere on the row in the Output Preview. If

30 Binomial Superiority Regression

you click on  icon, some of the design details will be displayed in the upper pane. In the Output Preview toolbar, click the  icon, to save this design to workbook Wbk1 in the **Library**. If you hover the cursor over Des1 in the **Library**, a tooltip will appear that summarizes the input parameters of the design.

| Wbk1:Des1 | |
|--------------------------|----------|
| Mnemonic | PN-RG-LR |
| Test Parameters | |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Effect Size (Δ) | 0.299 |
| Proportion Successes P0 | 0.2 |
| Proportion Successes P1 | 0.252 |
| Odds Ratio | 1.349 |
| Sample Size | |
| Maximum | 733 |

With Des1 selected in the **Library**, click  icon to see the detailed output as shown below:

Design: Discrete Endpoint: Regression Model - Parallel Design - Logistic Regression - Odds Ratio

| Test Parameters | |
|--|----------------|
| Design ID | Des1 |
| Test Type: | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Sample Size (n) | To be Computed |
| Model Parameters | |
| Proportion Successes at $X=\mu$, (P_0) | 0.2 |
| Proportion Successes at $X=\mu + \sigma$, (P_1) | 0.252 |
| $\Psi = [P_1(1-P_0)] / [P_0(1-P_1)]$ | |
| Under H0 | 1 |
| Under H1 | 1.349 |
| Odds Ratio | 1.349 |
| Coefficient ($B=\ln(\text{Odds Ratio})$) | 0.299 |

Sample Size Information

Sample Size (n) 733

Critical Points

Lower Critical Point -1.96
 Upper Critical Point 1.96

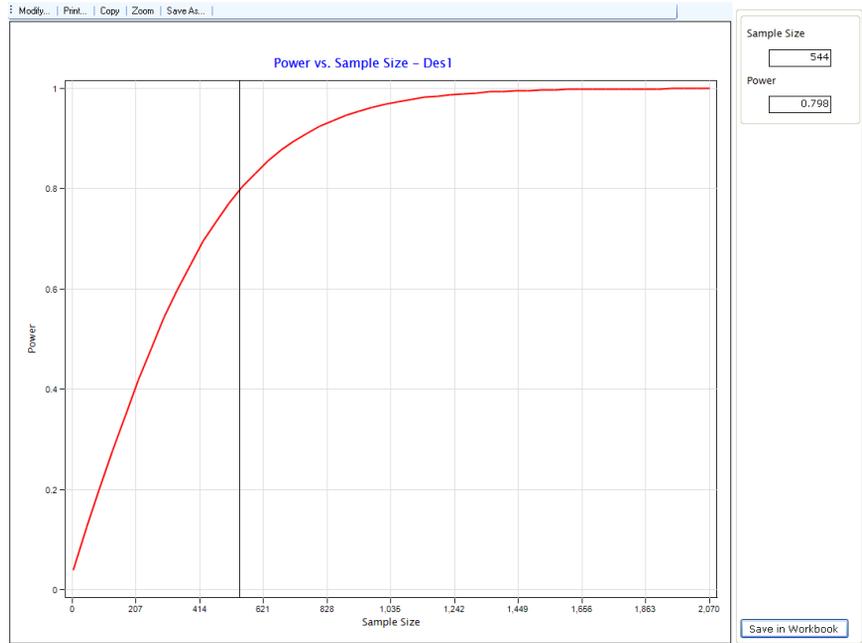
Summary

A total sample size of 733 is required in a study to detect a significant effect of the continuous covariate on response proportions between the treatment and control group with desired power 90.01%. The test is based on a 2-Sided test for significance of the covariate in a logistic regression with significance level 0.05. The proportion of successes expected at a covariate value equal to its mean is 0.2. The proportion of successes expected at a covariate value equal to its mean plus one standard deviation is 0.252. The odds ratio is 1.349 and the log odds ratio is 0.299, equal to the coefficient of the covariate in the logistic regression under the alternative hypothesis.

Observe that this kind output gives us the summary of the output as well.

With Des1 selected in the **Library**, click  icon, on the **Library** toolbar, and then click **Power vs. Sample Size**. The resulting power curve for this design is shown. You can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As....** For now, you may close the chart before continuing

30 Binomial Superiority Regression



31 Agreement

31.1 Cohen's Kappa

31.1.1 Trial Design

In some experimental situations, to check inter-rater reliability, independent sets of measurements are taken by more than one rater and the responses are checked for agreement. For a binary response, Cohen's Kappa test for binary ratings can be used to check inter-rater reliability. Conventionally, the kappa coefficient is used to express the degree of agreement between two raters when the same two raters rate each of a sample of n subjects independently, with the ratings being on a categorical scale consisting of k categories (Fleiss, 1981). A simple example is given in the below table where two tests **Test 1** and **Test 2** ($k = 2$) were performed. In the below table, π_{ij} denotes the true population proportion in the i -th row and the j -th column category.

Table 31.1: Table of proportions of two raters

| Test 1 \ Test 2 | Test 2(+) | Test 2(-) | Marginal Probability |
|----------------------|------------|------------|----------------------|
| Test 1(+) | π_{11} | π_{12} | $\pi_{1.}$ |
| Test 1(-) | π_{21} | π_{22} | $\pi_{2.}$ |
| Marginal Probability | $\pi_{.1}$ | $\pi_{.2}$ | 1 |

The Kappa coefficient (κ) is defined by

$$\kappa = \frac{\pi_0 - \pi_e}{1 - \pi_e} \tag{31.1}$$

where $\pi_0 = \sum_{i=1}^2 \pi_{ii}$ and $\pi_e = \sum_{i=1}^2 \pi_{i.}\pi_{.i}$.

We want to test the null hypothesis $H_0 : \kappa \leq \kappa_0$ against $H_1 : \kappa > \kappa_0$ where $\kappa_0 > 0$. The total sample size required for a test with type-I error rate of α to have a power $1 - \beta$ is

$$n = \frac{(z_\alpha + z_\beta)^2 (E + F - G)}{[(1 - \pi_e)^2 (\kappa - \kappa_0)]^2} \tag{31.2}$$

31 Agreement

where

$$E = \sum_{i=1}^2 \pi_{ii} [(1 - \pi_e) - (\pi_{.i} + \pi_{i.})(1 - \pi_0)]^2 \quad (31.3)$$

$$F = (1 - \pi_0)^2 \sum_{i=1}^2 \sum_{j \neq i} \pi_{ij} (\pi_{.i} + \pi_{j.})^2 \quad (31.4)$$

and

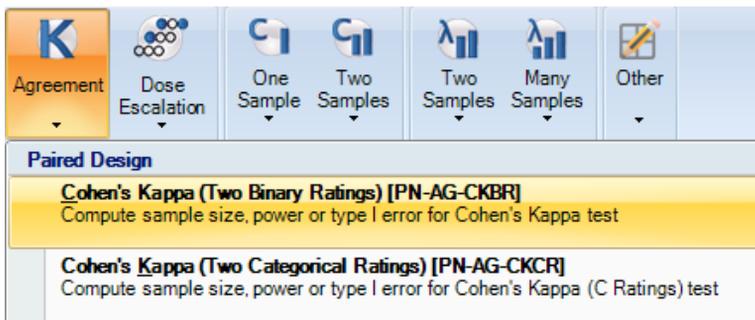
$$G = [\pi_0(1 + \pi_e) - 2\pi_e]^2 \quad (31.5)$$

We can calculate power, sample size or level of significance for your Cohen's Kappa test for two ratings.

31.1.1 Trial Design

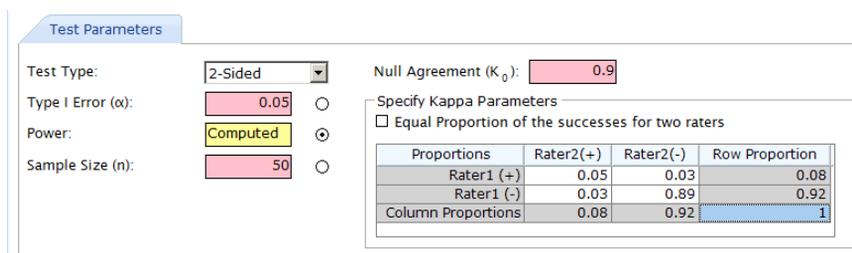
Consider responses from two raters. The example is based on a study to develop and validate a set of clinical criteria to identify patients with minor head injury who do not undergo a CT scan (Haydel, et al., 2000). In the study, CT scan was first reviewed by a staff neuroradiologist. An independent staff radiologist then reviewed 50 randomly selected CT scans and the two sets of responses checked for agreement. Let κ denote the level of agreement. The null hypothesis is $H_0 : \kappa = 0.9$ versus the one-sided alternative hypothesis $H_1 : \kappa < 0.9$. We wish to compute the power of the test at the alternative value $\kappa_1 = 0.6$. We expect each rater to identify 8% of CT scans to be positive. Also we expect 5% of the positive CT scans were rated by both the raters.

Start East. Click **Design** tab, then click Agreement in the **Discrete** group, and then click **Cohen's Kappa (Two Binary Ratings)** .



The input dialog box, with default input values will appear in the upper pane of this window.

Enter 0.9 in **Null Agreement** (κ_0) field. Specify the $\alpha = 0.05$, sample size and the kappa parameter values as shown below. Enter the rest of the inputs as shown below and click **Compute**.



The design output will be displayed in the Output Preview, with the computed power highlighted in yellow. The power of the test is 64.9% given a sample size of 50 scans to establish agreement of ratings by the two radiologists. Besides power, one can also compute the sample size for this study design.

| ID | Test Type | Specified α | Power | Sample Size | Prop Rater 1 | Prop Rater 2 | Kappa 0 | Kappa 1 |
|------|-----------|--------------------|-------|-------------|--------------|--------------|---------|---------|
| Des2 | 2-Sided | 0.05 | 0.649 | 50 | 0.08 | 0.08 | 0.9 | 0.592 |

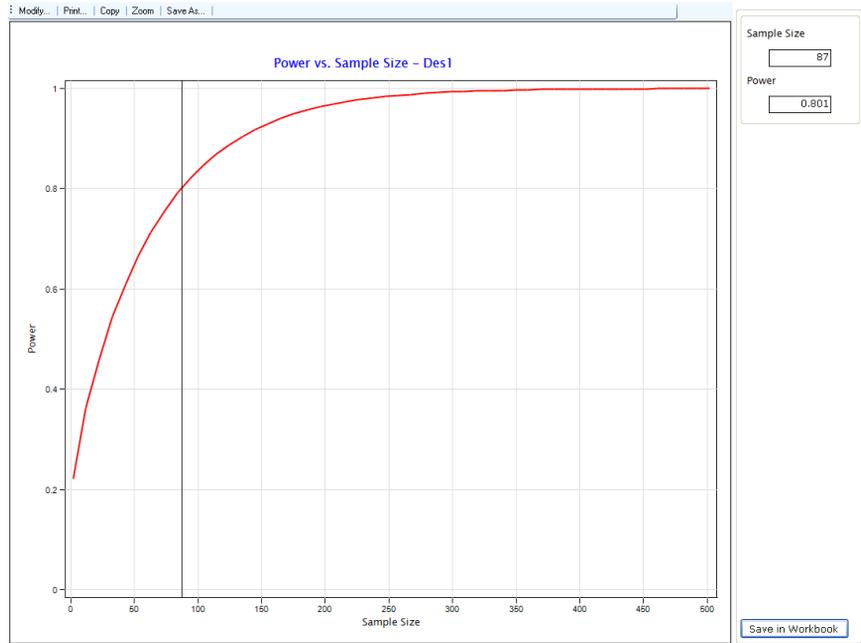
You can select this design by clicking anywhere on the row in the Output Preview. If you click  icon, some of the design details will be displayed in the upper pane. In

31 Agreement

the Output Preview toolbar, click  icon, to save this design to workbook Wbk1 in the **Library**. If you hover the cursor over Des1 in the **Library**, a tooltip will appear that summarizes the input parameters of the design.

| | Des 1 |
|-------------------------|--------------|
| Mnemonic | PN-AG-CKBR |
| Test Parameters | |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.649 |
| Model Parameters | |
| Proportions Rater 1 | 0.08 |
| Proportions Rater 2 | 0.08 |
| Kappa 0 | 0.9 |
| Kappa 1 | 0.592 |
| Sample Size | |
| Maximum | 50 |

With Des1 selected in the **Library**, click  icon on the **Library** toolbar, and then click **Power vs. Sample Size**. The resulting power curve for this design is shown. You can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As....** For now, you may close the chart before continuing



31.2 Cohen's Kappa (C Ratings)

Let κ denotes the measure of agreement between two raters who each classify n objects into C mutually exclusive ratings (categories). Here the null hypothesis is $H_0 : \kappa = \kappa_0$ is tested against two-sided hypothesis $H_1 : \kappa \neq \kappa_0$ or one sided hypothesis $H_1 : \kappa > \kappa_0$ or $H_1 : \kappa < \kappa_0$. The total sample size required for a test with type-I error rate of α to have a power $1 - \beta$ when $\kappa = \kappa_1$ is

$$n \geq \left\lceil \frac{Z_{1-\alpha} \max \tau(\hat{\kappa} | \kappa = \kappa_0) + Z_{1-\beta} \max \tau(\hat{\kappa} | \kappa = \kappa_1)}{\kappa_1 - \kappa_0} \right\rceil \quad (31.6)$$

Where

$$\tau(\hat{\kappa}) = \frac{(Q_1 + Q_2 - 2Q_3 - Q_4)^{\frac{1}{2}}}{(1 - \pi_e)^2} \quad (31.7)$$

and

31 Agreement

$$\begin{aligned}
 Q_1 &= \pi_0(1 - \pi_e)^2, \\
 Q_2 &= (1 - \pi_0)^2 \sum_{i=1}^C \sum_{j=1}^C \pi_{ij}(\pi_i + \pi_j)^2, \\
 Q_3 &= 2(1 - \pi_0)(1 - \pi_e) \sum_{i=1}^C \pi_{ij}(\pi_i + \pi_j), \\
 Q_4 &= (\pi_0\pi_e - 2\pi_e + \pi_0)^2.
 \end{aligned}$$

π_{ij} is the proportion of subjects that Rater 1 places in category i but Rater 2 places in category j , π_0 is the proportion of agreement and π_e is the expected proportion of agreement.

The power of the test is given by

$$\text{Power} = \Phi\left[\frac{\sqrt{n}(\kappa_1 - \kappa_0) - Z_{1-\alpha} \max \tau(\hat{\kappa}|\kappa = \kappa_0)}{\max \tau(\hat{\kappa}|\kappa = \kappa_1)}\right] \quad (31.8)$$

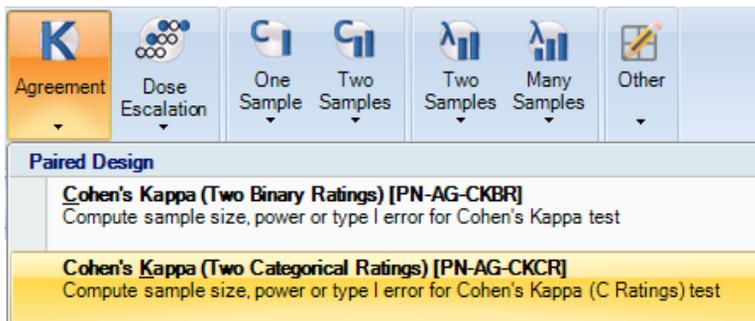
31.2.1 Trial Design

Consider a hypothetical problem of physical health ratings from two different raters-health instructor and subject's general practitioner. 360 subjects were randomly selected and the two sets of responses were checked for agreement. Let κ denote the level of agreement. The null hypothesis is $H_0 : \kappa = 0.6$ versus the one-sided alternative hypothesis $H_1 : \kappa < 0.6$. We wish to compute the power of the test at the alternative value $\kappa_1 = 0.5$.

Table 31.2: Table: Contingency Table

| General Petitioner \ Health Instructor | Poor | Fair | Good | Excellent | Total |
|--|------|------|------|-----------|-------|
| Poor | 2 | 12 | 8 | 0 | 22 |
| Fair | 9 | 35 | 43 | 7 | 94 |
| Good | 4 | 36 | 103 | 40 | 183 |
| Excellent | 1 | 8 | 30 | 22 | 61 |
| Total | 16 | 91 | 184 | 69 | 360 |

Start East. Click **Design** tab, then click Agreement in the **Discrete** group, and then click **Cohen's Kappa (Two Categorical Ratings)**.



The input dialog box, with default input values will appear in the upper pane of this window.

Enter **Number of Ratings (C)** as 4. Enter 0.6 in **Null Agreement (κ_0)** field and 0.5 in **Alternative Agreement (κ_1)**. Click **Marginal Probabilities** and specify the marginal probabilities calculated from the above table. Specify the **sample size**. Leave all other values as defaults, and click **Compute**.

Test Parameters

Test Type: 2-Sided

Type I Error (α): 0.05

Power: Computed

Sample Size (n): 360

Number of Ratings (C): 4

Specify Kappa Parameters

Alternative Agreement (κ_1): 0.5

Null Agreement (κ_0): 0.6

Marginal Probabilities

Cell Probabilities

| Ratings | 1 | 2 | 3 | 4 | Marginal Prop |
|---------------|-------|-------|-------|-------|---------------|
| 1 | 0.044 | 0 | 0 | 0 | 0.044 |
| 2 | 0 | 0.094 | 0.159 | 0 | 0.253 |
| 3 | 0 | 0.159 | 0.352 | 0 | 0.511 |
| 4 | 0 | 0 | 0 | 0.192 | 0.192 |
| Marginal Prop | 0.044 | 0.253 | 0.511 | 0.192 | 1 |

The design output will be displayed in the Output Preview, with the computed power highlighted in yellow. The power of the test is 73.3% given a sample size of 360 subjects to establish agreement of ratings by the two raters. Besides power, one can also compute the sample size for this study design.

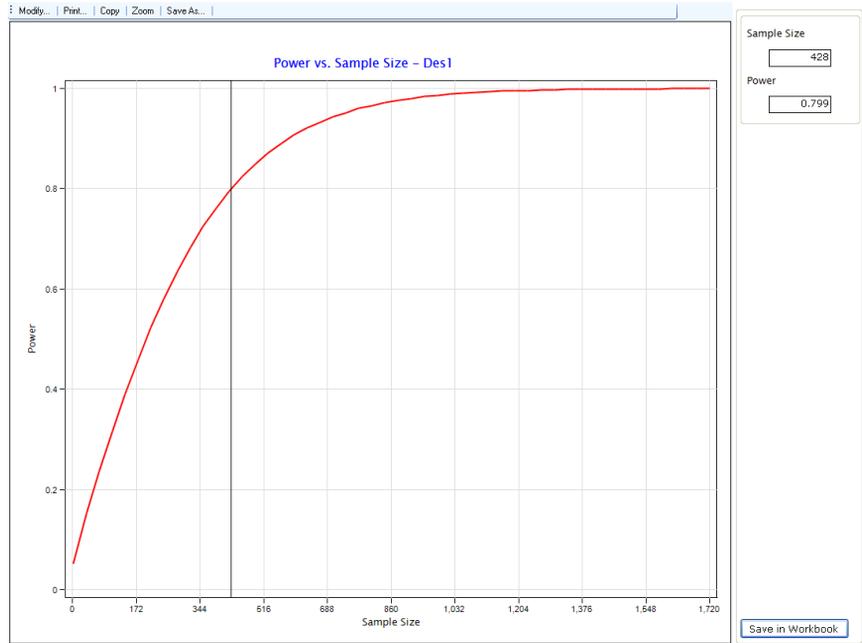
| ▲ | ID | Test Type | Specified α | Power | Sample Size | Kappa 0 | Kappa 1 | Kappa Rating |
|---|------|-----------|--------------------|-------|-------------|---------|---------|--------------|
| K | Des2 | 2-Sided | 0.05 | 0.733 | 360 | 0.6 | 0.5 | 4 |

31 Agreement

You can select this design by clicking anywhere on the row in the Output Preview. If you click  icon, some of the design details will be displayed in the upper pane. In the Output Preview toolbar, click  icon to save this design to workbook Wbk1 in the **Library**. If you hover the cursor over Des1 in the **Library**, a tooltip will appear that summarizes the input parameters of the design.

| | Des2 |
|-------------------------|-------------|
| Mnemonic | PN-AG-CKCR |
| Test Parameters | |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.733 |
| Model Parameters | |
| Kappa Rating | 4 |
| Kappa 0 | 0.6 |
| Kappa 1 | 0.5 |
| Sample Size | |
| Maximum | 360 |

With Des1 selected in the **Library**, click  icon on the **Library** toolbar, and then click **Power vs. Sample Size**. The resulting power curve for this design is shown. You can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...** For now, you may close the chart before continuing.



32

Dose Escalation

This chapter deals with the design, simulation, and interim monitoring of Phase 1 dose escalation trials. A brief overview of the designs is given below; more technical details are available in the Appendix N.

One of the primary goals of Phase I trials in oncology is to find the maximum tolerated dose (MTD). Currently, the vast majority of such trials have employed traditional dose escalation methods such as the 3+3 design. The 3+3 design starts by allocating three patients typically to the lowest dose level, and then adaptively moves up and down in subsequent cohorts until either the MTD is obtained, or the trial is stopped for excessive toxicity. In addition to the 3+3, East also provides the Continual Reassessment Method (CRM), the modified Toxicity Probability Interval (mTPI) method, and the Bayesian logistic regression model (BLRM) for single agent designs. Compared to the 3+3, these modern methods may offer a number of advantages, which can be explored systematically via simulation and interim monitoring.

The CRM (Goodman et al., 1995; O’Quigley et al., 1990) is a Bayesian model-based method that uses all available information from all doses to guide dose assignment. One first specifies a target toxicity, a one-parameter dose response curve and corresponding prior distribution. The posterior mean and predictions for the probability of toxicity at each dose are updated as the trial progresses. The next recommended dose is the one whose toxicity probability is closest to the target toxicity.

The mTPI method (Ji et al., 2010) is Bayesian like the CRM, but rule-based like the 3+3. In this way, the mTPI represents a useful compromise between the other methods. An independent beta distribution is assumed for the probability of toxicity at each dose. A set of decision intervals are specified, and subsequent dosing decisions (up, down, or stay) are determined by computing the normalized posterior probability in each interval at the current dose. The normalized probability for each interval is known as the unit probability mass (UPM).

A more advanced version of the CRM is the BLRM (Neuenschwander et al., 2008; Sweeting et al., 2013), which assumes a two-parameter logistic dose response curve. In addition to a target toxicity, one specifies a set of decision intervals, and optional associated losses, for guiding dosing decisions.

For dual-agent combination designs, East provides a combination version of the BLRM (Neuenschwander et al., 2014), as well as the PIPE (product of independent beta probabilities escalation) method (Mander & Sweeting, 2015).

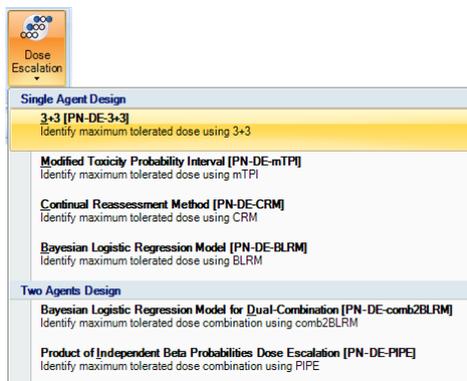
32.1 3+3

32.1.1 Simulation

32.1.2 Interim Monitoring

32.1.1 Simulation

Click **Discrete: Dose Escalation** on the Design tab, and then click **Single Agent Design: 3+3**.



This window is the Input dialog box, which is separated into three tabs: **Design Parameters**, **Response Generation**, and **Simulation Control**. First, you may specify the **Max. Number of Doses** as 7.

In the **Design Parameters** tab, enter 30 as the **Max. Sample Size**. For the 3+3 design, the **Cohort Size** is fixed at 3.

There are three variants of 3+3 offered: L and H and L(modified). The key differences between these variants can be seen in the respective **Decision Rules** table. Select 3+3

32 Dose Escalation

L.

Max. Number of Doses:

Design Parameters
Response Generation
Simulation Controls

Max. Sample Size:

Cohort Size:

Start With Accelerated Titration

Decision Rules

3+3 variant:

Target Probability of Toxicity (P_T): 1/6

| # Subjects with DLTs at Current Dose D_i | Decision 3+3 L |
|--|--|
| 0/3 | Escalate to $D_{[i+1]}$ |
| 1/3 | Stay at $D_{[i]}$ & treat 3 more |
| 2/3 or 3/3 | De-escalate to $D_{[i-1]}$; If $D_{[i-1]}$ already has 6 subjects, stop & choose $D_{[i-1]}$ as MTD |
| 0/3 + {0/3 or 1/3} | Stop & choose $D_{[i]}$ as MTD |
| 1/3 + 0/3 | Escalate to $D_{[i+1]}$ |
| {1/3 + 1/3} or {0/3 + 2/3} | De-escalate to $D_{[i-1]}$; If $D_{[i-1]}$ already has 6 subjects, stop & choose $D_{[i-1]}$ as MTD |
| 1/3 + {2/3 or 3/3} | De-escalate to $D_{[i-1]}$; If $D_{[i-1]}$ already has 6 subjects, stop & choose $D_{[i-1]}$ as MTD |

You also have the option of starting with an Accelerated Titration design (Simon et al., 1997), which escalates with single-patient cohorts until the first DLT is observed, after which the cohort is expanded at the current dose level with two more patients.

In the **Response Generation** tab, you can specify a set of true dose response curves from which to simulate. For the **Starting Dose**, select the lowest dose (5). In the row titled *Dose*, you can specify the dose levels (e.g., in mg). In the row titled *GCI*, you can edit the true probabilities of toxicity at each dose. You can also rename the profile by directly editing that cell. For now, leave all entries at their default values.

You can add a new profile generated from a parametric curve family. For example, click on the menu **Curve Family** and select **Emax**. You may construct a

four-parameter Emax curve by adjusting its parameters, then click **Add Profile**.

Specify True Probability of Toxicity

Curve Family: **Emax**

Parameters

| | | | |
|------|------|------|------|
| E0 | EMax | ED50 | Hill |
| 0.05 | 0.75 | 50 | 1 |

Add Profile

True Probability of Toxicity (Profiles)

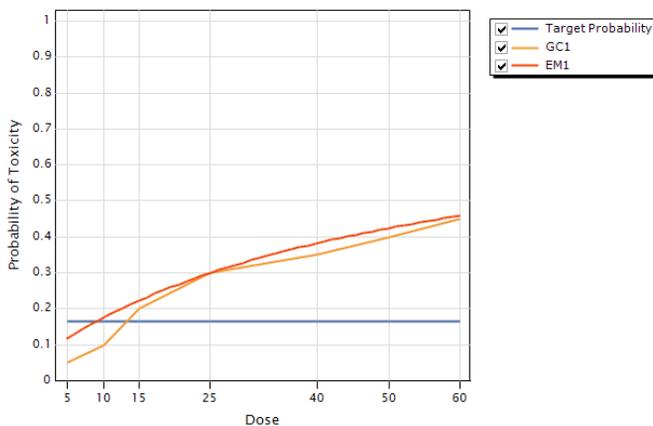
| Profile\Dose ID | D1 | D2 | D3 | D4 | D5 | D6 | D7 |
|-----------------|-------|-------|-------|-----|-------|-------|-------|
| Dose | 5 | 10 | 15 | 25 | 40 | 50 | 60 |
| GC1 | 0.05 | 0.1 | 0.2 | 0.3 | 0.35 | 0.4 | 0.45 |
| EM1 | 0.118 | 0.175 | 0.223 | 0.3 | 0.383 | 0.425 | 0.459 |

Starting Dose: **5**

Plot Profiles

Click **Plot Profiles** to plot the two dose toxicity curves in this grid.

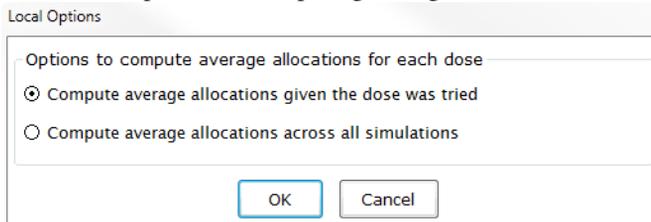
Dose Toxicity Curve



In the **Simulation Control** tab, check the boxes corresponding to **Save summary statistics** and **Save subject-level data**. These options will provides access to several charts derived from these more detailed levels of simulated data. If you wish to display subject-level plots for more than one simulation, you can increase the number. For now, leave this at 1 to save computation time.

32 Dose Escalation

You may also like to examine the **Local Options** button of the input window toolbar. This gives you different options for computing average allocations for each dose.



Click **Simulate**. East will simulate data generated from the two profiles you specified, and apply the 3+3 design to each simulation data set. Once completed, the two simulations will appear as two rows in the **Output Preview**. Select both rows in the **Output Preview** and click the  icon in the toolbar. The two simulations will be displayed side by side in the **Output Summary**.

| | Sim1 | Sim2 |
|-----------------------------------|-----------|-----------|
| Mnemonic | PN-DE-3+3 | PN-DE-3+3 |
| Simulation Parameters | | |
| Max. Sample Size | 30 | 30 |
| No. of Doses | 7 | 7 |
| Cohort Size | 3 | 3 |
| Starting Dose | 5 | 5 |
| Variant | 3+3 L | 3+3 L |
| Response Generation | | |
| Curve Family | General | Emax |
| Curve Label | GC1 | EM1 |
| Summary Statistics | | |
| Median Sample Size | 18 | 15 |
| Mean Sample Size | 17.658 | 15.528 |
| Median No. of DLTs | 3 | 3 |
| Mean No. of DLTs | 3.186 | 3.196 |
| Median Prop. of DLTs | 0.167 | 0.2 |
| Mean Prop. of DLTs | 0.188 | 0.232 |
| True MTD Analysis | | |
| True MTD | 10 | 5 |
| % Sims. True MTD Selected | 27.5 | 22.4 |
| Avg. Allocs. at True MTD | 4.401 | 4.401 |
| Avg. No. of DLTs at True MTD | 0.419 | 0.519 |
| Avg. Prop. of Allocs. at True MTD | 0.271 | 0.381 |
| Avg. Prop. of DLTs at True MTD | 0.076 | 0.097 |
| MTD Analysis | | |
| Median MTD | 15 | 10 |
| Mean MTD | 16.469 | 14.337 |
| Mode of MTD | 15 | 5 |
| Mode (%) | 32.7 | 22.4 |
| % of MTD Below Lowest Dose | 3 | 13.6 |
| % of MTD Above Highest Dose | 1.3 | 0.5 |

In the **Output Preview** toolbar, click the  icon to save both simulations to the

Library. Double-click Sim1 in the **Library** to display the simulation output details.

Simulation: Discrete Endpoint: Dose Escalation - 3 + 3

| Simulation Parameters | |
|-----------------------|---------|
| Simulation ID | Sim1 |
| Maximum Sample Size | 30 |
| Cohort Size | 3 |
| Starting Dose | 5 |
| Variant | 3+3 L |
| Response Generation | |
| Number of Doses | 7 |
| Curve Family | General |
| Curve Label | GC1 |
| Simulation Control | |
| Starting Seed | Fixed |
| Number of Simulations | 1000 |

Decision Rules

| Subjects | Toxicity | Decision |
|----------|----------|-------------|
| 3 | 0 | Escalate |
| 3 | 1 | Stay |
| 3 | >=2 | De-escalate |
| 6 | 0 | MTD |
| 6 | 1 | Escalate |
| 6 | 2 | De-escalate |
| 6 | >2 | De-escalate |

Summary Statistics

| | Min. | Q1 | Median | Mean | Q3 | Max. | Std. Dev. |
|---------------|------|-------|--------|--------|-------|-------|-----------|
| Sample Size | 3 | 15 | 18 | 17.658 | 21 | 30 | 5.191 |
| No. of DLTs | 0 | 2 | 3 | 3.186 | 4 | 8 | 1.153 |
| Prop. of DLTs | 0 | 0.143 | 0.167 | 0.188 | 0.208 | 0.667 | 0.069 |

MTD Analysis

| | True | Median | Mean | Mode |
|-----|------|--------|--------|------|
| MTD | 10 | 15 | 16.469 | 15 |

% of Simulations selecting each dose as MTD

| Dose ID | Below lowest dose | D1 | D2 | D3 | D4 | D5 | D6 | D7 | Above highest dose |
|------------|-------------------|-----|------|------|------|-----|-----|----|--------------------|
| Dose | | 5 | 10 | 15 | 25 | 40 | 50 | 60 | |
| Percentage | 3 | 9.7 | 27.5 | 32.7 | 17.2 | 6.1 | 1.4 | 0 | 1.3 |

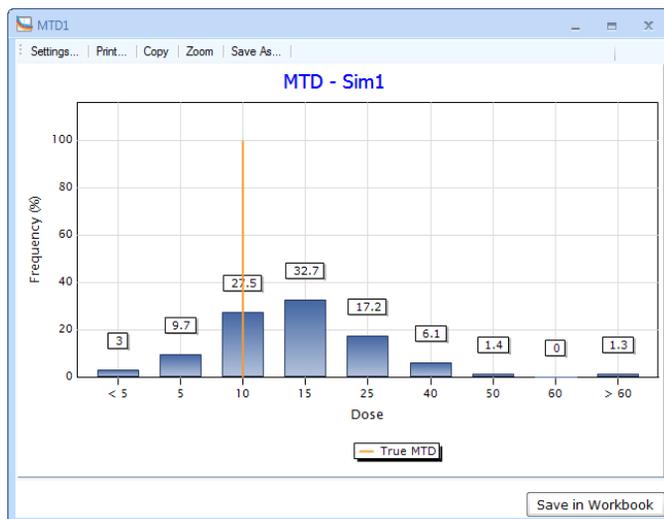
- % Simulations MTD selected
- using decision rules 94.6
- % Simulations MTD Not selected
- below lowest dose: 3
- above highest dose: 1.3
- inadequate sample size 1.1

Dose-wise Summary

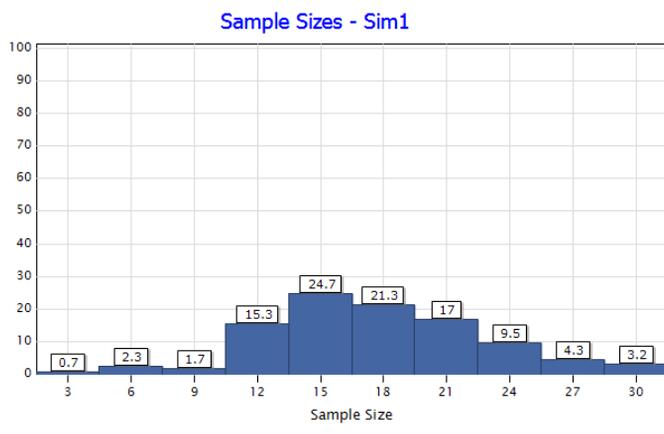
With Sim1 selected in the **Library**, click the Plots icon to access a wide range of available plots. Below is an example of the MTD plot, showing the percentage of simulations that each dose level was selected as the MTD. The "true" MTD is displayed as the second dose level. This is the dose whose true probability of DLT

32 Dose Escalation

(0.1) was closest to and below the target probability (1/6).



Another useful plot is that showing the possible sample sizes, shown as percentages over all simulations.



Close each plot after viewing, or save them by clicking **Save in Workbook**.

Finally, to save the workbook to disk, right-click Wbk1 in the **Library** and then **Save**

As...

32.1.2 Interim Monitoring

Right-click one of the Simulation nodes in the Library, and select **Interim Monitoring**. This will open an empty interim monitoring dashboard.

| Cohort # | Dose Assigned | #Subjects | #DLTs | Recommended Dose |
|----------|---------------|-----------|-------|------------------|
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| Total | | | | |

Click **Enter Interim Data** to open a window in which to enter data for the first cohort: in particular, the **Dose Assigned** and the **DLTs Observed**. Click **OK** to continue.

Enter Interim Data

Editing Cohort #1

Dose Assigned: 5

#Subjects Allocated: 3

#DLTs Observed: 0

OK Cancel

32 Dose Escalation

The dashboard will be updated accordingly, and the next **Recommended Dose** is 10.

| Cohort # | Dose Assigned | #Subjects | #DLTs | Recommended Dose |
|----------|---------------|-----------|-------|------------------|
| 1 | 5 | 3 | 0 | 10 |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |

Click **Enter Interim Data** again, with 10 selected as **Dose Assigned**, enter 2 for **DLTs Observed**, and click **OK**.

Enter Interim Data

Editing Cohort #2

Dose Assigned: 10

#Subjects Allocated: 3

#DLTs Observed: 2

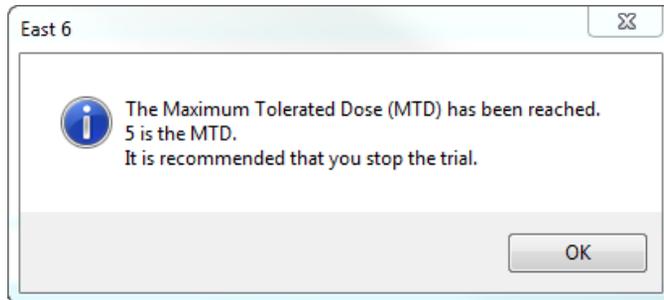
OK Cancel

East now recommends de-escalation to 5.

| Cohort # | Dose Assigned | #Subjects | #DLTs | Recommended Dose |
|----------|---------------|-----------|-------|------------------|
| 1 | 5 | 3 | 0 | 10 |
| 2 | 10 | 3 | 2 | 5 |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |

Click **Enter Interim Data**, with 5 selected as **Dose Assigned**, enter 1 for **DLTs Observed**, and click **OK**.

East recommends that you stop the trial.



Click **Stop Trial** to generate a table for final inference.

| Final Inference | |
|-------------------------|---|
| Final Output at Cohort# | 3 |
| MTD | 5 |

32 Dose Escalation

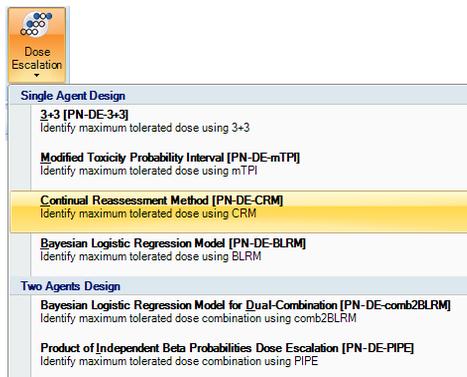
32.2 Continual Reassessment Method (CRM)

32.2.1 Simulation

32.2.2 Interim Monitoring

32.2.1 Simulation

Click **Discrete: Dose Escalation** on the Design tab, and then click **Single Agent Design: Continual Reassessment Method**.



This window is the Input dialog box, which is separated into four tabs: **Design Parameters**, **Stopping Rules**, **Response Generation**, and **Simulation Control**.

In the **Design Parameters** tab, enter **30** as the **Maximum Sample Size**, and **3** for **Cohort Size**. If you were to check the box **Start With**, then you would be simulating from the 3+3 or Accelerated Titration design first, before switching to the CRM. For this tutorial, however, leave the box unchecked.

Enter **0.25** for the **Target Probability of Toxicity**, and **0.3** for the **Target Probability Upper Limit**. This will ensure that the next dose assignment is that whose posterior mean toxicity probability is closest to **0.25**, and below **0.3**.

Target Probability of Toxicity (P_T):

Toxicity Probability Upper Limit (UL):

Assign the dose whose toxicity probability is closest to P_T , and below UL.

Click the **Posterior Sampling...** button. By default, CRM requires the posterior mean only. If instead you wish to sample from the posterior distribution (using a

Metropolis-Hastings algorithm), you will be able to compute and plot the posterior probabilities of being the MTD for each dose. Note that this option will increase the simulation time.

Click the **Dose Skipping...** button. As was recommended in later variations of CRM, in the interests of promoting safety, leave the default options: No untried doses will be skipped while escalating, and no dose escalation will occur when the most recent subject experienced a DLT.

For **Model Type**, select **Power**, with a Gamma($\alpha = 1, \beta = 1$) prior for θ . Other model types available include the **Logistic** and the **Hyperbolic Tangent**. Finally, for the prior probabilities of toxicity of all doses (known as the *skeleton*), enter: 0.05, 0.1,

32 Dose Escalation

0.2, 0.3, 0.35, 0.4, and 0.45.

Model Type: Power

Prior

Distribution: Gamma θ

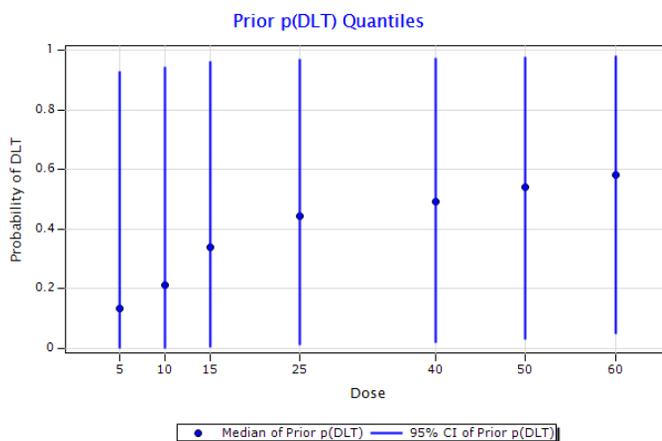
α (shape):

β (rate):



| Dose ID | Probability of Toxicity |
|---------|-------------------------|
| D1 | 0.050 |
| D2 | 0.100 |
| D3 | 0.200 |
| D4 | 0.300 |
| D5 | 0.350 |
| D6 | 0.400 |
| D7 | 0.450 |

Click the  icon to generate a chart of the 95% prior intervals at each dose for probability of DLT.



In the **Stopping Rules** tab, you may specify various rules for stopping the trial. Enter

the following inputs as below.

| | | |
|--|---|----------------|
| Stop Trial Early (MTD not determined) if | | Min SS on Dose |
| <input checked="" type="checkbox"/> | MTD Below Lowest Dose: Prob.($P_l > P_T$ data) > | 0.8 6 |
| <input checked="" type="checkbox"/> | MTD Above Highest Dose: Prob.($P_h < P_T$ data) > | 0.9 6 |
| Stop Trial Early (MTD determined) if | | Min SS on Dose |
| <input type="checkbox"/> | Minimum SS Observed in the Trial >= | |
| <input checked="" type="checkbox"/> | Allocation Rule: SS Already Allocated at Current MTD >= | 9 |

The early stopping rules are divided into two types: Those where the MTD is not determined, and those where the MTD is determined. The former case may arise when the MTD is estimated to be below the lowest dose or above the highest dose. Thus, if the posterior probability of overdosing (toxicity at the lowest dose is greater than target toxicity) exceeds 0.8, then the trial will be stopped. Similarly, if the posterior probability of underdosing (toxicity at the highest dose is lower than target toxicity) exceeds 0.9, then the trial will be stopped. A minimum of 6 subjects will need to be observed on a dose before either of these two rules is activated. A further stopping rule is based on the **Allocation Rule**: If the number of subjects already allocated to the current MTD is at least 9, the trial will be stopped.

In the **Response Generation** tab, you can specify a set of true dose response curves from which to simulate. For the **Starting Dose**, select the lowest dose (5). Leave the default profile as shown below. If you wish to edit or add additional profiles (dose

32 Dose Escalation

response curves), see the corresponding section for the 3+3 design.

Specify True Probability of Toxicity

Curve Family:
 General

Add Profile

True Probability of Toxicity (Profiles)

| Profile\Dose ID | D1 | D2 | D3 | D4 | D5 | D6 | D7 |
|-----------------|------|-----|-----|-----|------|-----|------|
| Dose | 5 | 10 | 15 | 25 | 40 | 50 | 60 |
| GC1 | 0.05 | 0.1 | 0.2 | 0.3 | 0.35 | 0.4 | 0.45 |

Starting Dose: 5

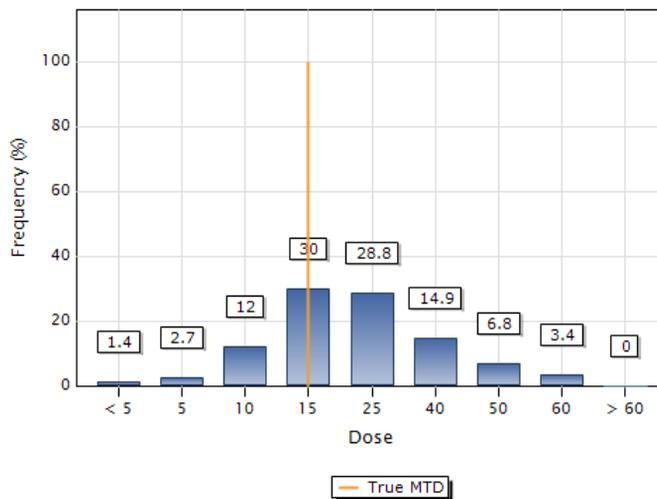
Plot Profiles

In the **Simulation Control** tab, check the boxes corresponding to **Save summary statistics** and **Save subject-level data**. These options will provides access to several charts derived from these more detailed levels of simulated data. If you wish to display subject-level plots for more than one simulation, you can increase the number. For now, leave this at 1 to save computation time.

Click **Simulate** to simulate the CRM design. In the **Output Preview** toolbar, click the  icon to save the simulation to the **Library**. Double-click the simulation node in the **Library** to display the simulation output details. Click the Plots icon in the **Library** to access a wide range of available plots.

Below is an example of the MTD plot, showing the percentage of simulations that each dose level was selected as the MTD. The true MTD is displayed as the third dose level (15). This is the dose whose true probability of DLT (0.2) was closest to and below the

target probability (0.25).



32.2.2 Interim Monitoring

Right-click the Simulation node in the Library, and select **Interim Monitoring**. This will open an empty interim monitoring dashboard.

Click **Enter Interim Data** to open a window in which to enter data for the first cohort: in particular, the **Dose Assigned** and the **DLTs Observed**. Click **OK** to continue.

Continue in this manner by clicking **Enter Interim Data**, entering the following doses, and the corresponding number of DLTs.

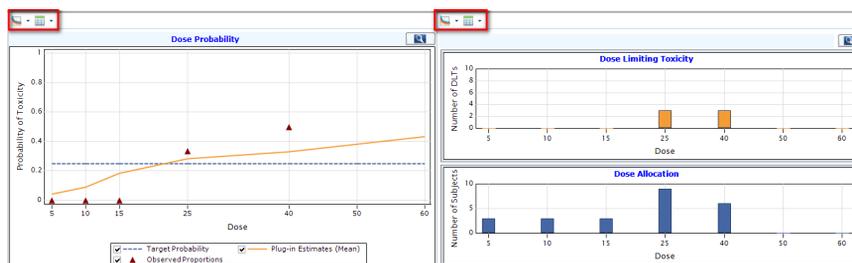
| Cohort # | Dose Assigned | #Subjects | #DLTs | Recommended Dose | Posterior Mean (Θ) |
|----------|---------------|-----------|-------|------------------|--------------------|
| 1 | 5 | 3 | 0 | 10 | 1.493 |
| 2 | 10 | 3 | 0 | 15 | 1.773 |
| 3 | 15 | 3 | 0 | 25 | 2.074 |
| 4 | 25 | 3 | 1 | 40 | 1.509 |
| 5 | 40 | 3 | 1 | 40 | 1.375 |
| 6 | 40 | 3 | 2 | 25 | 1.105 |
| 7 | 25 | 3 | 1 | 25 | 1.072 |
| 8 | 25 | 3 | 1 | 25 | 1.049 |

32 Dose Escalation

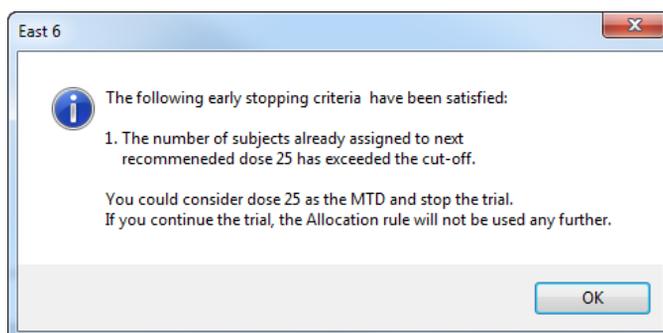
If you click **Display by Dose**, you will see the data grouped by dose level. You may click **Display by Cohort** to return to the original view.

| Dose # | Dose Assigned | Standardized Dose | #Subjects | Prop. Of Subjects | #DLTs | Prop. Of DLTs | Posterior Probability of Toxicity |
|--------|---------------|-------------------|-----------|-------------------|-------|---------------|-----------------------------------|
| 1 | 5 | 0.05 | 3 | 0.125 | 0 | 0 | 0.043 |
| 2 | 10 | 0.1 | 3 | 0.125 | 0 | 0 | 0.089 |
| 3 | 15 | 0.2 | 3 | 0.125 | 0 | 0 | 0.185 |
| 4 | 25 | 0.3 | 9 | 0.375 | 3 | 0.333 | 0.283 |
| 5 | 40 | 0.35 | 6 | 0.25 | 3 | 0.5 | 0.333 |
| Total | | | 24 | | 6 | | |

After each cohort, East will update the Interim Monitoring Dashboard. You may replace the IM dashboard plots with any other plots or corresponding tables, by clicking on the associated icons at the top left of each panel.



At this point, East recommends that you stop the trial.



Click **Stop Trial** to generate a table for final inference.

| Final Inference | |
|-------------------------|---------------------------|
| Final Output at Cohort# | 8 |
| MTD | 25 (using stopping rules) |
| Fitted MTD | 21.662 |

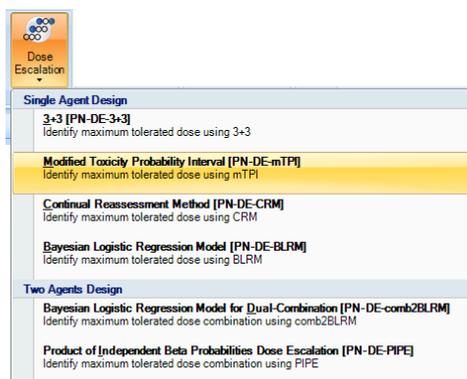
32.3 modified Toxicity Probability Interval (mTPI)

32.3.1 Simulation

32.3.2 Interim Monitoring

32.3.1 Simulation

Click **Discrete: Dose Escalation** on the Design tab, and then click **Single Agent Design: Modified Toxicity Probability Interval**.



This window is the Input dialog box, which is separated into five tabs: **Design Parameters**, **Stopping Rules**, **Trial Monitoring Table**, **Response Generation**, and **Simulation Control**.

In the **Design Parameters** tab, enter 30 as the **Maximum Sample Size**, and 3 for **Cohort Size**. If you were to check the box **Start With**, then you would be simulating from the 3+3 or Accelerated Titration design first, before switching to the mTPI. For this tutorial, however, leave the box unchecked.

Enter 0.25 for the **Target Probability of Toxicity**, 0.2 for the upper limit of the **Under**

32 Dose Escalation

dosing interval, and 0.3 for the upper limit of **Proper dosing** interval.

Target Probability of Toxicity (P_T):

| Toxicity Intervals | Lower Limit | Upper Limit |
|--------------------|-------------|-------------|
| Under dosing | 0.000 | 0.2 |
| Proper dosing | 0.200 | 0.3 |
| Over dosing | 0.300 | 1 |

These entries imply that toxicity probabilities within this interval [0.2 to 0.3] can be regarded as equivalent to the target toxicity (0.25) as far as dosing decisions are concerned. Finally, we will assume a uniform Beta($a = 1, b = 1$) prior distribution for all doses.

Prior

$P_i \sim \text{Beta}(a, b)$ 

P_i : True Toxicity Probability at Dose i

a (Prior Toxicity):

b (Prior Non-Toxicity):

In the **Stopping Rules** tab, enter the following inputs as below.

Design Parameters
Stopping Rules
Trial Monitoring Table
Response Generation
Simulation Controls

Stop Trial Early (MTD not determined) if

| | | |
|---|-----------------------------------|--------------------------------|
| Dose Exclusion Rule: $\text{Prob.}(P_i > P_T \text{data}) >$ | <input type="text" value="0.95"/> | <input type="text" value="3"/> |
| <input checked="" type="checkbox"/> MTD Above Highest Dose: $\text{Prob.}(P_h < P_T \text{data}) >$ | <input type="text" value="0.9"/> | <input type="text" value="3"/> |

Stop Trial Early (MTD determined) if

| | |
|---|----------------------|
| <input type="checkbox"/> Minimum SS Observed in the Trial \geq <input type="text"/> | <input type="text"/> |
| <input type="checkbox"/> Target Rule: $\text{Prob.}(\text{Target Toxicity}) >$ <input type="text"/> | <input type="text"/> |
| <input type="checkbox"/> CI Rule: <input type="text"/> % CI for P(DLT) at MTD $>$ <input type="text"/> and $<$ <input type="text"/> | <input type="text"/> |
| <input type="checkbox"/> Allocation Rule: SS Already Allocated at Current MTD \geq <input type="text"/> | <input type="text"/> |

Stopping Rules Combination

R1 Or R2

| # | Stopping Rule | And/Or |
|----|------------------------|-------------------------------------|
| R1 | Dose Exclusion Rule | Or <input type="button" value="v"/> |
| R2 | MTD Above Highest Dose | <input type="button" value="v"/> |

P_i : Probability of Toxicity at Dose i
 P_h : Probability of Toxicity at Highest Dose
 P_T : Target Probability of Toxicity

For the mTPI design, the stopping rule is based on dose exclusion rules. This states that if there is greater than a 0.95 posterior probability that toxicity for a given dose is greater than the target toxicity, that dose and all higher doses will be excluded in subsequent cohorts. When this dose exclusion rule applies to the lowest dose, then all doses are excluded, and hence the trial will be stopped for excessive toxicity.

Furthermore, the dose exclusion rule is not activated until at least 3 subjects are observed on a dose. A similar idea can be applied to the highest dose: If there is a greater than 95% posterior probability that the toxicity at the highest dose is less than the target toxicity, then stop the trial early.

The remaining stopping rules allow one to stop the trial early with MTD determined. The **Allocation Rule** requires a certain number of subjects already allocated to the next recommended dose. The **CI Rule** requires that the credible interval for probability of DLT at the MTD is within some range. The **Target Rule** requires that the posterior probability of being in the target toxicity, or proper dosing interval, exceeds some threshold. Finally, any of these rules can be combined with **Minimum Ss Observed in the Trial**.

In the **Trial Monitoring Table** tab, you can view the decision table corresponding to the inputs entered in the previous tabs.

Design Parameters | Stopping Rules | Trial Monitoring Table | Response Generation | Simulation Controls

Edit Trial Monitoring Table

Number of patients treated at current dose

| r/n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|-----|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 0 | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E |
| 1 | D | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| 2 | | | DU | D | D | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S |
| 3 | | | | DU | DU | DU | D | S | S | S | S | S | S | S | S | S | S | S | S | S |
| 4 | | | | | DU | DU | DU | DU | D | S | S | S | S | S | S | S | S | S | S | S |
| 5 | | | | | | DU | DU | DU | DU | D | S | S | S | S | S | S | S | S | S | S |
| 6 | | | | | | | DU | DU | DU | DU | D | S | S | S | S | S | S | S | S | S |
| 7 | | | | | | | | DU | DU | DU | DU | DU | D | S | S | S | S | S | S | S |
| 8 | | | | | | | | | DU | DU | DU | DU | DU | D | S | S | S | S | S | S |
| 9 | | | | | | | | | | DU | DU | DU | DU | DU | D | S | S | S | S | S |
| 10 | | | | | | | | | | | DU | DU | DU | DU | DU | D | S | S | S | S |
| 11 | | | | | | | | | | | | DU | DU | DU | DU | DU | D | S | S | S |
| 12 | | | | | | | | | | | | | DU | DU | DU | DU | DU | D | S | S |
| 13 | | | | | | | | | | | | | | DU | DU | DU | DU | DU | D | S |
| 14 | | | | | | | | | | | | | | | DU | DU | DU | DU | DU | D |

Number of Toxicities

- E = Escalate to the next higher dose
- S = Stay at the current dose
- D = De-escalate to the next lower dose
- DU = The current dose is unacceptably toxic

Target Toxicity (%) = 25%
Sample Size = 30

East also provides the option of creating and simulating from a customized trial monitoring table. If you click **Edit Trial Monitoring Table**, you can click on any cell

32 Dose Escalation

in the grid to edit and change the dose assignment rule for that cell.

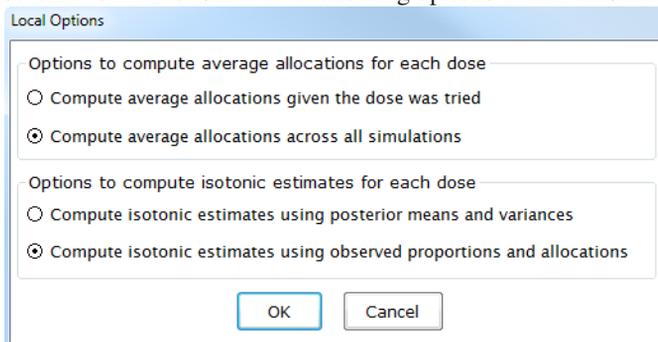
Number of patient

| r\n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-----|----|----|----|----|----|----|----|----|
| 0 | E | E | E | E | E | E | E | E |
| 1 | E | S | S | S | S | S | E | E |
| 2 | D | DU | D | D | S | S | S | S |
| 3 | S | | DU | DU | DU | D | S | S |
| 4 | DU | | | DU | DU | DU | DU | DU |
| 5 | | | | | DU | DU | DU | DU |

In the **Response Generation** tab, you can specify a set of true dose response curves from which to simulate. For the **Starting Dose**, select the lowest dose (5). Leave the default profile as shown below. If you wish to edit or add additional profiles (dose response curves), see the corresponding section for the 3+3 design.

In the **Simulation Control** tab, check the boxes corresponding to **Save summary statistics** and **Save subject-level data**. These options will provide access to several charts derived from these more detailed levels of simulated data. If you wish to display subject-level plots for more than one simulation, you can increase the number. For now, leave this at 1 to save computation time.

Click the **Local Options** button at the top left corner of the input window toolbar. This gives you different options for computing average allocations for each dose, and for computing isotonic estimates. Select the following options and click OK.



Click **Simulate** to simulate the mTPI design. In the **Output Preview** toolbar, click the

 icon to save the simulation to the **Library**. Double-click the simulation node in the **Library** to display the simulation output details. For example, the true MTD was D3 (15), and this dose was selected as MTD the most often (43% of the time).

⊖ **MTD Analysis**

| | True | Median | Mean | Mode |
|------------|------|--------|--------|------|
| MTD | 15 | 15 | 20.756 | 15 |
| Fitted MTD | 20 | 18.992 | 21.538 | NA |

⊖ **% of Simulations selecting each dose as MTD**

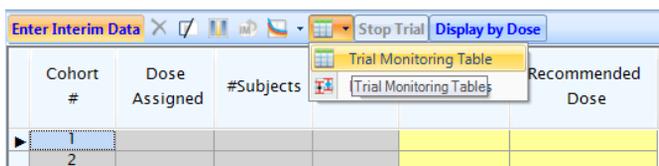
| Dose ID | Below lowest dose | D1 | D2 | D3 | D4 | D5 | D6 | D7 | Above highest dose |
|------------|-------------------|------|-------|-------|-------|------|-----|------|--------------------|
| Dose | | 5 | 10 | 15 | 25 | 40 | 50 | 60 | |
| Percentage | 0.01 | 0.94 | 13.29 | 43.46 | 29.57 | 9.41 | 2.9 | 0.42 | 0 |

Click the Plots icon in the **Library** to access a wide range of available plots.

32.3.2 Interim Monitoring

Right-click one of the Simulation nodes in the Library, and select **Interim Monitoring**. This will open an empty interim monitoring dashboard.

In the interim monitoring toolbar, click the chart icon, and **Trial Monitoring Table** to generate a table to guide dosing decisions for this trial.



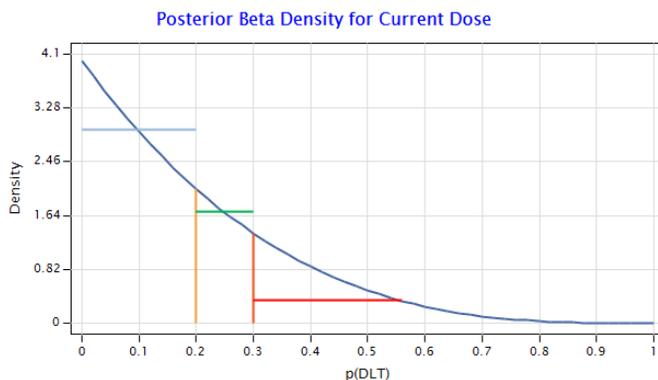
Click **Enter Interim Data** to open a window in which to enter data for the first cohort:

32 Dose Escalation

in particular, the **Dose Assigned** and the **DLTs Observed**. Click **OK** to continue.

The screenshot shows a dialog box titled "Enter Interim Data". Inside, there is a section for "Editing Cohort #1". It contains three input fields: "Dose Assigned" is a dropdown menu showing "5"; "#Subjects Allocated" is a text box containing "3"; and "#DLTs Observed" is a text box containing "0". At the bottom of the dialog are two buttons: "OK" and "Cancel".

The dashboard will be updated accordingly. The decision for the next cohort is based on the highest Unit Probability Mass (UPM): the posterior probability for each toxicity interval divided by the length of the interval. The underdosing interval corresponds to an E (Escalate) decision, the proper dosing interval corresponds to an S (Stay) decision, and the overdosing interval corresponds to a D (De-escalate) decision. In this case, the UMP for underdosing is highest.



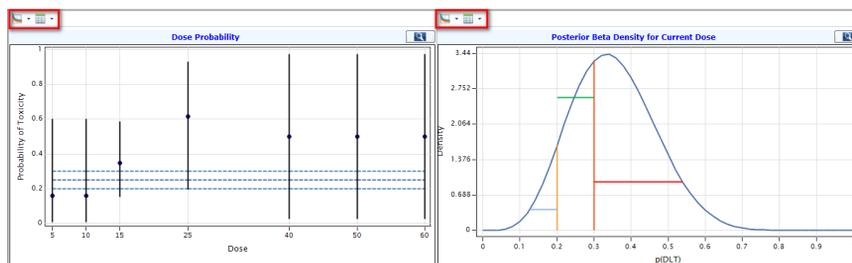
Thus, the recommendation is to escalate to dose 10.

Continue in this manner by entering data for each subsequent cohort, and observe how the interim monitoring dashboard updates.

One example is given below.

| Cohort # | Dose Assigned | #Subjects | #DLTs | Decision | Recommended Dose |
|----------|---------------|-----------|-------|----------|------------------|
| 1 | 5 | 3 | 0 | E | 10 |
| 2 | 10 | 3 | 0 | E | 15 |
| 3 | 15 | 3 | 0 | E | 25 |
| 4 | 25 | 3 | 2 | D | 15 |
| 5 | 15 | 3 | 2 | S | 15 |
| 6 | 15 | 3 | 1 | S | 15 |
| 7 | 15 | 3 | 2 | S | 15 |
| 8 | 15 | 3 | 0 | S | 15 |

After each cohort, East will update the Interim Monitoring Dashboard. You may replace the IM dashboard plots with any other plots or corresponding tables, by clicking on the associated icons at the top left of each panel.



Suppose we wished to end the study after 8 cohorts (24 patients). Click **Stop Trial** to end the study and generate a table of final inference.

| Final Inference | |
|-------------------------|-------|
| Final Output at Cohort# | 8 |
| MTD | 15 |
| Fitted MTD | 13.75 |

32.4 Bayesian logistic regression model (BLRM)

32.4.1 Simulation

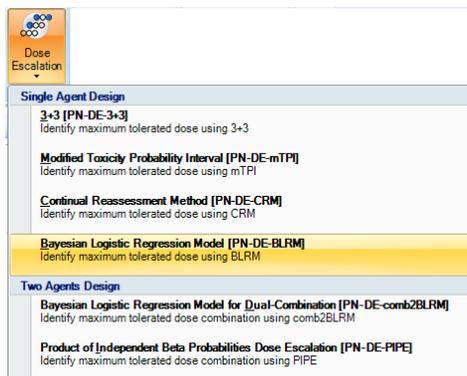
32.4.2 Interim Monitoring

32.4.1 Simulation

Click **Discrete: Dose Escalation** on the Design tab, and then click **Single Agent**

32 Dose Escalation

Design: Bayesian Logistic Regression Model.



This window is the Input dialog box, which is separated into four tabs: **Design Parameters**, **Stopping Rules**, **Response Generation**, and **Simulation Control**.

In the **Design Parameters** tab, enter 30 as the **Maximum Sample Size**, and 3 for **Cohort Size**. If you were to check the box **Start With**, then you would be simulating from the 3+3 or Accelerated Titration design first, before switching to the BLRM. For this tutorial, however, leave the box unchecked.

The next step is to choose a **Dose Selection Method**: either by **Bayes Risk** or by **Max Target Toxicity**. For the next cohort, the Bayes risk method selects the dose that minimizes the posterior expected loss, aka Bayes risk. In contrast, Max Target Toxicity method selects the dose that maximizes the posterior probability of targeted toxicity. For both methods, the dose selected must not exceed the EWOC (Escalation With Overdose Control) threshold: the posterior probability of overdosing,

either excessive or unacceptable toxicity, is less than the threshold (e.g., 0.25).

Target Probability of Toxicity (P_T):

Dose Selection Method
 Max Targeted Toxicity Bayes Risk

| Toxicity Intervals | Lower Limit | Upper Limit |
|-----------------------|-------------|-------------|
| Under dosing | 0.000 | 0.200 |
| Targeted toxicity | 0.200 | 0.350 |
| Excessive toxicity | 0.350 | 1.000 |
| Unacceptable toxicity | | |

EWOC: Prob. (Overdosing) <

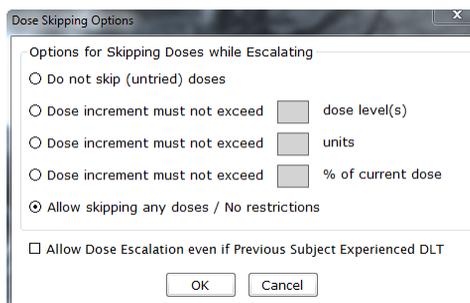
Reference Dose (D^*):

Recall that the BLRM method applies the following model:

$$\text{logit}(\pi_d) = \log(\alpha) + \beta \log(d/d^*) \tag{32.1}$$

The **Reference Dose (D^*)** is the dose at which the odds of observing a DLT is α .

Click the **Dose Skipping** button, and select **Allow skipping any doses / No Restrictions**.



You can specify the prior directly in terms of a bivariate normal distribution for $\log(\alpha)$

32 Dose Escalation

and $\log(\beta)$.

Prior Specification

Prior Calculator...

| | | | |
|--------------|-------------------------------------|------------------------------------|---|
| | ln (α) | ln (β) | |
| Mean: | <input type="text" value="-0.847"/> | <input type="text" value="0.265"/> |  |
| SD: | <input type="text" value="1.275"/> | <input type="text" value="1.981"/> | |
| Correlation: | <input type="text" value="0"/> | | |

Alternatively, if you click **Prior Calculator**, a calculator will appear allowing you to specify a prior indirectly by one of three methods: (1) lowest dose and reference dose, (2) lowest dose and highest dose, or (3) lowest dose and MTD. Click **Recalc** to convert the prior inputs into matching bivariate normal parameter values, and click **OK** to paste these values into the input window. Appendix N of the manual, and Appendix A of Neuenschwander et al. (2008) describes some of these methods.

Prior Distribution Calculator x

Specify Prior on: Lowest Dose and Reference Dose

Input

Prob. (DLT) at D1:

Prob. (DLT) at D*:

$P[\pi(D1) \leq \text{ }] = \text{ }$

$P[\pi(D*) \leq \text{ }] = \text{ }$

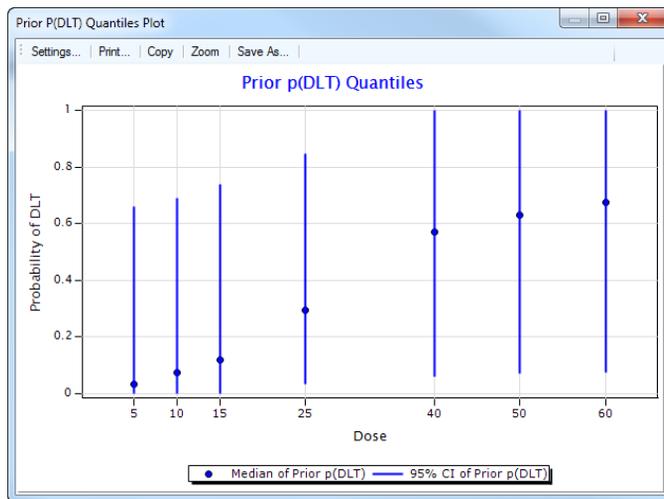
D1: Lowest Dose D*: Reference Dose

Output

| | | | |
|--------------|-------------------------------------|------------------------------------|--|
| | ln (α) | ln (β) | |
| Mean: | <input type="text" value="-0.847"/> | <input type="text" value="0.265"/> |  |
| SD: | <input type="text" value="1.275"/> | <input type="text" value="1.981"/> | |
| Correlation: | <input type="text" value="0"/> | | |

Click the  icon to generate a chart of the 95% prior intervals at each dose for

probability of DLT.



Click **Posterior Sampling Methods** to select from one of two methods: Metropolis Hastings, or direct Monte Carlo. For this tutorial, click **OK** to select **Direct**.

The dialog box is titled "Posterior Sampling Methods". It has two radio buttons: "Direct" (selected) and "Metropolis Hastings". Below the radio buttons are two input fields: "In (α)" with value "30" and "In (β)" with value "30". There are also two input fields: "# Points:" with value "30" and "Steady state simulations:" with value "1000". At the bottom are "Ok" and "Cancel" buttons.

In the **Stopping Rules** tab, you can specify multiple rules for stopping the trial. The trial is stopped early and MTD not determined if there is evidence of underdosing. This rule is identical to that from mTPI: If there is a greater than some threshold posterior probability that the toxicity at the highest dose is less than the target toxicity, then stop the trial early.

The remaining stopping rules allow one to stop the trial early with MTD determined. The **Allocation Rule** requires a certain number of subjects already allocated to the

32 Dose Escalation

next recommended dose. The **CI Rule** requires that the credible interval for probability of DLT at the MTD is within some range. The **Target Rule** requires that the posterior probability of being in the target toxicity exceeds some threshold. Finally, any of these rules can be combined with **Minimum Ss Observed in the Trial**. Check the appropriate boxes and enter values as below.

Stop Trial Early (MTD not determined) if

MTD Above Highest Dose: Prob.($P_n < P_T$ | data) > Min SS on Dose

Stop Trial Early (MTD determined) if

Minimum SS Observed in the Trial >= Min SS on Dose

Target Rule: Prob.(Target Toxicity) > Min SS on Dose

CI Rule: % CI for P(DLT) at MTD > and < Min SS on Dose

Allocation Rule: SS Already Allocated at Current MTD >=

Stopping Rules Combination
(R1 Or R2)

| # | Stopping Rule | And/Or |
|----|-----------------|--------|
| R1 | Target Rule | Or |
| R2 | Allocation Rule | |

In the **Response Generation** tab, you can specify a set of true dose response curves from which to simulate. For the **Starting Dose**, select the lowest dose (5). Leave the default profile as shown below. If you wish to edit or add additional profiles (dose response curves), see the corresponding section for the 3+3 design.

In the **Simulation Control** tab, check the boxes corresponding to **Save summary statistics**, **Save subject-level data**, and **Save final posterior samples**. These options will provides access to several charts derived from these more detailed levels of simulated data. If you wish to display subject-level plots, or posterior distribution plots, for more than one simulation, you can increase the number. For now, leave both of these at 1 to save computation time.

Design Parameters
Stopping Rules
Response Generation
Simulation Control

Number of Simulations:

Refresh Frequency:

Random Number Seed

Clock

Fixed

Suppress All Intermediate Output

Pause after Refresh

Stop at End

Output Options

Output Type:

Save summary statistics for every simulation run

Save subject level data for simulation runs

Save final posterior samples for simulation runs

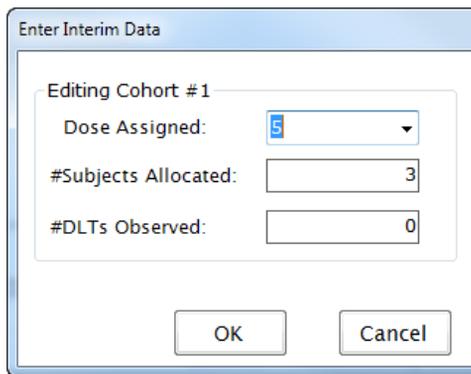
Note: Max. 100,000 records will be saved.

Click **Simulate** to simulate the BLRM design. In the **Output Preview** toolbar, click

the  icon to save the simulation to the **Library**. Double-click the simulation node in the **Library** to display the simulation output details. Click the Plots icon in the **Library** to access a wide range of available plots.

32.4.2 Interim Monitoring

Right-click the Simulation node in the Library, and select **Interim Monitoring**. This will open an empty interim monitoring dashboard. Click **Enter Interim Data** to open a window in which to enter data for the first cohort: in particular, the **Dose Assigned** and the **DLTs Observed**. Click **OK** to continue.



The dialog box titled "Enter Interim Data" contains the following fields and controls:

- Editing Cohort #1
- Dose Assigned:
- #Subjects Allocated:
- #DLTs Observed:
- OK button
- Cancel button

The dashboard will be updated accordingly.

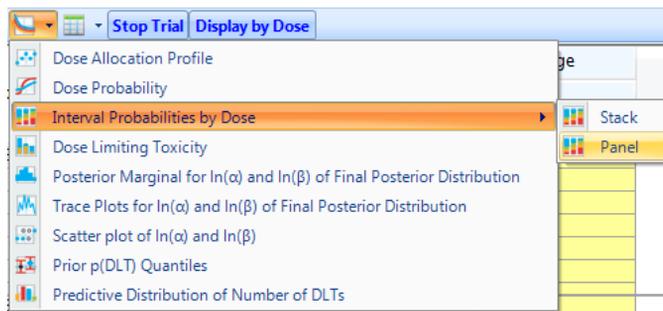
The acceptable dose range is on a continuous scale between the minimum and maximum doses. The upper limit of the acceptable dose range is the largest dose whose probability of overdosing is less than the EWOC threshold. The lower limit of the acceptable range is the dose whose DLT rate is equal to the lower limit of the targeted toxicity interval. When the computed lower limit exceeds the recommended

32 Dose Escalation

dose, it is set to the recommended dose.

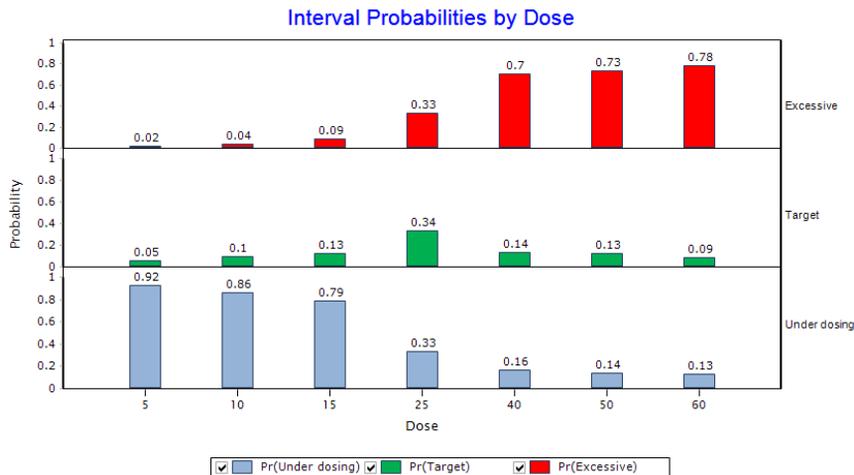
| Cohort # | Dose Assigned | #Subjects | #DLTs | Recommended Dose | Acceptable Dose Range | |
|----------|---------------|-----------|-------|------------------|-----------------------|-------------|
| | | | | | Lower Limit | Upper Limit |
| 1 | 5 | 3 | 0 | 15 | 15 | 23.221 |
| 2 | | | | | | |
| 3 | | | | | | |
| 4 | | | | | | |
| 5 | | | | | | |
| 6 | | | | | | |
| 7 | | | | | | |
| Total | | 3 | 0 | | | |

In the IM toolbar, click the Plots icon, then **Interval Probabilities by Dose** and **Panel**.

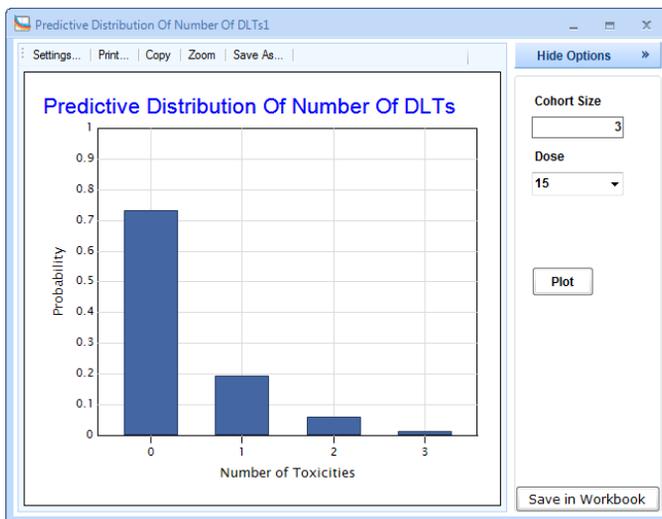


Notice that for all doses greater than or equal to 25, the posterior probability of overdosing exceeds the EWOC threshold (0.25). Of the remaining doses, dose 15 maximizes the probability of targeted toxicity, and is therefore the next recommended

dose.

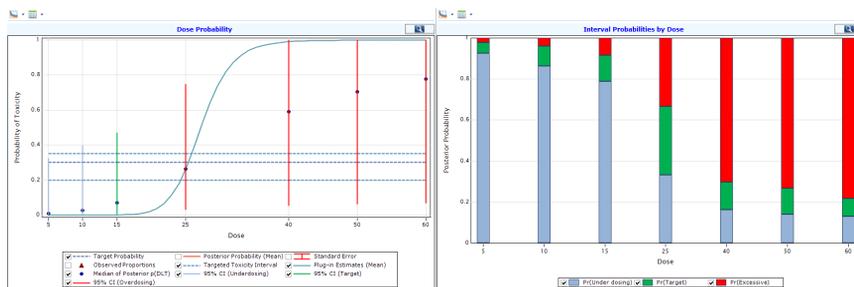


In the IM toolbar, click the Plots icon, then **Predictive Distribution of Number of DLTs**. You can enter a planned cohort size and select a next dose, to plot the posterior predictive probability of the number of DLTs to be observed in next cohort.



32 Dose Escalation

After each cohort, East will update the Interim Monitoring Dashboard. You may replace the IM dashboard plots with any other plots or corresponding tables, by clicking on the associated icons at the top left of each panel.



Continue entering data for each subsequent cohort, and observe how the interim monitoring dashboard updates. One example is given below.

| | Cohort # | Dose Assigned | #Subjects | #DLTs | Recommended Dose |
|---|----------|---------------|-----------|-------|------------------|
| | 1 | 5 | 3 | 0 | 15 |
| | 2 | 15 | 3 | 0 | 25 |
| | 3 | 25 | 3 | 0 | 25 |
| | 4 | 25 | 3 | 0 | 25 |
| | 5 | 25 | 3 | 1 | 25 |
| ▶ | 6 | | | | |
| | 7 | | | | |
| | Total | | 15 | 1 | |

Click **Stop Trial** to generate final inference table.

32.5 Bayesian logistic regression model for dual-combination (comb2BLRM)

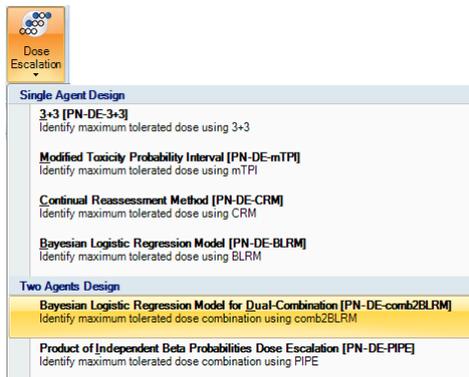
32.5.1 Simulation

32.5.2 Interim Monitoring

32.5.1 Simulation

Click **Discrete: Dose Escalation** on the Design tab, and then click **Two Agents**

Design: Bayesian Logistic Regression Model for Dual-Combination.



Set the **Max. Number of Doses** as 4 for both Agent 1 and Agent 2, the **Max. Sample Size** as 60, the **Cohort Size** as 3.

Set the target toxicity interval to 16-35%, with an EWOC criterion of 0.25. Set the reference doses to 290 and 20 for Agents 1 and 2, respectively.

Design Parameters Stopping Rules Response Generation Simulation Controls

Max. Sample Size:

Cohort Size:

| Toxicity Intervals | Lower Limit | Upper Limit |
|--------------------|-------------|-------------|
| Under dosing | 0.000 | 0.160 |
| Targeted toxicity | 0.160 | 0.350 |
| Overdosing | 0.350 | 1.000 |

EWOC: Prob. (Overdosing) <

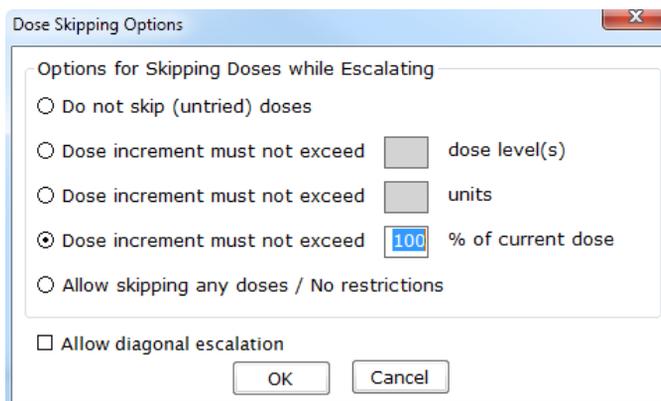
Reference Dose for Agent 1:

Reference Dose for Agent 2:

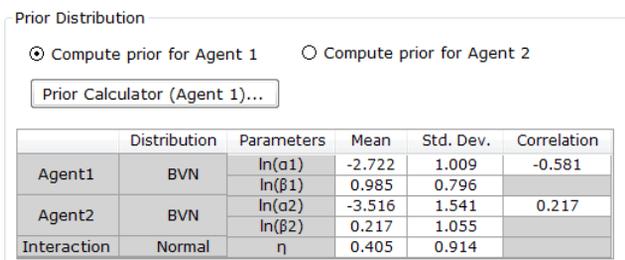
Click the button for **Dose Skipping**. These options imply that the dose of only one compound can be increased for the next cohort (no diagonal escalation), with a

32 Dose Escalation

maximum increment of 100



The prior distribution is an extension of that for the single-agent BLRM, but includes a normal prior for the interaction term. As with the single-agent BLRM, you can use the calculator to transform prior information on particular dose levels to a bivariate normal for either Agent 1 or Agent 2. In this tutorial, we will simply enter the following values adapted from Neuenschwander et al. (2015).



In the **Stopping Rules** tab, you may specify various rules for stopping the trial. The logical operators (And/Or) follow left-to-right precedence, beginning with the top-most rule in the table. The order of the stopping rules is determined by the order of selection.

Enter the following inputs as below. Select the **Minimum Ss** rule first, followed by the **Target Rule**, followed by the **Allocation Rule**. Be sure to select the appropriate logical operators. This combination of rules implies the MTD dose combination declared will meet the following conditions: (1) At least 6 patients have already been allocated to this combination, and (2) This dose satisfies one of the following: (i) The

probability of targeted toxicity at this combination exceeds 0.5, or (ii) A minimum of 15 subjects have already been observed in the trial.

Stop Trial Early (MTD determined) if

Minimum SS Observed in the Trial >=

Target Rule: Prob.(Target Toxicity) >

Allocation Rule: SS Already Allocated at Current MTD >=

Stopping Rules Combination
 ((R1 Or R2) And R3)

| # | Stopping Rule | And/Or |
|----|---------------------------|--------|
| R1 | Minimum Subjects in Trial | Or |
| R2 | Target Rule | And |
| R3 | Allocation Rule | |

In the **Response Generation** tab, enter the following inputs. Make sure that the starting dose combination is the lowest dose level for each agent.

Design Parameters Stopping Rules **Response Generation** Simulation Controls

Specify True Probability of Toxicity (Agent1)

Curve Family: General

| Profile\Dose ID | D1 | D2 | D3 | D4 |
|-----------------|------|-----|-----|-----|
| Dose | 100 | 200 | 300 | 400 |
| GC1 | 0.05 | 0.1 | 0.2 | 0.3 |

Starting Dose: 100

Interaction (η): Recalc>>

Specify True Probability of Toxicity (Agent2)

Curve Family: General

| Profile\Dose ID | D1 | D2 | D3 | D4 |
|-----------------|------|-----|-----|-----|
| Dose | 10 | 20 | 30 | 40 |
| GC1 | 0.05 | 0.1 | 0.2 | 0.3 |

Starting Dose: 10

True Probability of Toxicity (Profiles)

Label: Profile1 Add Delete << 0/0 >>

Agent2

| Doses | 10 | 20 | 30 | 40 |
|-------|-------|-------|------|-------|
| 100 | 0.098 | 0.145 | 0.24 | 0.335 |
| 200 | 0.145 | 0.19 | 0.28 | 0.37 |
| 300 | 0.24 | 0.28 | 0.36 | 0.44 |
| 400 | 0.335 | 0.37 | 0.44 | 0.51 |

Plot

In the **Simulation Control** tab, select the following options. In this tutorial, we will run only 1000 simulations. Click **Simulate**.

Design Parameters Stopping Rules Response Generation **Simulation Controls**

Number of Simulations:

Refresh Frequency:

Random Number Seed

Clock

Fixed

Suppress All Intermediate Output

Pause after Refresh

Stop at End

Output Options

Output Type: Case Data

Save summary statistics for every simulation run

Save subject level data for simulation runs

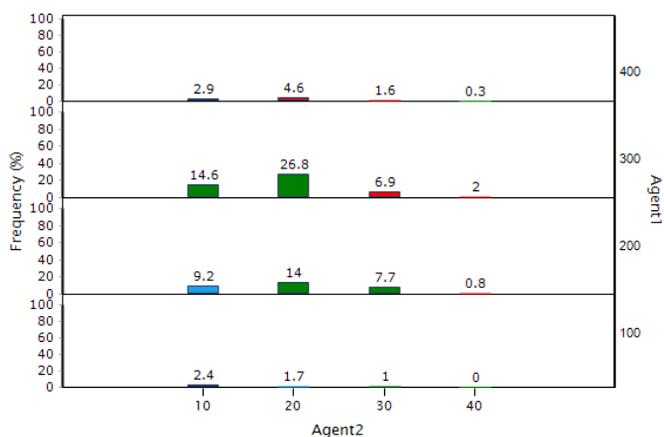
Save final posterior samples for simulation runs

Note: Max. 100,000 records will be saved.

32 Dose Escalation

In the **Output Preview** toolbar, click the  icon to Sim1 to the **Library**. Double-click Sim1 in the **Library** to display the simulation output details.

With Sim1 selected in the **Library**, click the Plots icon to access a wide range of available plots. Below is an example of the MTD plot, showing the percentage of simulations that each dose combination was selected as the MTD. The combinations whose true DLT rates were below, within, and above the target toxicity interval (0.16 – 0.35) are colored blue, green, and red, respectively.



32.5.2 Interim Monitoring

Right-click the Simulation node in the Library, and select **Interim Monitoring**. This will open an empty interim monitoring dashboard. Click **Enter Interim Data** to open a window in which to enter data for the first cohort: in particular, the **Dose Assigned** for

each agent, the **Subjects Allocated** and the **DLTs Observed**. Click **OK** to continue.

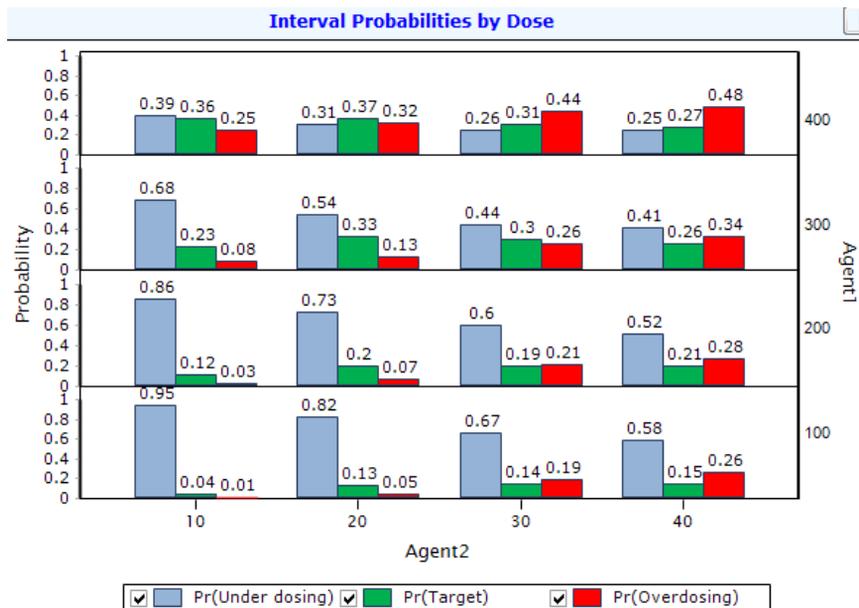
The next recommended dose is 100 mg for Agent 1 and 20 mg for Agent 2.

| Cohort # | Dose Assigned | | #Subjects | #DLTs | Recommended Dose | |
|----------|---------------|--------|-----------|-------|------------------|--------|
| | Agent1 | Agent2 | | | Agent1 | Agent2 |
| 1 | 100 | 10 | 3 | 0 | 100 | 20 |
| 2 | | | | | | |
| 3 | | | | | | |
| 4 | | | | | | |
| 5 | | | | | | |
| 6 | | | | | | |
| 7 | | | | | | |
| Total | | | 3 | 0 | | |

Recall that the dose skipping constraints are that the dose increment cannot exceed 100% of the current dose, and that only one compound can be increased. Of the eligible dose combinations, the recommended one has the highest probability of

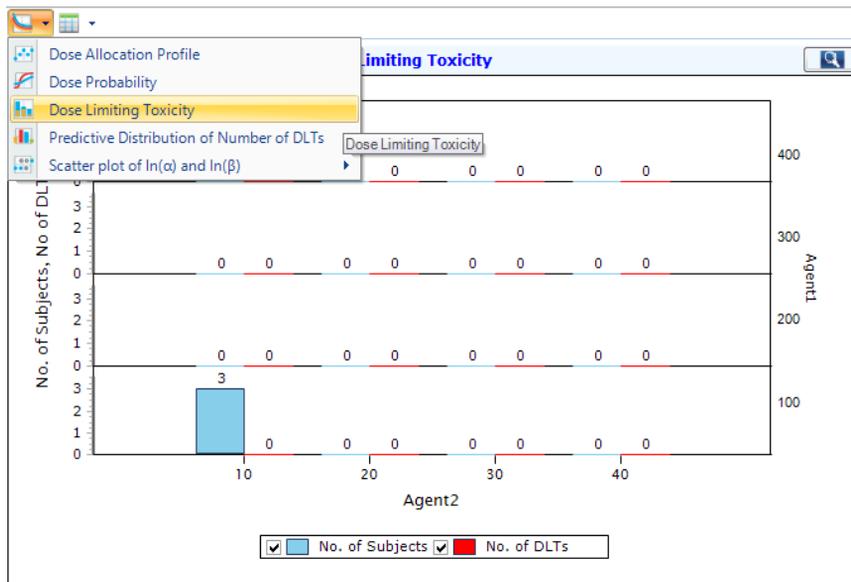
32 Dose Escalation

targeted toxicity.



You may replace the IM dashboard plots with any other plots or corresponding tables, by clicking on the associated icons at the top left of each panel. For example, change the left-hand plot to **Dose Limiting Toxicity** to view the number of subjects and DLTs

observed at each dose combination.



Continue in this manner by clicking **Enter Interim Data**, entering the following doses, and the corresponding number of DLTs.

| Cohort # | Dose Assigned | | #Subjects | #DLTs | Recommended Dose | |
|----------|---------------|--------|-----------|-------|------------------|----------|
| | Agent1 | Agent2 | | | Agent1 | Agent2 |
| 1 | 100 | 10 | 3 | 0 | 100 | 20 |
| 2 | 100 | 20 | 3 | 0 | 100 | 40 |
| 3 | 100 | 40 | 3 | 1 | 200 | 40 |
| 4 | 200 | 40 | 3 | 2 | 200 | 30 |
| 5 | 200 | 30 | 3 | 1 | 200 | 30 |
| 6 | 200 | 30 | 3 | 1 | 100 | 30 |
| 7 | 100 | 30 | 3 | 0 | MTD = 200 | MTD = 30 |
| 8 | | | | | | |
| 9 | | | | | | |
| 10 | | | | | | |
| 11 | | | | | | |
| Total | | | 21 | 5 | | |

The recommended MTD combination is 200 mg for Agent 1 and 30 mg for Agent 2.

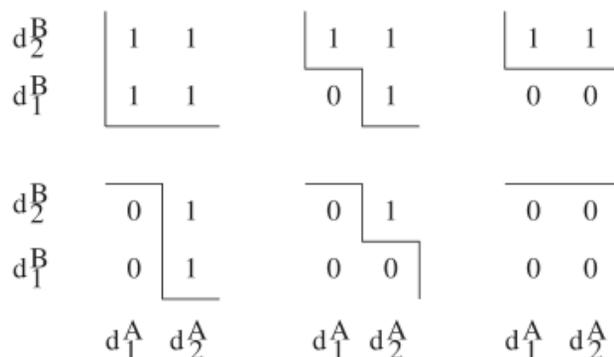
32 Dose Escalation

32.6 Product of Independent beta Probabilities dose Escalation (PIPE)

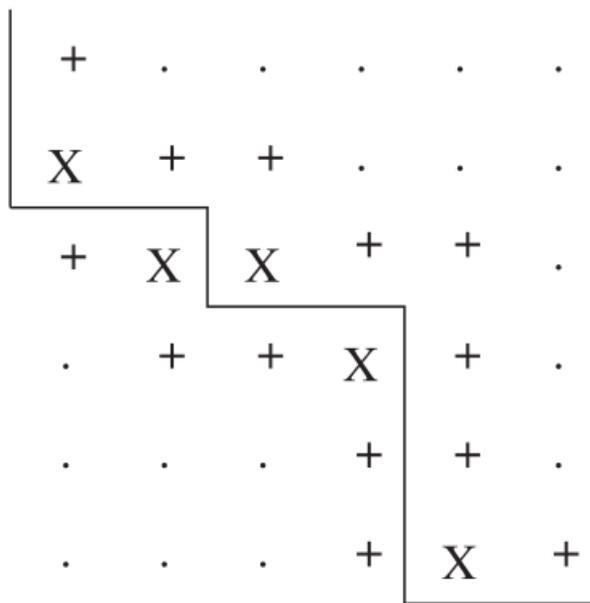
32.6.1 Simulation

One of the core concepts underlying the PIPE method is the maximum tolerated contour (MTC), a line partitioning the dose combination space into toxicity probabilities either less than or greater than the target. The recommended dose combination at the end of the trial is the dose combination closest from below to the MTC. The following figures from Mander and Sweeting (2015) illustrate the MTC, and the related concepts of admissible dose combinations (adjacent or closest) and dose skipping options (neighborhood vs non-neighborhood constraint).

This figure below shows six monotonic MTCs for two agents, each with two dose levels.



After each cohort, the most likely contour is selected before applying a dose selection strategy. The next dose combination is chosen from a set of admissible doses, which are either *closest* to the most likely contour, or *adjacent*. In the figure below, all the (X) and (+) symbols are considered *adjacent*. Of these, the (X) symbols represent the *closest* doses.

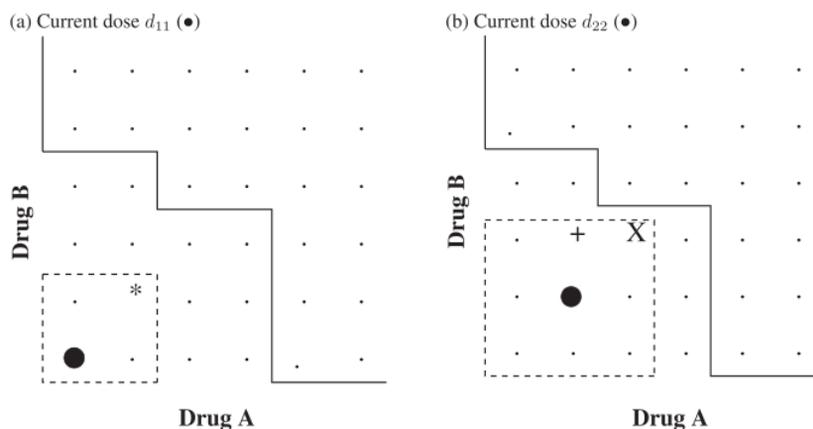


Of the admissible doses, the next dose combination chosen is that with the **minimum sample size**, where sample size is defined as the prior and trial sample size combined. The **weighted randomization method** selects one of the admissible doses at random, with selection probabilities weighted by the inverse of their sample size.

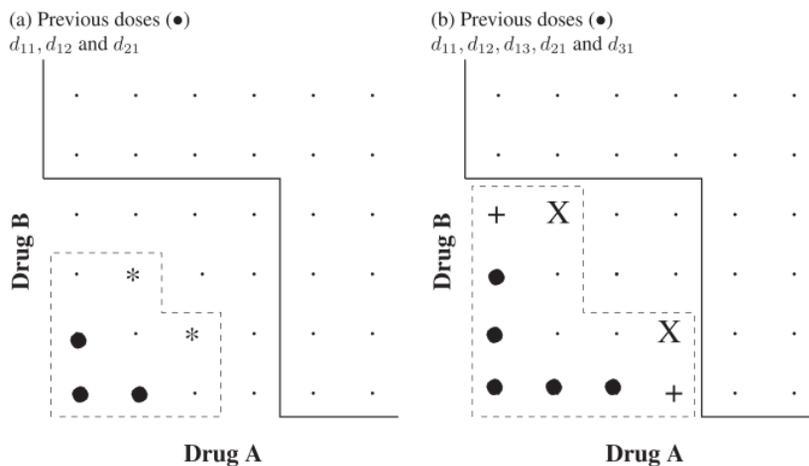
For dose skipping options, one can choose between a neighborhood constraint, or a non-neighborhood constraint. The **neighborhood constraint** restricts the set of admissible doses to those a single dose level higher or lower than the current dose for both agents, while the **non-neighborhood constraint** restricts the set of admissible doses to a single dose level higher or lower than any previously administered dose combination.

This figure below illustrates the neighborhood constraint, at two different cohorts. Only those combinations within the dashed box are admissible. The asterisk symbol on the left represents the admissible dose combination closest to the MTC.

32 Dose Escalation

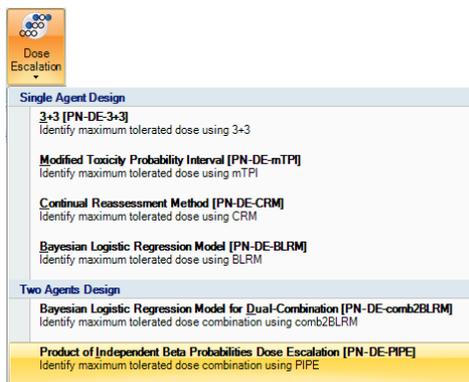


This figure below illustrates the non-neighborhood constraint. The set of admissible doses is now larger because all previously administered doses are included.

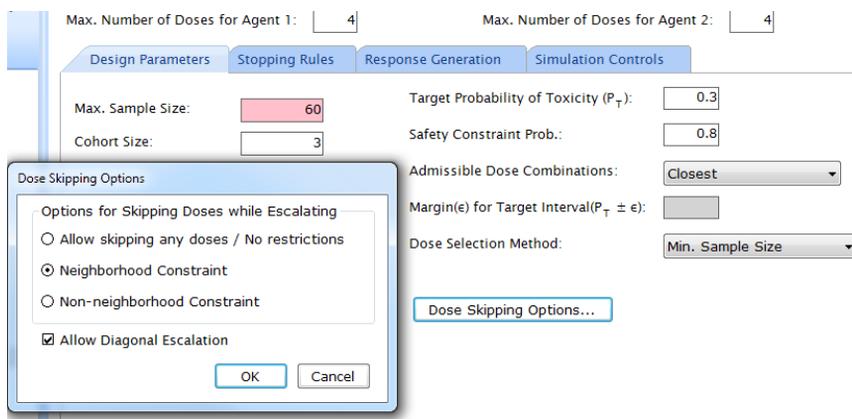


Finally, there is a safety constraint threshold to avoid overdosing. Averaging over the posterior distribution of all monotonic contours, the expected probability of being above the MTC is calculated for all dose combinations. Those dose combinations whose expected probabilities exceed the safety threshold are excluded from the admissible set.

Click **Discrete: Dose Escalation** on the Design tab, and then click **Two Agents Design: Product of Independent Beta Probabilities Dose Escalation**.



In the **Design Parameters** tab, select the following options.



In addition to the **Closest** and **Adjacent** options for **Admissible Dose Combinations**, there is also an **Interval** option. This allows you to specify a margin ϵ around the target toxicity level to define the admissible dose set, rather than relying on the MTC. Dose combinations whose posterior mean toxicity risk lies in the specified target interval ($P_T \pm \epsilon$) are considered admissible.

For the prior specification, enter the following values. When entering the same prior

32 Dose Escalation

sample size for each dose combination, a value of 1 considered a *strong* prior, whereas a value of 1 divided by the number of combinations can be considered a *weak* prior (Mander & Sweeting, 2015).

Specify Prior Using: Prior P(DLT) and Prior SS

Prior Specification

Prior P(DLT):

| | | Agent2 | | | |
|--------|----|--------|-------|-------|-------|
| Doses | | D1 | D2 | D3 | D4 |
| Agent1 | D1 | 0.098 | 0.145 | 0.240 | 0.335 |
| | D2 | 0.145 | 0.190 | 0.280 | 0.370 |
| | D3 | 0.240 | 0.280 | 0.360 | 0.440 |
| | D4 | 0.335 | 0.370 | 0.440 | 0.510 |

Prior Sample Size:

| | | Agent2 | | | |
|--------|----|--------|-------|-------|-------|
| Doses | | D1 | D2 | D3 | D4 |
| Agent1 | D1 | 1.000 | 1.000 | 1.000 | 1.000 |
| | D2 | 1.000 | 1.000 | 1.000 | 1.000 |
| | D3 | 1.000 | 1.000 | 1.000 | 1.000 |
| | D4 | 1.000 | 1.000 | 1.000 | 1.000 |

In the **Stopping Rules** tab, there are a number of options similar to those from other designs. However, for this tutorial, leave these options unchecked.

Design Parameters | **Stopping Rules** | Response Generation | Simulation Controls

Stop Trial Early (MTD not determined) if

MTD Below Lowest Dose: Prob.($P_1 > P_T$ | data) > Min SS on Dose

Stopping Rule | And/Or

| Stopping Rule | And/Or |
|---------------|--------|
| | ▼ |
| | ▼ |

Stop Trial Early (MTD determined) if

Minimum SS Observed in the Trial >=

Allocation Rule: SS Already Allocated to Next Recommended Dose >=

Similarly, leave the default options in the **Response Generation** tab. In this tutorial, the true probabilities of toxicity will be in agreement with the prior medians specified

above.

The screenshot shows the 'Simulation Controls' tab with two sections for 'Specify True Probability of Toxicity'. The first section is for 'Agent1' and the second for 'Agent2'. Both sections have a 'Curve Family' dropdown set to 'General' and a 'Starting Dose' dropdown. The 'Agent1' starting dose is 100, and the 'Agent2' starting dose is 10. A 'Recalc>>' button is located between the two sections. To the right, a 'True Probability of Toxicity (Profiles)' window shows a table for 'Agent2' with columns for Doses (10, 20, 30, 40) and rows for Dose (100, 200, 300, 400). The table contains numerical values representing toxicity probabilities.

| | | Agent2 | | | | |
|--------|------|--------|-------|-------|------|-------|
| | | Doses | 10 | 20 | 30 | 40 |
| Agent1 | Dose | 100 | 0.098 | 0.145 | 0.24 | 0.335 |
| | GC1 | 200 | 0.145 | 0.19 | 0.28 | 0.37 |
| | 300 | 0.24 | 0.28 | 0.36 | 0.44 | |
| | 400 | 0.335 | 0.37 | 0.44 | 0.51 | |

In the **Simulation Controls** tab, you can run 1000 simulations, although the PIPE method runs relatively quickly.

The screenshot shows the 'Simulation Controls' tab with 'Number of Simulations' set to 1000 and 'Refresh Frequency' set to 100. Under 'Random Number Seed', the 'Clock' radio button is selected. There are checkboxes for 'Suppress All Intermediate Output', 'Pause after Refresh', and 'Stop at End'. On the right, the 'Output Options' section shows 'Output Type' set to 'Case Data' and checkboxes for 'Save summary statistics for every simulation run' and 'Save subject level data for 1 simulation runs'. A note states: 'Note: Max. 100,000 records will be saved.'

In the **Output Preview** toolbar, click the  icon to save the simulation to the **Library**. Double-click the simulation node in the **Library** to display the simulation output details.

In the **MTD Analysis** table, you can see that the (Agent 1, Agent 2) dose combinations selected most often as MTD were: (300, 10) at 22.1% and (300, 20) at 20.8%. The true

32 Dose Escalation

probabilities of toxicity at these combinations were 0.24 and 0.28, respectively.

⊖ MTD Analysis

| Dose_1 | Dose_2 | True Prob. of Toxicity | % of Sims selected as MTD | Average Allocations | Average No. of DLTs |
|--------|--------|------------------------|---------------------------|---------------------|---------------------|
| 100 | 10 | 0.098 | 0.1 | 9 | 0 |
| 100 | 20 | 0.145 | 0.4 | 10.5 | 1.75 |
| 100 | 30 | 0.24 | 0.7 | 12.429 | 1.857 |
| 100 | 40 | 0.335 | 0.4 | 15.75 | 3.75 |
| 200 | 10 | 0.145 | 5 | 6.78 | 0.8 |
| 200 | 20 | 0.19 | 9.7 | 10.763 | 1.546 |
| 200 | 30 | 0.28 | 8.1 | 14.889 | 2.667 |
| 200 | 40 | 0.37 | 1.8 | 17.5 | 3.889 |
| 300 | 10 | 0.24 | 22.1 | 9.339 | 1.412 |
| 300 | 20 | 0.28 | 20.8 | 12.649 | 2.188 |
| 300 | 30 | 0.36 | 5.6 | 13.018 | 2.5 |
| 300 | 40 | 0.44 | 0.3 | 14 | 3.667 |
| 400 | 10 | 0.335 | 16.1 | 9.783 | 1.665 |
| 400 | 20 | 0.37 | 8 | 9.938 | 1.688 |
| 400 | 30 | 0.44 | 0.7 | 9.429 | 1.143 |
| 400 | 40 | 0.51 | 0 | NA | NA |

32.6.2 Interim Monitoring

Right-click the Simulation node in the Library, and select **Interim Monitoring**. This will open an empty interim monitoring dashboard. Click **Enter Interim Data** to open a window in which to enter data for the first cohort: in particular, the **Dose Assigned** for

each agent, the **Subjects Allocated** and the **DLTs Observed**. Click **OK** to continue.

The next recommended dose is 200 mg for Agent 1 and 20 mg for Agent 2.

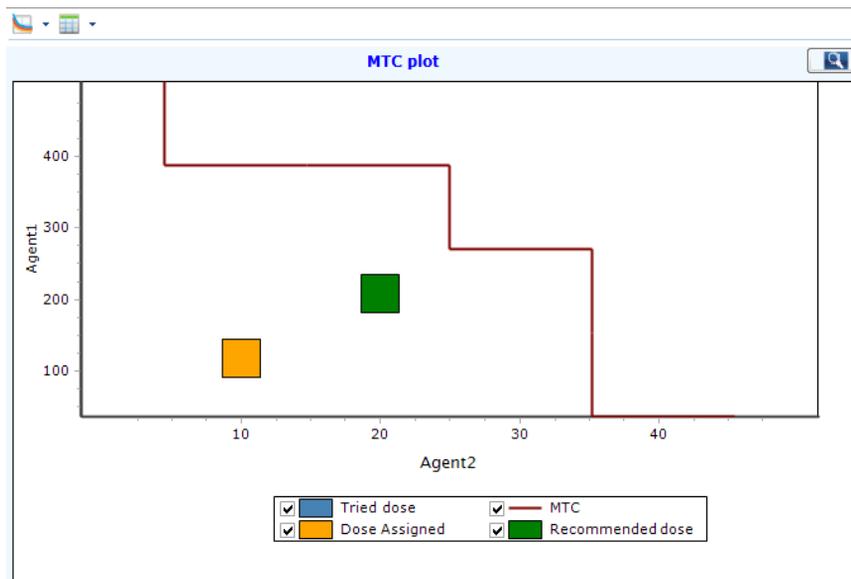
| Cohort # | Dose Assigned | | #Subjects | #DLTs | Recommended Dose | |
|----------|---------------|--------|-----------|-------|------------------|--------|
| | Agent1 | Agent2 | | | Agent1 | Agent2 |
| 1 | 100 | 10 | 3 | 0 | 200 | 20 |
| 2 | | | | | | |
| 3 | | | | | | |
| 4 | | | | | | |
| 5 | | | | | | |
| 6 | | | | | | |
| 7 | | | | | | |
| Total | | | 3 | 0 | | |

Recall that the dose skipping constraints allow for **diagonal escalation** (that is, escalation on both agents at the same time), but the **neighborhood constraint** restricts the set of admissible doses to a single dose level higher or lower than the current dose. Given these constraints, the dose combination (200, 10) is the only combination **closest** to the most probable MTC.

The **MTC plot** allows you to view the most probable MTC, the current dose, the

32 Dose Escalation

recommended dose, and all tried doses.



You may replace the IM dashboard plots with any other plots or corresponding tables, by clicking on the associated icons at the top left of each panel.

Continue in this manner by clicking **Enter Interim Data**, entering the following doses, and the corresponding number of DLTs.

| Cohort # | Dose Assigned | | #Subjects | #DLTs | Recommended Dose | |
|----------|---------------|--------|-----------|-------|------------------|--------|
| | Agent1 | Agent2 | | | Agent1 | Agent2 |
| 1 | 100 | 10 | 3 | 0 | 200 | 20 |
| 2 | 200 | 20 | 3 | 1 | 300 | 20 |
| 3 | 300 | 20 | 3 | 1 | 300 | 10 |
| 4 | 300 | 10 | 3 | 1 | 200 | 20 |
| 5 | 200 | 20 | 3 | 2 | 200 | 10 |
| 6 | 200 | 10 | 3 | 1 | 100 | 20 |
| 7 | 100 | 20 | 3 | 1 | 100 | 30 |
| 8 | 100 | 30 | 3 | 2 | 100 | 20 |
| 9 | | | | | | |
| Total | | | 24 | 9 | | |

Click **Stop Trial**. The recommended MTD combination is 200 mg for Agent 1 and 10

mg for Agent 2. The recommended MTD combination must meet three criteria: (i) closest to MTC from below, (ii) have been experimented on, and (iii) below safety threshold. If there is no dose combination satisfying all three criteria, the MTD will be undetermined.

Volume 4 *Exact Binomial Designs*

| | |
|---|------------|
| <i>33 Introduction to Volume 8</i> | <i>709</i> |
| <i>34 Binomial Superiority One-Sample – Exact</i> | <i>714</i> |
| <i>35 Binomial Superiority Two-Sample – Exact</i> | <i>736</i> |
| <i>36 Binomial Non-Inferiority Two-Sample – Exact</i> | <i>751</i> |
| <i>37 Binomial Equivalence Two-Sample – Exact</i> | <i>767</i> |
| <i>38 Binomial Simon’s Two-Stage Design</i> | <i>774</i> |

33 *Introduction to Volume 8*

This volume describes various cases of clinical trials using binomial endpoints where the asymptotic normal approximation to the test statistic may fail. This is often the case in situations where the trial sample size is too small, however testing and analysis based on the exact binomial distribution would provide valid results. Asymptotic tests may also fail when proportions are very close to the boundary, namely zero or one. These exact methods can be applied in situations where the normal approximation is adequate, in which case the solutions to both the exact and asymptotic method would converge to the same result.

Using exact computations, Chapter 34 deals with the design and interim monitoring of a one sample test of superiority for proportion. The first section discusses a fixed and group sequential design in which an observed binomial response rate is compared to a fixed response rate. The following section illustrates how, for a fixed sample, McNemar's conditional test can be used to compare matched pairs of binomial responses.

Chapters 35 through 37 illustrates how to use East to design two-sample exact tests of superiority, non-inferiority and equivalence, including examples for both the difference and ratio of proportions.

Chapter 38 describes Simon's two stage design in an exact setting, which computes the expected minimal sample size of a trial that may be stopped due to futility or continue to a second stage to further study efficacy and safety.

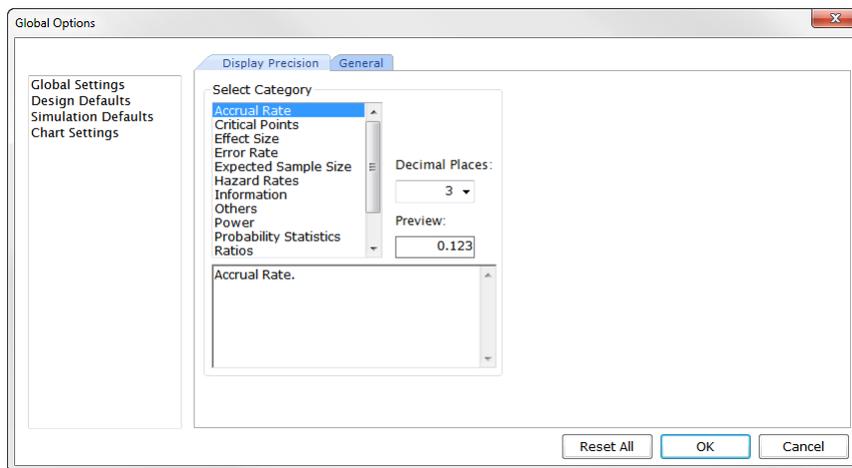
It is important to note that all exact tests work with only integer values for sample size, and will override the **Design Defaults - Common: Do not round off sample size/events** flag in the **Options** menu. Whenever the **Perform Exact Computations** check box is selected in the **Design Input Output** dialog box, resulting sample sizes will be converted to an integer value for all computations, including power and chart/table values.

33 Introduction to Volume 8

33.1 Settings



Click the icon in the **Home** menu to adjust default values in East 6.

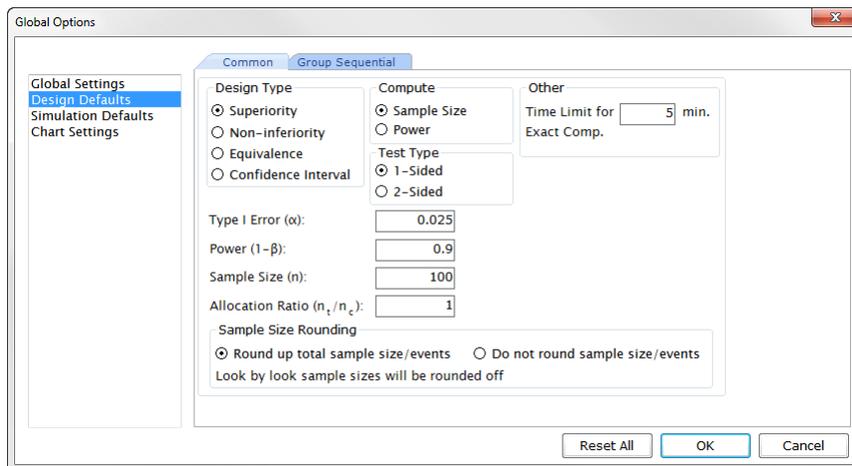


The options provided in the **Display Precision** tab are used to set the decimal places of numerical quantities. The settings indicated here will be applicable to all tests in East 6 under the **Design** and **Analysis** menus.

All these numerical quantities are grouped in different categories depending upon their usage. For example, all the average and expected sample sizes computed at simulation or design stage are grouped together under the category "Expected Sample Size". So to view any of these quantities with greater or lesser precision, select the corresponding category and change the decimal places to any value between 0 to 9.

The **General** tab has the provision of adjusting the paths for storing workbooks, files, and temporary files. These paths will remain throughout the current and future sessions even after East is closed. This is the place where we need to specify the installation directory of the R software in order to use the feature of R Integration in East 6.

The **Design Defaults** is where the user can change the settings for trial design:



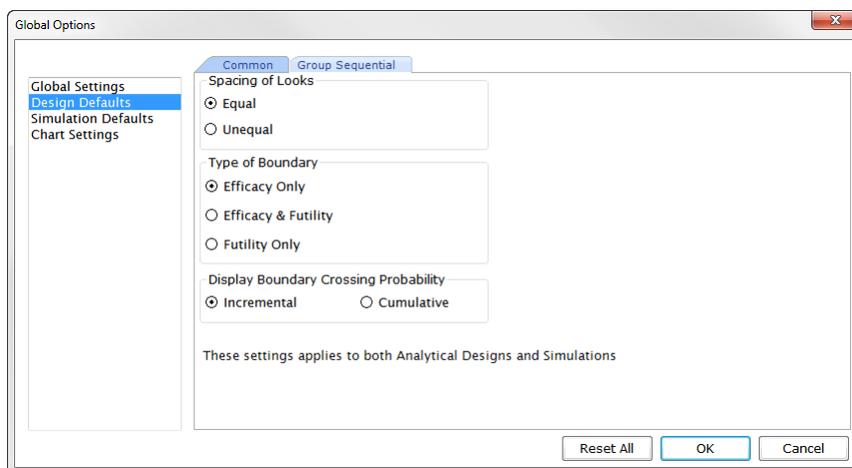
Under the **Common** tab, default values can be set for input design parameters.

You can set up the default choices for the design type, computation type, test type and the default values for type-I error, power, sample size and allocation ratio. When a new design is invoked, the input window will show these default choices.

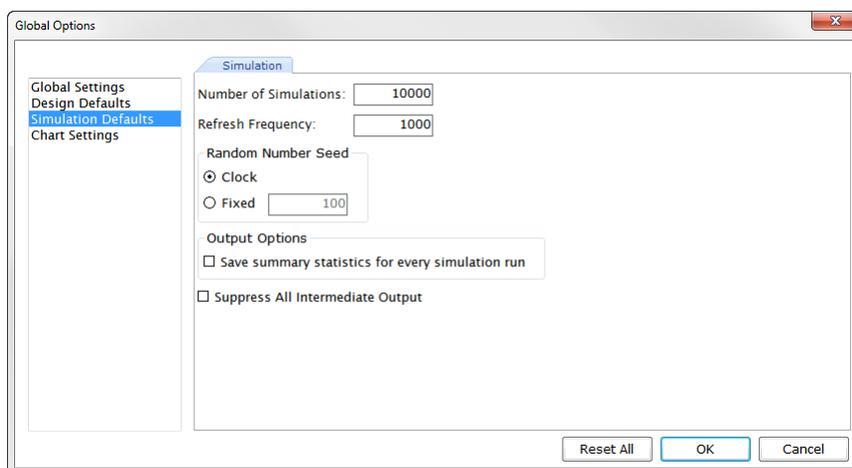
- **Time Limit for Exact Computation**
 This time limit is applicable only to exact designs and charts. Exact methods are computationally intensive and can easily consume several hours of computation time if the likely sample sizes are very large. You can set the maximum time available for any exact test in terms of minutes. If the time limit is reached, the test is terminated and no exact results are provided. Minimum and default value is 5 minutes.
- **Type I Error for MCP**
 If user has selected 2-sided test as default in global settings, then any MCP will use half of the alpha from settings as default since MCP is always a 1-sided test.
- **Sample Size Rounding**
 Notice that by default, East displays the integer sample size (events) by rounding up the actual number computed by the East algorithm. In this case, the look-by-look sample size is rounded off to the nearest integer. One can also see the original floating point sample size by selecting the option "Do not round sample size/events".

33 Introduction to Volume 8

Under the **Group Sequential** tab, defaults are set for boundary information. When a new design is invoked, input fields will contain these specified defaults. We can also set the option to view the Boundary Crossing Probabilities in the detailed output. It can be either Incremental or Cumulative.



Simulation Defaults is where we can change the settings for simulation:

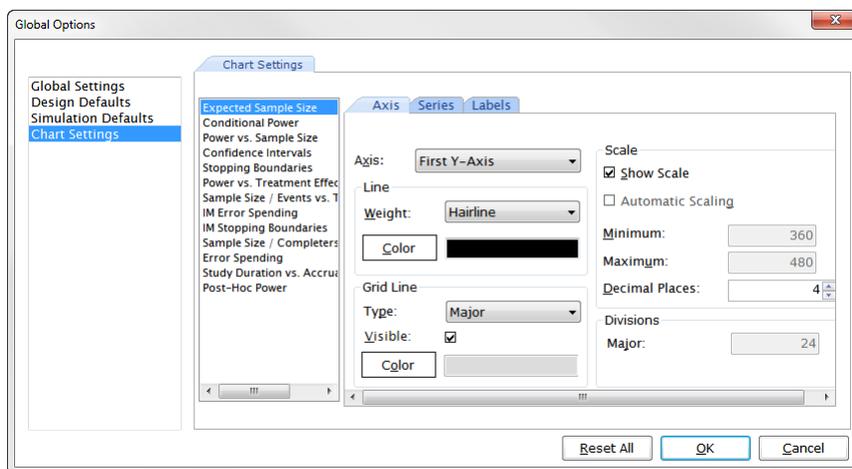


If the checkbox for "Save summary statistics for every simulation" is checked, then East simulations will by default save the per simulation summary data for all the

simulations in the form of a case data.

If the checkbox for "Suppress All Intermediate Output" is checked, the intermediate simulation output window will be always suppressed and you will be directed to the **Output Preview** area.

The **Chart Settings** allows defaults to be set for the following quantities on East6 charts:



We suggest that you do not alter the defaults until you are quite familiar with the software.

34 Binomial Superiority One-Sample – Exact

This chapter deals with the design and interim monitoring of tests involving binomial response rates using exact computations. Section 34.1 discusses a fixed sample and group sequential design in which an observed binomial response rate is compared to a fixed response rate. In Section 34.2, McNemar’s conditional test for comparing matched pairs of binomial responses for a fixed sample is discussed.

34.1 Binomial One-Sample

34.1.1 Trial Design

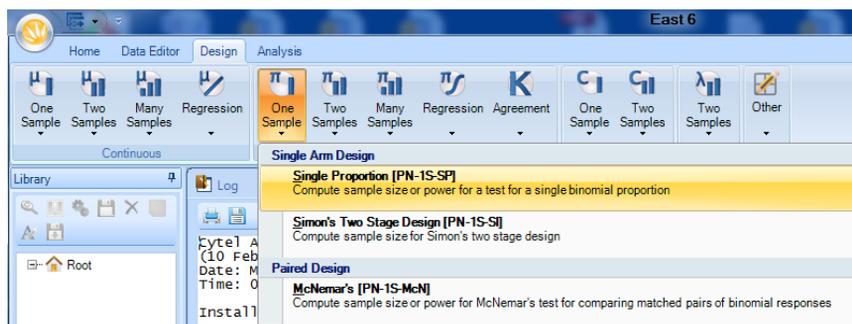
34.1.2 Interim Monitoring

In experimental situations where the variable of interest has a binomial distribution, it may be of interest to determine whether the response rate π differs from a fixed value π_0 . Specifically, we wish to test the null hypothesis $H_0: \pi = \pi_0$ against one-sided alternatives of the form $H_1: \pi > \pi_0$ or $H_1: \pi < \pi_0$. Either the sample size or power is determined for a specified value of π which is consistent with the alternative hypothesis, denoted as π_1 .

34.1.1 Trial Design

Consider a single-arm oncology trial designed to determine if the tumor response rate for a new cytotoxic agent is at least 15%. Thus it is desired to test the null hypothesis $H_0: \pi = 0.15$ against the one-sided alternative hypothesis $H_1: \pi > 0.15$. The trial will be designed using a one-sided test that achieves 80% power at $\pi = \pi_1 = 0.25$ with a level $\alpha = 0.05$ test.

Single-Look Design To illustrate this example, in **East** under the **Design** ribbon for **Discrete** data, click **One Sample** and then choose **Single Arm Design: Single Proportion**:



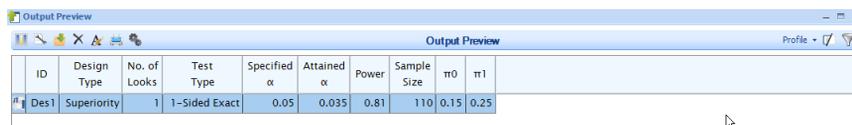
This will launch the following input window:

Leave the default values of **Design Type: Superiority** and **Number of Looks: 1**. In the **Design Parameters** dialog box, select the **Perform Exact Computations** checkbox and enter the following parameters:

- Test Type: 1 sided
- Type I Error (α): 0.05
- Power: 0.8
- Sample Size (n): Computed (select radio button)
- Prop. Response under Null (π_0): 0.15
- Prop. Response under Alt (π_1): 0.25

34 Binomial Superiority One-Sample – Exact

Click **Compute**. The sample size for this design is calculated and the results are shown as a row in the **Output Preview** window:

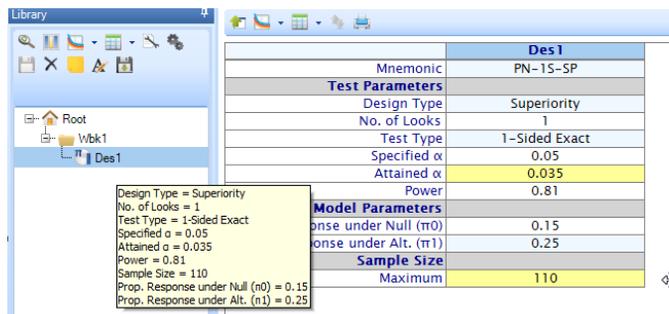


| ID | Design Type | No. of Looks | Test Type | Specified α | Attained α | Power | Sample Size | π_0 | π_1 |
|------|-------------|--------------|---------------|--------------------|-------------------|-------|-------------|---------|---------|
| Des1 | Superiority | 1 | 1-Sided Exact | 0.05 | 0.035 | 0.81 | 110 | 0.15 | 0.25 |

The sample size required in order to achieve 80% power is 110 subjects. Note that because of the discreteness involved in performing exact computations, the attained type-1 error is 0.035, less than the specified value of 0.05. Similarly, the attained power is 0.81, slightly larger than the specified value of 0.80.

As is standard in **East**, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the **Output Preview** toolbar. The design details will be displayed in the upper pane, labeled **Output Summary**. This can be saved to the **Library** by selecting **Des 1** and

clicking the  icon.



| Des1 | |
|---------------------------------------|---------------|
| Mnemonic | PN-15-SP |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided Exact |
| Specified α | 0.05 |
| Attained α | 0.035 |
| Power | 0.81 |
| Model Parameters | |
| Prop. Response under Null (π_0) | 0.15 |
| Prop. Response under Alt. (π_1) | 0.25 |
| Sample Size | |
| Maximum | 110 |

Design Type = Superiority
No. of Looks = 1
Test Type = 1-Sided Exact
Specified α = 0.05
Attained α = 0.035
Power = 0.81
Sample Size = 110
Prop. Response under Null (π_0) = 0.15
Prop. Response under Alt. (π_1) = 0.25

The design details can be displayed by clicking the  icon.

Design: Discrete Endpoint: One-Sample Exact Test - Single Arm Design - Single Proportion

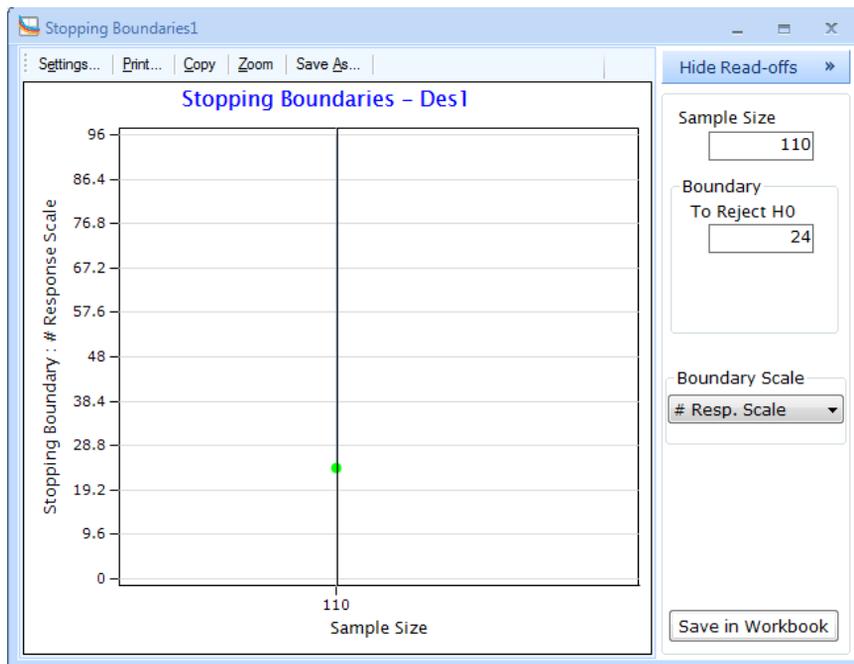
| Test Parameters | |
|---------------------------------------|---------------|
| Design ID | Des1 |
| Design Type | Superiority |
| Number of Looks | 1 |
| Test Type | 1-Sided Exact |
| Specified α | 0.05 |
| Attained α | 0.035 |
| Power | 0.81 |
| Model Parameters | |
| Prop. Response under Null (π_0) | 0.15 |
| Prop. Response under Alt. (π_1) | 0.25 |

- ⊖ Sample Size Information
 - Sample Size (n) 110
- ⊖ Critical Points
 - Critical Point 24

The critical point, or the boundary set for the rejection of H_0 is 24 (on the # response scale). Therefore out of 110 subjects, if the observed number of patients responding to the new treatment exceeds 24, the null hypothesis will be rejected in favor of declaring the new treatment to be superior. This can also be seen using both a response scale and proportion scale in either the **Stopping Boundaries** chart or table, available in the

34 Binomial Superiority One-Sample – Exact

Library



Stopping Boundaries .00 .00

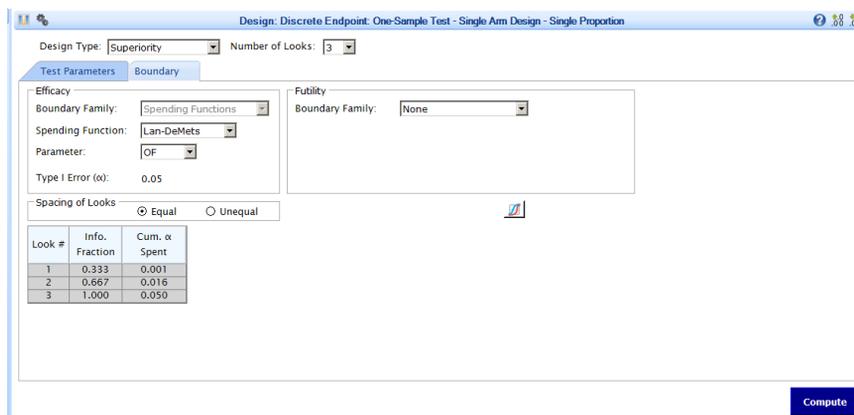
Boundary Scales : # Resp. Scale Settings...

Des1

| Look # | Info. Fraction | Sample Size | Cum. α Spent | Boundaries | |
|--------|----------------|-------------|---------------------|-------------------|--|
| | | | | Efficacy Boundary | |
| 1 | 1 | 110 | 0.05 | 24 | |

Three-Look Design In order to reach an early decision and enter into comparative trials, conduct this single-arm study as a group sequential trial with a maximum of 3

looks. Create a new design by selecting **Des1** in the **Library**, and clicking the  icon on the **Library** toolbar. To generate a study with two interim looks and a final analysis, change the **Number of Looks** from 1 to 3. A **Boundary Info** tab will appear, which allows the specification of parameters for the **Efficacy** and **Futility** boundary families. By default, an efficacy boundary to reject H_0 is selected, however there is no futility boundary to reject H_1 . The **Boundary Family** specified is of the **Spending Functions** type and the default **Spending Function** is the Lan-DeMets (Lan & DeMets, 1983), with **Parameter** OF (O'Brien-Fleming). The default **Spacing of Looks** is **Equal**, therefore the interim analyses will be equally spaced by the number of patients accrued between looks.



Return to the the **Design Parameters** dialog box. The binomial parameters $\pi_0 = 0.15$

34 Binomial Superiority One-Sample – Exact

and $\pi_1 = 0.25$ are already specified. Click **Compute** to generate this exact design:

| Output Preview | | | | | | | | | | | | | |
|----------------|-------------|--------------|---------------|--------------------|-------------------|-------|-------------|---------|---------|------------------|-------------------|------------------|------------------|
| ID | Design Type | No. of Looks | Test Type | Specified α | Attained α | Power | Sample Size | π_0 | π_1 | Spacing of Looks | Efficacy Boundary | Expected SS (H0) | Expected SS (H1) |
| Des1 | Superiority | 1 | 1-Sided Exact | 0.05 | 0.035 | 0.81 | 110 | 0.15 | 0.25 | | | | |
| Des2 | Superiority | 3 | 1-Sided Exact | 0.05 | 0.038 | 0.814 | 110 | 0.15 | 0.25 | Equal | LD (OF) | 109.604 | 90.706 |

The maximum sample size is again 110 subjects with 110 also expected under the null hypothesis $H_0: \pi = 0.15$ and 91 expected when the true value is $\pi=0.25$. Save this design to the **Library**.

| | Des1 | Des2 |
|---------------------------------------|---------------|---------------|
| Mnemonic | PN-1S-SP | PN-1S-SP |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 3 |
| Test Type | 1-Sided Exact | 1-Sided Exact |
| Specified α | 0.05 | 0.05 |
| Attained α | 0.035 | 0.038 |
| Power | 0.81 | 0.814 |
| Model Parameters | | |
| Prop. Response under Null (π_0) | 0.15 | 0.15 |
| Prop. Response under Alt. (π_1) | 0.25 | 0.25 |
| Boundary Parameters | | |
| Spacing of Looks | | Equal |
| Efficacy Boundary | | LD (OF) |
| Sample Size | | |
| Maximum | 110 | 110 |
| Expected Under H0 | | 109.604 |
| Expected Under H1 | | 90.706 |

The details for **Des2** can be displayed by clicking the  icon.

Design: Discrete Endpoint: One-Sample Exact Test - Single Arm Design - Single Proportion

| Test Parameters | |
|---------------------------------------|------------------|
| Design ID | Des2 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided Exact |
| Specified α | 0.05 |
| Attained α | 0.038 |
| Power | 0.814 |
| Model Parameters | |
| Prop. Response under Null (π_0) | 0.15 |
| Prop. Response under Alt. (π_1) | 0.25 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |

☰ Sample Size Information

| Sample Size (n) | Total |
|-----------------|---------|
| Maximum | 110 |
| Expected H1 | 90.706 |
| Expected H0 | 109.604 |

☰ Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries Efficacy # Resp. | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|--------------------------------|---|----------|
| | | | | | Under H0 | Under H1 |
| 1 | 0.336 | 37 | 7.264E-4 | 14 | 5.746E-4 | 0.058 |
| 2 | 0.664 | 73 | 0.016 | 19 | 0.01 | 0.408 |
| 3 | 1 | 110 | 0.05 | 24 | 0.028 | 0.348 |

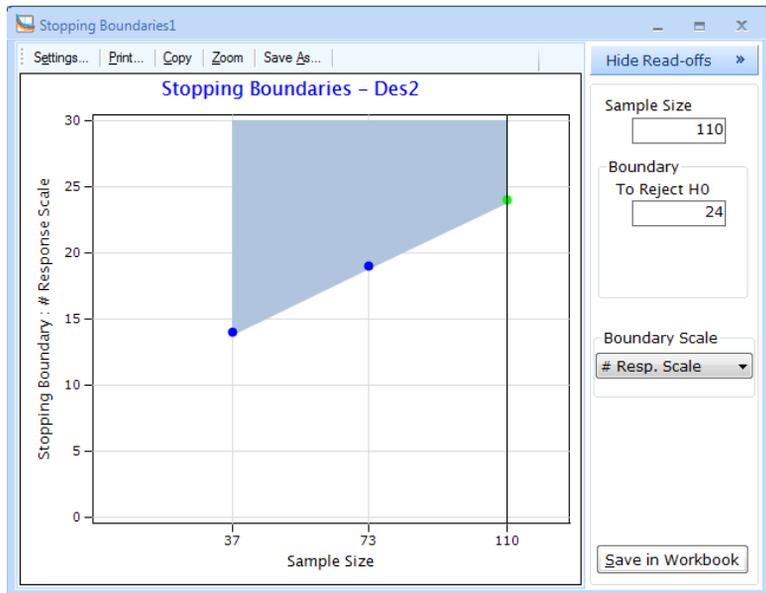
Here we can see the cumulative sample size and cumulative type 1 error (α) spent at each of the three looks. The boundaries set for the rejection of H_0 at each look are 14, 19 and 24 (on the # response scale). For example, at the second look with a cumulative 73 subjects, if the observed number of patients responding to the new treatment exceeds 19, the null hypothesis would be rejected in favor of declaring the new treatment to be superior. In addition, the incremental boundary crossing probabilities under both the null and alternative are displayed for each look.

The cumulative boundary stopping probabilities can also be seen in the **Stopping Boundaries** chart and table. Select **Des 2** in the **Library**, click the  icon and

34 Binomial Superiority One-Sample – Exact

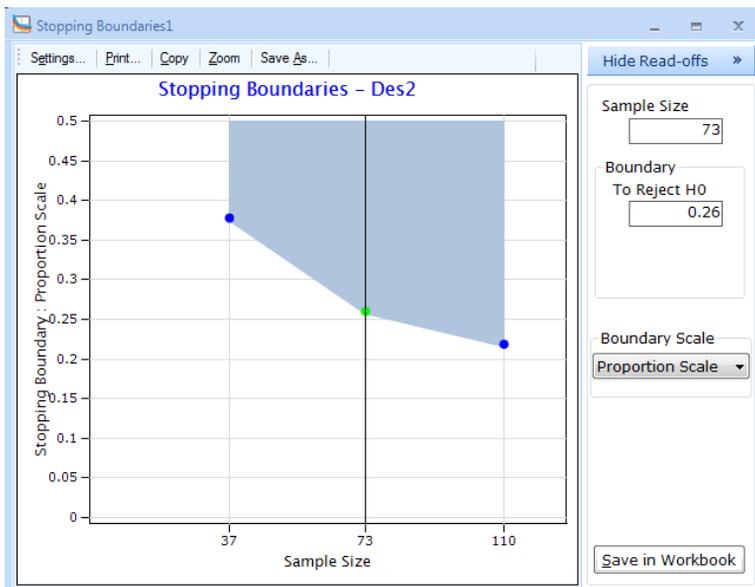
choose **Stopping Boundaries**.

| | Wbk1:Des1 | Wbk1:Des2 |
|---------------------|---------------|---------------|
| Stopping Boundaries | | PN-1S-SP |
| Superiority | 1 | 3 |
| Test Type | 1-Sided Exact | 1-Sided Exact |
| Specified α | 0.05 | 0.05 |



The default scale is **# Response Scale**. The **Proportion Scale** can also be

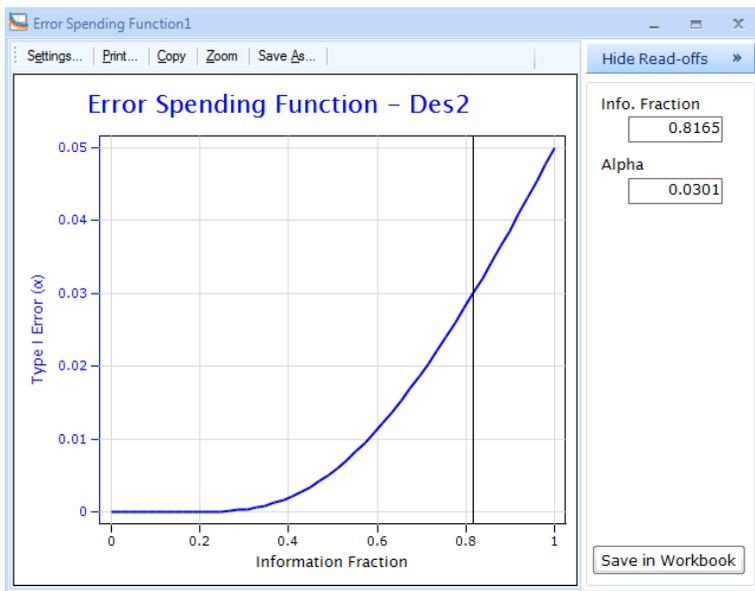
chosen from the drop-down list **Boundary Scale** in the chart.



To examine the **Error Spending** function click the  icon in the **Library** and

34 Binomial Superiority One-Sample – Exact

choose **Error Spending**.



When the sample size for a study is subject to external constraints, power can be computed for a specified maximum sample size. Suppose for the previous design the total sample size is constrained to be at most 80 subjects. Create a new design by editing **Des2** in the **Library**. Change the parameters so that the trial is now designed to compute power for a maximum sample size of 80 subjects, as shown below.

Design: Discrete Endpoint, One-Sample Test - Single Arm Design - Single Proportion

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary

Test Type: 1-Sided

Type I Error (α): 0.05

Power: 0.739

Sample Size (n): 80

Specify Proportion Response

Prop. Response under Null (π_0): 0.15

Prop. Response under Alt (π_1): 0.25

Perform Exact Computations

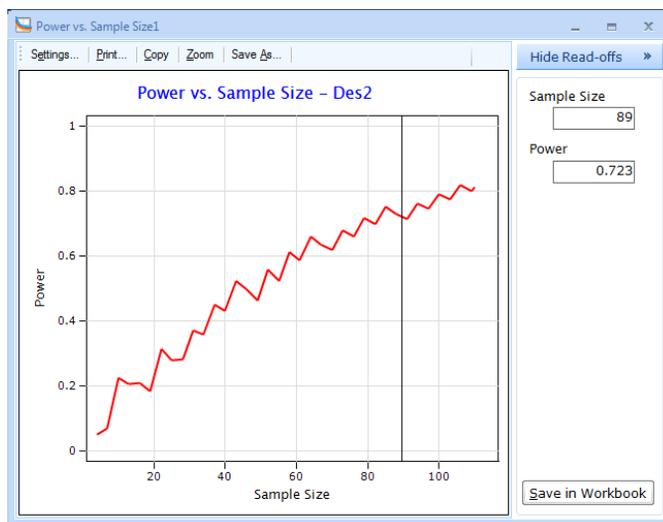
Output Preview

| ID | Design Type | No. of Looks | Test Type | Specified α | Attained α | Power | Sample Size | π_0 | π_1 | Spacing of Looks | Efficacy Boundary | Expected SS (H0) | Expected SS (H1) |
|------|-------------|--------------|---------------|--------------------|-------------------|-------|-------------|---------|---------|------------------|-------------------|------------------|------------------|
| Des1 | Superiority | 1 | 1-Sided Exact | 0.05 | 0.038 | 0.814 | 110 | 0.15 | 0.25 | Equal | LD (OF) | 109.604 | 90.706 |
| Des2 | Superiority | 3 | 1-Sided Exact | 0.05 | 0.038 | 0.814 | 110 | 0.15 | 0.25 | Equal | LD (OF) | 109.604 | 90.706 |
| Des3 | Superiority | 3 | 1-Sided Exact | 0.05 | 0.049 | 0.739 | 80 | 0.15 | 0.25 | Equal | LD (OF) | 79.737 | 70.306 |

The trial now attains only 73.9% power.

| | Des1 | Des2 | Des3 |
|---------------------------------------|---------------|---------------|---------------|
| Mnemonic | PN-1S-SP | PN-1S-SP | PN-1S-SP |
| Test Parameters | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 1 | 3 | 3 |
| Test Type | 1-Sided Exact | 1-Sided Exact | 1-Sided Exact |
| Specified α | 0.05 | 0.05 | 0.05 |
| Attained α | 0.035 | 0.038 | 0.049 |
| Power | 0.81 | 0.814 | 0.739 |
| Model Parameters | | | |
| Prop. Response under Null (π_0) | 0.15 | 0.15 | 0.15 |
| Prop. Response under Alt. (π_1) | 0.25 | 0.25 | 0.25 |
| Boundary Parameters | | | |
| Spacing of Looks | | Equal | Equal |
| Efficacy Boundary | | LD (OF) | LD (OF) |
| Sample Size | | | |
| Maximum | 110 | 110 | 80 |
| Expected Under H0 | | 109,604 | 79,737 |
| Expected Under H1 | | 90,706 | 70,306 |

Power vs Sample size-Sawtooth paradigm Generate the **Power vs. Sample Size** graph for **Des 2**. You will get the following power chart which is commonly described in the literature as a *sawtooth* chart.



This chart illustrates that it is possible to have designs with different sample sizes but all with the same power. What is not apparent is that for designs with the same power, the attained significance level may vary. Upon examination, the sample sizes of 43 and 55 seem to have the same power of about 0.525. The data can also be displayed in a

34 Binomial Superiority One-Sample – Exact

chart form by selecting the  icon in the **Library**, and can be printed from here as well. Compute the power for two new designs based on **Des 2** with sample sizes of 43 and 55 respectively.

| | Des4 | Des5 |
|---------------------------------------|---------------|---------------|
| Mnemonic | PN-1S-SP | PN-1S-SP |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 3 | 3 |
| Test Type | 1-Sided Exact | 1-Sided Exact |
| Specified α | 0.05 | 0.05 |
| Attained α | 0.049 | 0.031 |
| Power | 0.525 | 0.524 |
| Model Parameters | | |
| Prop. Response under Null (π_0) | 0.15 | 0.15 |
| Prop. Response under Alt. (π_1) | 0.25 | 0.25 |
| Boundary Parameters | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) |
| Sample Size | | |
| Maximum | 43 | 55 |
| Expected Under H0 | 42.89 | 54.877 |
| Expected Under H1 | 40.508 | 51.11 |

Although sample sizes of 43 and 55 attain nearly same power, the attained significance levels are different, at 0.049 and 0.031 respectively. Though both are less than the design specification of 0.05, the plan with lower sample size of 43 pays a higher penalty in terms of type-1 error than the plan with a larger sample size of 55.

34.1.2 Interim Monitoring

Consider the interim monitoring of **Des 2**, which has 80% power. Select this design

from the **Library** and click the **IM** icon.

The screenshot shows the 'Interim Monitoring: Des2' window. At the top, there is a toolbar with an 'Enter Interim Data' button. Below the toolbar is a table for data entry:

| Look # | Information Fraction | Cumulative Sample Size | Cumulative Response | Efficacy | CP |
|--------|----------------------|------------------------|---------------------|----------|----|
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |

Below the table, the hypothesis is defined as $H1: \pi_1 = 0.1$. A red instruction reads: 'Select the Look # 1 row for which data entry is desired and click the "Enter Interim Data" button on the toolbar.'

The interface is divided into four main monitoring panels:

- Stopping:** Contains a plot of cumulative response vs. sample size and a table with columns 'Sample Size' and 'Efficacy'.
- Conditional:** Contains a plot of cumulative response vs. sample size and a table with columns 'Prop.' and 'CP'.
- Error Spending:** Contains a plot of error spending vs. information fraction and a table with columns 'Info. Fraction' and ' α '.

Suppose at the first interim look, when 40 subjects have enrolled, the observed cumulative response is 12. Click the **Enter Interim Data** button at the top left of the **IM Interim Monitoring** window. Enter 40 for the **Cumulative Sample Size** and 12

34 Binomial Superiority One-Sample – Exact

for the **Cumulative Response** in the **Test Statistic Calculator** window.

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Cumulative Sample Size:

Cumulative Response:

OK Cancel

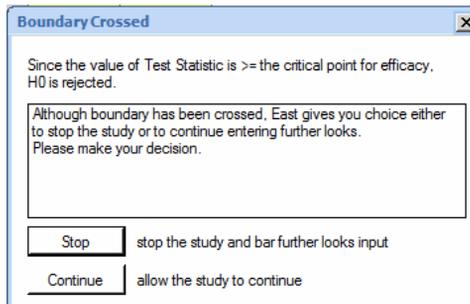
| Look # | Information Fraction | Cumulative Sample Size | Cumulative Response | Efficacy | CP |
|--------|----------------------|------------------------|---------------------|----------|-------|
| 1 | 0.364 | 40 | 12 | 15 | 0.996 |
| 2 | | | | | |
| 3 | | | | | |

Click the "Enter Interim Data" button to enter the Look # 2 data.

| Stopping Boundaries | Sample Size | Efficacy | Conditional Power | Prop. | CP |
|---------------------|-------------|----------|-------------------|-------|-------|
| | 40 | 15 | | 0.15 | 0.404 |
| | | | | 0.165 | 0.539 |
| | | | | 0.184 | 0.687 |
| | | | | 0.202 | 0.806 |
| | | | | 0.22 | 0.889 |
| | | | | 0.239 | 0.942 |

| Error Spending Function | Info. Fraction | α |
|-------------------------|----------------|----------|
| | 0.364 | 0 |

At the second interim monitoring time point when 80 subjects have enrolled, suppose the cumulative responses increases to 20. Again click the **Enter Interim Data** button at the top left of the **IM Interim Monitoring** window. Enter 80 for the **Cumulative Sample Size** and 20 for the **Cumulative Response** in the **Test Statistic Calculator** window. This will result in the following message:



It can be concluded that $\pi > 0.15$ and the trial should be terminated. Clicking on **Stop** results in the final analysis.

Wbk1:Des2:Interim Monitoring

| Look # | Information Fraction | Cumulative Sample Size | Cumulative Response | Efficacy | CP |
|--------|----------------------|------------------------|---------------------|----------|-------|
| 1 | 0.364 | 40 | 12 | 15 | 0.996 |
| 2 | 0.727 | 80 | 20 | 20 | NA |

Click the "Edit Interim Data" button to edit the Look # 2 data.

Stopping Boundaries

| Sample Size | Efficacy |
|-------------|----------|
| 40 | 15 |
| 80 | 20 |

Error Spending Function

| Info. Fraction | α |
|----------------|----------|
| 0.364 | 0 |
| 0.727 | 0.013 |

Final Inference

| | |
|---|-------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.013 |
| Adj. Pt. Est. for π | 0.245 |
| Adj. 90% CI for π | |
| Upper Confidence Bound | 0.342 |
| Lower Confidence Bound | 0.172 |

34.2 McNemar's Conditional Exact Test

McNemar's conditional test is used in experimental situations where paired comparisons are observed. In a typical application, two binary response measurements are made on each subject – perhaps from two different treatments, or from two different time points. For example, in a comparative clinical trial, subjects are matched on baseline demographics and disease characteristics and then randomized with one subject in the pair receiving the experimental treatment and the other subject receiving the control. Another example is the crossover clinical trial in which each subject receives both treatments. By random assignment, some subjects receive the experimental treatment followed by the control while others receive the control followed by the experimental treatment. Let π_c and π_t denote the response probabilities for the control and experimental treatments, respectively. The probability parameters for this test are displayed in Table 34.1.

34 Binomial Superiority One-Sample – Exact

Table 34.1: A 2 x 2 Table of Probabilities for McNemar’s Conditional Exact Test

| Control | Experimental | | Total Probability |
|-------------------|--------------|------------|-------------------|
| | No Response | Response | |
| No Response | π_{00} | π_{01} | $1 - \pi_c$ |
| Response | π_{10} | π_{11} | π_c |
| Total Probability | $1 - \pi_t$ | π_t | 1 |

The null hypothesis

$$H_0: \pi_c = \pi_t$$

is tested against the alternative hypothesis

$$H_1: \pi_c > \pi_t$$

(or $H_1: \pi_c < \pi_t$) for the one-sided testing problem. Since $\pi_t = \pi_c$ if and only if $\pi_{01} = \pi_{10}$, the null hypothesis is also expressed as

$$H_0: \pi_{01} = \pi_{10} ,$$

is tested against corresponding one-sided alternative. The power of this test depends on two quantities:

1. The difference between the two discordant probabilities (which is also the difference between the response rates of the two treatments)

$$\delta = \pi_{01} - \pi_{10} = \pi_t - \pi_c ;$$

2. The sum of the two discordant probabilities

$$\xi = \pi_{10} + \pi_{01} .$$

East accepts these two parameters as inputs at the design stage.

34.2.1 Trial Design

Consider a trial in which we wish to determine whether a transdermal delivery system (TDS) can be improved with a new adhesive. Subjects are to wear the old TDS (control) and new TDS (experimental) in the same area of the body for one week each. A response is said to occur if the TDS remains on for the entire one-week observation

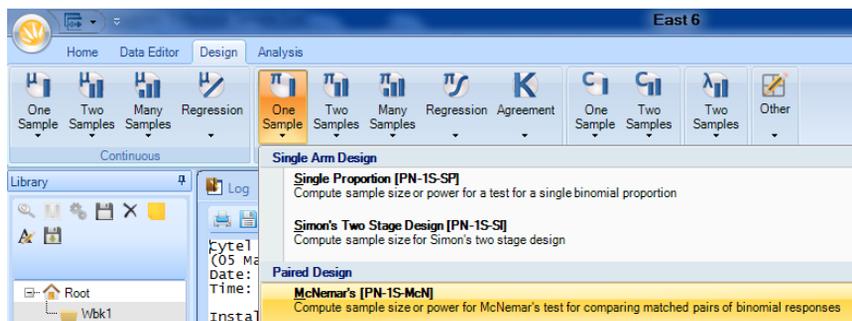
period. From historical data, it is known that control has a response rate of 85% ($\pi_c = 0.85$). It is hoped that the new adhesive will increase this to 95% ($\pi_t = 0.95$). Furthermore, of the 15% of the subjects who did not respond on the control, it is hoped that 87% will respond on the experimental system. That is, $\pi_{01} = 0.87 \times 0.15 = 0.13$. Based on these data, we can fill in all the entries of Table 34.1 as displayed in Table 34.2.

Table 34.2: McNemar Probabilities for the TDS Trial

| Control | Experimental | | Total Probability |
|-------------------|--------------|----------|-------------------|
| | No Response | Response | |
| No Response | 0.02 | 0.13 | 0.15 |
| Response | 0.03 | 0.82 | 0.85 |
| Total Probability | 0.05 | 0.95 | 1 |

As it is expected that the new adhesive will increase the adherence rate, the comparison is posed as a one-sided testing problem, testing $H_0: \pi_c = \pi_t$ against $H_1: \pi_c < \pi_t$ at the 0.05 level. We wish to determine the sample size to have 90% power for the values displayed in Table 34.2.

To illustrate this example, in East under the Design ribbon for Discrete data, click **One Sample** and then choose **Paired Design: McNemar's**:



34 Binomial Superiority One-Sample – Exact

This will launch the following input window:

Design: Discrete Endpoint: One-Sample Test - Paired Design - McNemar's

Design Type: Superiority Number of Looks: 1

Design Parameters

Test Type: 1-Sided

Type I Error (α): 0.025

Power: 0.9

Sample Size (n): Computed

Difference in Probabilities (δ_1): 0.1

Prop. of Discordant Pairs (ξ): 0.2

Perform Exact Computations

Probability Allocation: Row = Control, Column = Treatment

| | No Response | Response | Total |
|-------------|-------------|------------|-------------|
| No Response | π_{00} | π_{01} | $1 - \pi_c$ |
| Response | π_{10} | π_{11} | π_c |
| Total | $1 - \pi_t$ | π_t | 1 |

Compute

Leave the default values of **Design Type: Superiority** and **Number of Looks: 1**. In the **Design Parameters** dialog box, select the **Perform Exact Computations** checkbox and enter the following parameters:

Test Type: 1 sided
 Type I Error (α): 0.05
 Power: 0.9
 Sample Size (n): Computed (select radio button)
 Difference in Probabilities ($\delta_1 = \pi_t - \pi_c$): 0.1
 Prop. of Discordant Pairs ($\xi = \pi_{01} + \pi_{10}$): 0.16

Design: Discrete Endpoint: One-Sample Test - Paired Design - McNemar's

Design Type: Superiority Number of Looks: 1

Design Parameters

Test Type: 1-Sided

Type I Error (α): 0.05

Power: 0.9

Sample Size (n): Computed

Difference in Probabilities (δ_1): 0.1

Prop. of Discordant Pairs (ξ): 0.16

Perform Exact Computations

Probability Allocation: Row = Control, Column = Treatment

| | No Response | Response | Total |
|-------------|-------------|------------|-------------|
| No Response | π_{00} | π_{01} | $1 - \pi_c$ |
| Response | π_{10} | π_{11} | π_c |
| Total | $1 - \pi_t$ | π_t | 1 |

Compute

Click **Compute**. The sample size for this design is calculated and the results are shown

as a row in the **Output Preview** window:

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | Sample Size | δ_1 | ξ |
|-------|-------------|--------------|---------------|--------------------|-------|-------------|------------|-------|
| Des 1 | Superiority | 1 | 1-Sided Exact | 0.05 | 0.9 | 139 | 0.1 | 0.16 |

The sample size required in order to achieve 90% power is 139 subjects.

As is standard in **East**, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the icon in the **Output Preview** toolbar. The design details will be displayed in the upper pane, labeled **Output Summary**. This can be saved to the **Library** by selecting **Des 1** and clicking the icon.

| Des6 | |
|--|---------------|
| Mnemonic | PN-1S-McN |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided Exact |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Difference in Probabilities (δ_1) | 0.1 |
| Prop. of Discordant Pairs (ξ) | 0.16 |
| Sample Size | |
| Maximum | 139 |

34 Binomial Superiority One-Sample – Exact

The design details can be displayed by clicking the  icon.

Design: Discrete Endpoint: One-Sample Exact Test - Paired Design - McNemar's

| Test Parameters | |
|-------------------------------------|---------------|
| Design ID | Des6 |
| Design Type | Superiority |
| Number of Looks | 1 |
| Test Type | 1-Sided Exact |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| $\delta = \pi_t - \pi_c$ | |
| Under H0 | 0 |
| Under H1 | 0.1 |
| Prop. of Discordant Pairs (ξ) | 0.16 |

Sample Size Information

Sample Size (n) 139

Critical Points

Critical Point 1.645

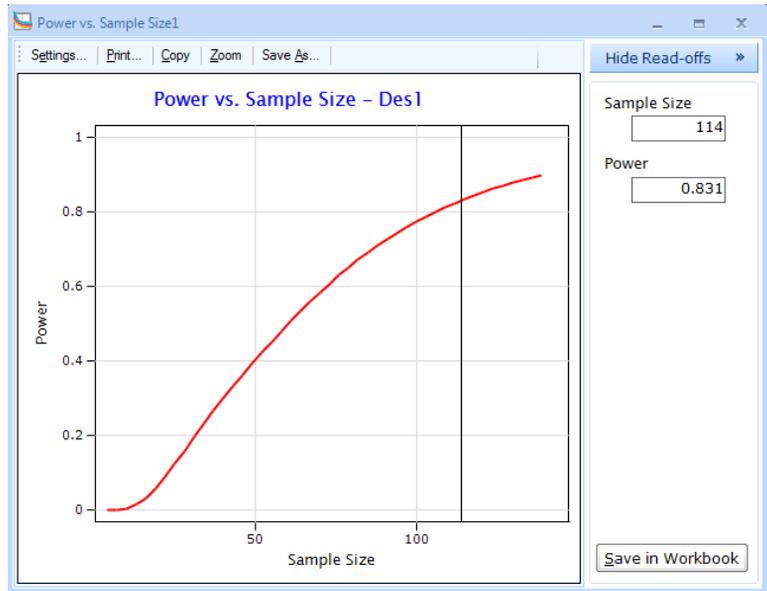
Sample sizes have been rounded.

Due to discreteness the displayed sample size may not be unique. Use the 'Power Vs Sample Size' chart from charts menu to find all other sample sizes that guarantee desired power.

The critical point, or the boundary set for the rejection of H_0 is 1.645

It is important to note that in this exact computation the displayed sample size may not be unique due to the discreteness of the distribution. This can be seen in the **Power Vs Sample Size** graph, which is a useful tool along with its corresponding table, and can be used to find all other sample sizes that guarantee the desired power. These visual

tools are available in the **Library** under the **Plots** and **Tables** menus.



35 *Binomial Superiority Two-Sample – Exact*

In many experiments based on binomial data, the aim is to compare independent samples from two populations in terms of the proportion of sampling units presenting a given trait. In medical research, outcomes such as the proportion of patients responding to a therapy, developing a certain side effect, or requiring specialized care, would satisfy this definition. Exact tests support the design and monitoring of clinical trials in which this comparison is based on either the difference of proportions or ratio of proportions of the two populations. These two cases are discussed in Sections 35.1, and 35.2 respectively.

Caution: The methods presented in this chapter are computationally intensive and could consume several hours of computer time if the exact sample sizes are very large. Here are some guidelines:

1. Estimate the likely sample size under the Exact method by first determining the asymptotic sample size
2. If the exact sample size is likely to be larger than 1000, computing power is preferable to computing the sample size

35.1 *Difference of Two Binomial Proportions*

35.1.1 *Trial Design*

Let π_c and π_t denote the binomial probabilities for the control and treatment arms, respectively, and let $\delta = \pi_t - \pi_c$. Interest lies in testing the null hypothesis that $\delta = 0$ against one and two-sided alternatives.

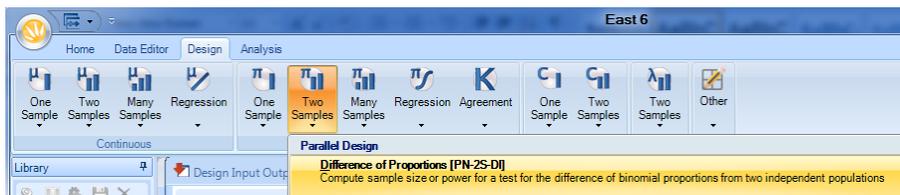
The technical details of the sample size computations for this option are given in Appendix V.

35.1.1 *Trial Design*

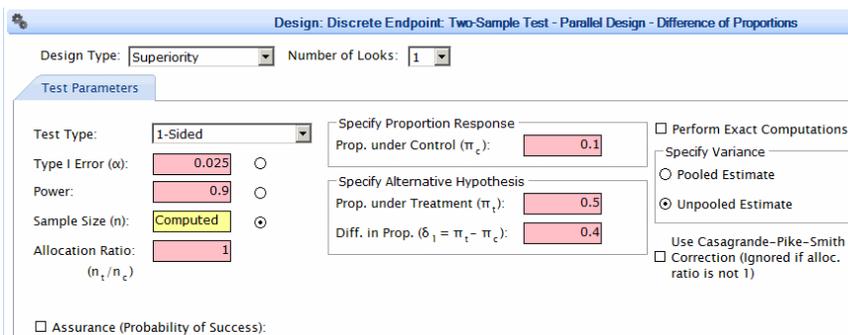
In a clinical study, an experimental drug coded Y73 is to be compared with a control drug coded X39 to treat chronic hepatitis B infected adult patients. The primary end point is histological improvement as determined by Knodell Scores at week 48 of treatment period. It is estimated that the proportion of patients who are likely to show histological improvement under treatment X39 to be 25% and under the treatment Y73, as much as 60%. A one-sided fixed sample study is to be designed with $\alpha = 0.05$ and 90% power.

Single Look Design To illustrate this example, in **East** under the **Design** ribbon for **Discrete** data, click **Two Samples** and then choose **Parallel Design: Difference of**

Proportions:



This will launch the following input window:



The goal of this study is to test the null hypothesis, H_0 , that the X39 and Y73 arms both have an event rate of 25%, versus the alternative hypothesis, H_1 , that Y73 increases the event rate by 35%, from 25% to 60%. This will be a one-sided test with a single fixed look at the data, a type-1 error of $\alpha = 0.05$ and a power of $(1 - \beta) = 0.9$.

Leave the default values of **Design Type: Superiority** and **Number of Looks: 1**. In the **Design Parameters** dialog box, select the **Perform Exact Computations** checkbox and enter the following parameters:

- Test Type: 1 sided (required)
- Type 1 Error (α): 0.05
- Power: 0.9
- Sample Size (n): Computed (select radio button)
- Prop. under Control (π_c): 0.25
- Prop. under Treatment (π_t): 0.6
- Diff. in Prop. ($\delta_1 = \pi_t - \pi_c$): (will be calculated)

35 Binomial Superiority Two-Sample – Exact

Click **Compute**. The sample size for this design is calculated and the results are shown as a row in the **Output Preview** window:

| Output Preview | | | | | | | | | | | | |
|----------------|-------------|--------------|---------------|--------------------|-------------------|-------|-------|-------------|---------|------------------------|------------|--|
| ID | Design Type | No. of Looks | Test Type | Specified α | Attained α | Power | nt/nc | Sample Size | π_c | Prop. Treatment (Alt.) | δ_1 | |
| Des 1 | Superiority | 1 | 1-Sided Exact | 0.05 | 0.049 | 0.905 | 1 | 68 | 0.25 | 0.6 | 0.35 | |

The sample size required in order to achieve 90% power is 68 subjects. Note that because of the discreteness involved in performing exact computations, the attained type-1 error is 0.049, slightly less than the specified value of 0.05. Similarly, the attained power is 0.905, slightly larger than the specified value of 0.90.

As is standard in **East**, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the **Output Preview** toolbar. The design details will be displayed in the upper pane, labeled **Output Summary**. This can be saved to the **Library** by selecting **Des 1** and

clicking the  icon.

| Des7 | |
|--|---------------|
| Mnemonic | PN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided Exact |
| Specified α | 0.05 |
| Attained α | 0.049 |
| Power | 0.905 |
| Model Parameters | |
| Allocation Ratio (n_t/n_c) | 1 |
| Proportion under Control (π_c) | 0.25 |
| Proportion under Treatment (π_t) | 0.6 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0.35 |
| Sample Size | |
| Maximum | 68 |

The design details can be displayed by clicking the  icon.

Design: Discrete Endpoint: Two-Sample Exact Test - Parallel Design - Difference of Proportions

| Test Parameters | |
|-----------------------------------|---------------|
| Design ID | Des1 |
| Design Type | Superiority |
| Number of Looks | 1 |
| Test Type | 1-Sided Exact |
| Specified α | 0.05 |
| Attained α | 0.049 |
| Power | 0.905 |
| Model Parameters | |
| Prop. under Control (π_c) | 0.25 |
| Prop. under Treatment (π_t) | 0.6 |
| $\delta = \pi_t - \pi_c$ | |
| Under H0 | 0 |
| Under H1 | 0.35 |
| Allocation Ratio (n_t/n_c) | 1 |

Sample Size Information

| | |
|-----------------------------|----|
| Sample Size (n) | 68 |
| Treatment (n _t) | 34 |
| Control (n _c) | 34 |

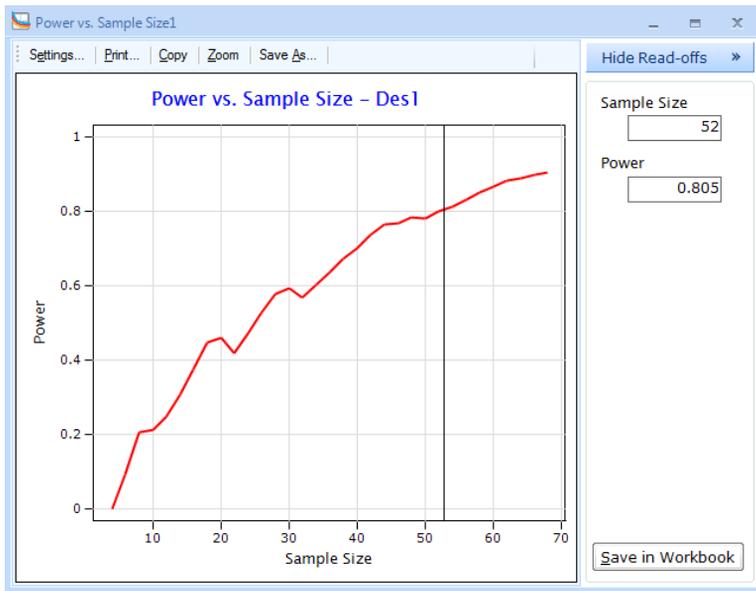
Critical Points

| | |
|---------------------|-------|
| Critical Point | 1.715 |
| Attained at π_c | 0.371 |

It is important to note that in this exact computation the displayed sample size may not be unique due to the discreteness of the distribution. This can be seen in the **Power Vs Sample Size** graph, which is a useful tool along with its corresponding table, and can be used to find all other sample sizes that guarantee the desired power. These visual

35 Binomial Superiority Two-Sample – Exact

tools are available in the **Library** under the **Plots** and **Tables** menus.



In tabular form:

 **Power Vs. Sample Size** ↑0 ↓0
.00 .00

Range for Sample Size

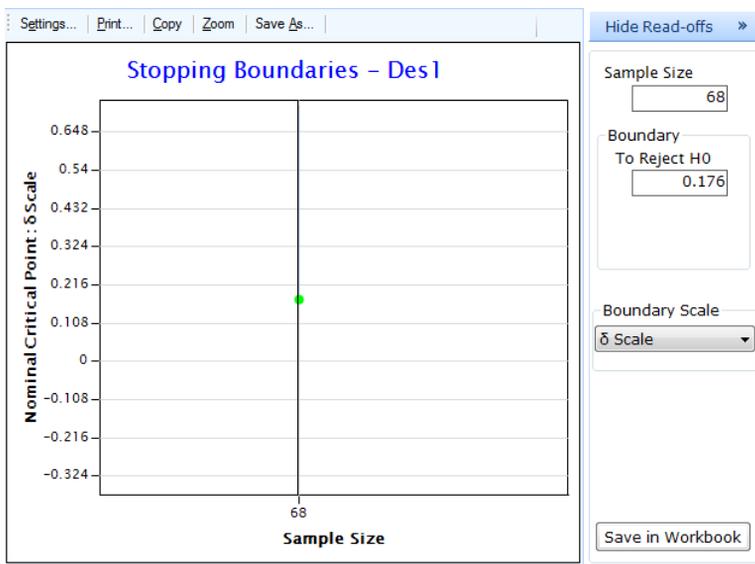
| | | | |
|------|----|-----------|-----------------|
| From | To | Step Size | Tabulate |
| 40 | 68 | 2 | |

| Sample Size | Power_Des 1 |
|-------------|-------------|
| 40 | 0.702 |
| 42 | 0.736 |
| 44 | 0.766 |
| 46 | 0.769 |
| 48 | 0.785 |
| 50 | 0.781 |
| 52 | 0.801 |
| 54 | 0.812 |
| 56 | 0.833 |
| 58 | 0.851 |
| 60 | 0.868 |
| 62 | 0.883 |
| 64 | 0.888 |
| 66 | 0.897 |
| 68 | 0.905 |

Save as Case Data

35 Binomial Superiority Two-Sample – Exact

The critical point, or the boundary set for the rejection of H_0 is 1.715 attained at $\pi_U = 0.371$ (on the Z scale) and 0.176 (on the δ scale). If the observed test statistic exceeds this boundary the null will be rejected in favor of declaring the new treatment to be superior. This can also be seen in the **Stopping Boundaries** chart and table, available in the **Library**.



Stopping Boundaries

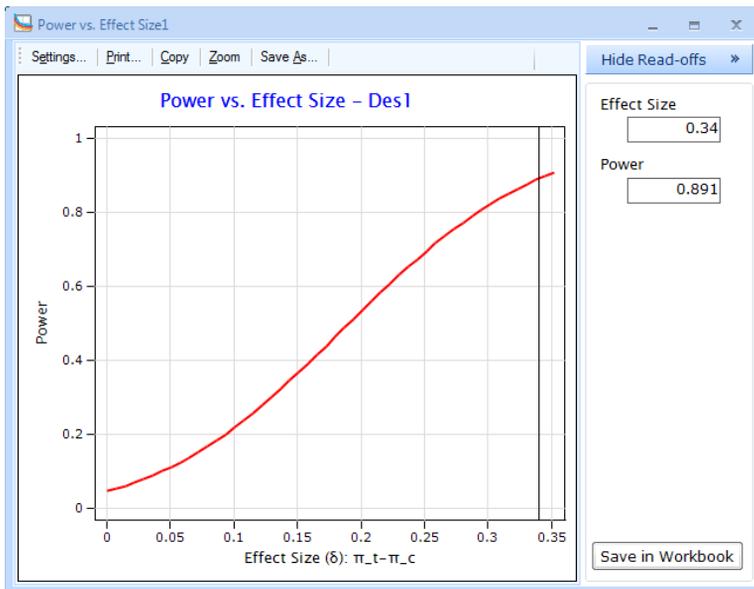
Boundary Scales : δ Scale

Des1

| Look # | Info. Fraction | Sample Size | Cum. α Spent | Boundaries | |
|--------|----------------|-------------|---------------------|------------|----------|
| | | | | Efficacy | Boundary |
| 1 | 1 | 68 | 0.05 | 0.176 | |

The **Power vs. Treatment Effect** chart dynamically generates power under this design for all values of treatment effect $\delta = \pi_t - \pi_c$. Here it is easy to see how as treatment effect size increases (H_1 : alternative treatment is superior) the power of the study

reaches the desired 90%. This is available in tabular form as well.



35.2 Ratio of Two Binomial Proportions

Let π_c and π_t denote the binomial probabilities for the control and treatment arms, respectively, and let $\rho = \pi_t/\pi_c$. It is of interest to test the null hypothesis that $\rho = 1$ against a one-sided alternative.

The technical details of the sample size computations for this option are given in Appendix V.

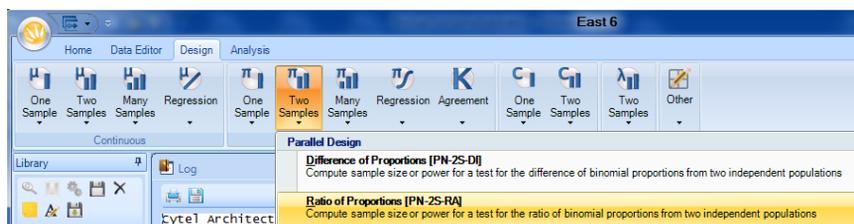
35.2.1 Trial Design

In a clinical study, an experimental drug coded Y73 is to be compared with a control drug coded X39 to treat chronic hepatitis B infected adult patients. The primary end point is histological improvement as determined by Knodell Scores at week 48 of treatment period. It is estimated that the proportion of patients who are likely to show histological improvement under treatment coded X39 to be 25% and under the treatment coded Y73 as much as 60%, that is 2.4 times the rate for X39. A single look, one-sided fixed sample study is to be designed with $\alpha = 0.05$ and 90% power.

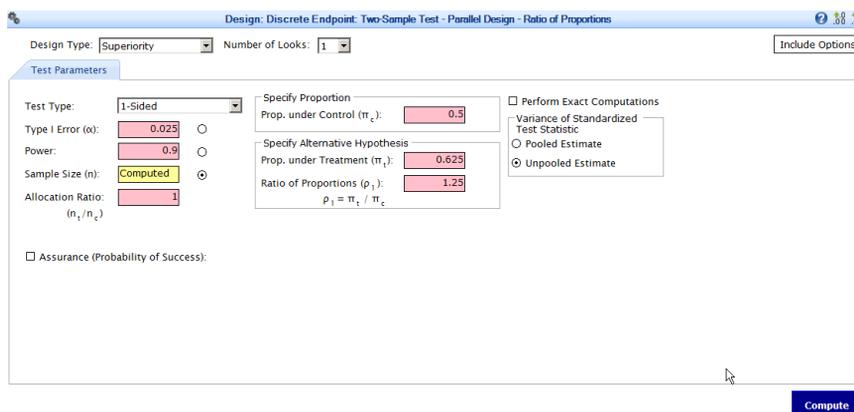
Single Look Design

35 Binomial Superiority Two-Sample – Exact

To illustrate this example, in **East** under the **Design** ribbon for **Discrete** data, click **Two Samples** and then choose **Parallel Design: Ratio of Proportions**:



This will launch the following input window:



Leave the default values of **Design Type: Superiority** and **Number of Looks: 1**. In the **Design Parameters** dialog box, select the **Perform Exact Computations** checkbox and enter the following parameters:

- Test Type: 1 sided (required)
- Type 1 Error (α): 0.05
- Power: 0.9
- Sample Size (n): Computed (select radio button)
- Prop. under Control (π_c): 0.25
- Prop. under Treatment (π_t): (will be calculated to be 0.6)
- Ratio of Proportions (ρ_1): 2.4

Click **Compute**. The sample size for this design is calculated and the results are shown as a row in the **Output Preview** window:

| Output Preview | | | | | | | | | | | | |
|----------------|-------------|--------------|---------------|--------------------|-------------------|-------|-------------|---------|---------|------------------------|----------|--|
| ID | Design Type | No. of Looks | Test Type | Specified α | Attained α | Power | Sample Size | nt / nc | π_c | Prop. Treatment (Alt.) | ρ_1 | |
| Des8 | Superiority | 1 | 1-Sided Exact | 0.05 | 0.046 | 0.903 | 72 | 1 | 0.25 | 0.6 | 2.4 | |

The sample size required in order to achieve 90% power is 72 subjects. Note that because of the discreteness involved in performing exact computations, the attained type-1 error is 0.046, less than the specified value of 0.05. Similarly, the attained power is 0.903, slightly larger than the specified value of 0.90.

As is standard in **East**, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the **Output Preview** toolbar. The design details will be displayed in the upper pane, labeled **Output Summary**. This can be saved to the **Library** by selecting **Des 1** and

35 Binomial Superiority Two-Sample – Exact

clicking the  icon.

| Des8 | |
|--|---------------|
| Mnemonic | PN-2S-RA |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided Exact |
| Specified α | 0.05 |
| Attained α | 0.046 |
| Power | 0.903 |
| Model Parameters | |
| Allocation Ratio (n_t/n_c) | 1 |
| Proportion under Control (π_c) | 0.25 |
| Proportion under Treatment (π_t) | 0.6 |
| Ratio of Proportions (π_t / π_c) | 2.4 |
| Sample Size | |
| Maximum | 72 |

Design details can be displayed by clicking the  icon.







Design: Discrete Endpoint: Two-Sample Exact Test - Parallel Design - Ratio of Proportions

| Test Parameters | |
|-----------------------------------|---------------|
| Design ID | Des1 |
| Design Type | Superiority |
| Number of Looks | 1 |
| Test Type | 1-Sided Exact |
| Specified α | 0.05 |
| Attained α | 0.046 |
| Power | 0.903 |
| Model Parameters | |
| Prop. under Control (π_c) | 0.25 |
| Prop. under Treatment (π_t) | 0.6 |
| $\rho = \pi_t / \pi_c$ | |
| Under H0 | 1 |
| Under H1 | 2.4 |
| Allocation Ratio (n_t/n_c) | 1 |

Sample Size Information

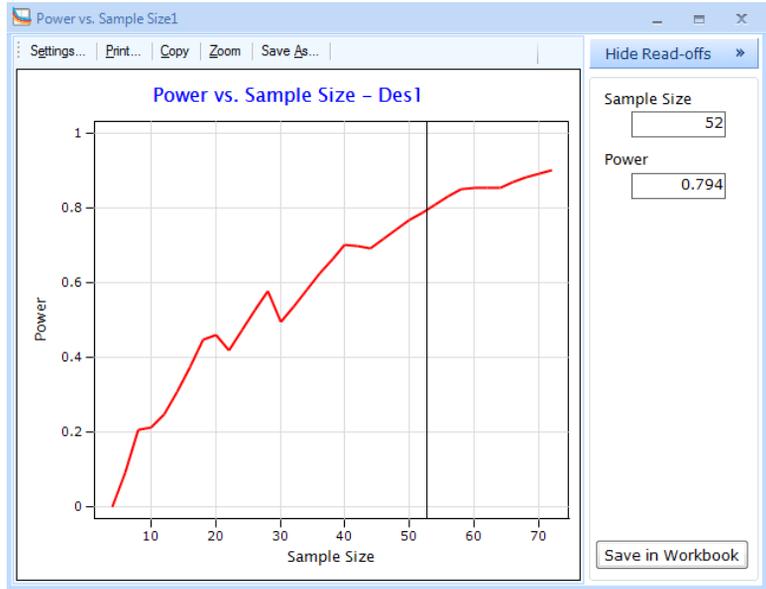
| | |
|-----------------------------|----|
| Sample Size (n) | 72 |
| Treatment (n _t) | 36 |
| Control (n _c) | 36 |

Critical Points

| | |
|---------------------|-------|
| Critical Point | 1.813 |
| Attained at π_c | 0.057 |

It is important to note that in this exact computation the displayed sample size may not be unique due to the discreteness of the distribution. This can be seen in the **Power Vs Sample Size** graph, which is a useful tool along with its corresponding table, and can be used to find all other sample sizes that guarantee the desired power. These visual

tools are available in the **Library** under the **Plots** and **Tables** menus.



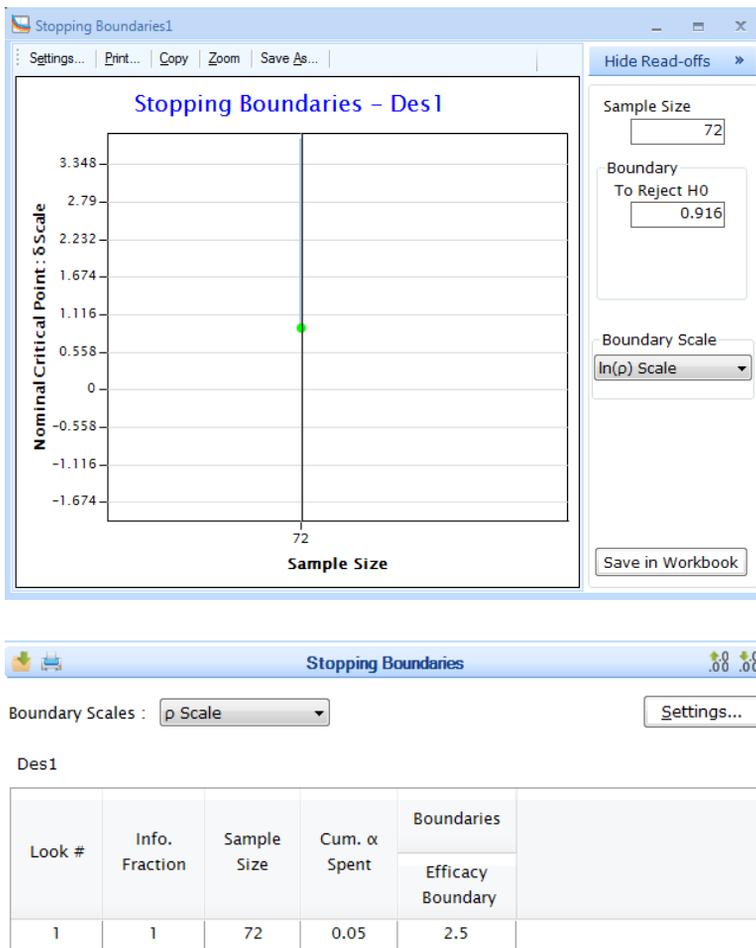
35 Binomial Superiority Two-Sample – Exact

In tabular form:

| Sample Size | Power_Des1 |
|-------------|------------|
| 38 | 0.665 |
| 40 | 0.702 |
| 42 | 0.699 |
| 44 | 0.693 |
| 46 | 0.718 |
| 48 | 0.743 |
| 50 | 0.767 |
| 52 | 0.786 |
| 54 | 0.809 |
| 56 | 0.831 |
| 58 | 0.85 |
| 60 | 0.853 |
| 62 | 0.854 |
| 64 | 0.855 |
| 66 | 0.868 |
| 68 | 0.881 |
| 70 | 0.891 |
| 72 | 0.903 |

The critical point, or the boundary set for the rejection of H_0 is 1.813 (on the Z scale). If the observed test statistic exceeds this boundary the null will be rejected in favor of declaring the new treatment to be superior. This boundary can be seen in terms of the observed ratio (0.916 on the $\ln(\rho)$ scale and 2.5 on the ρ scale) in the **Stopping**

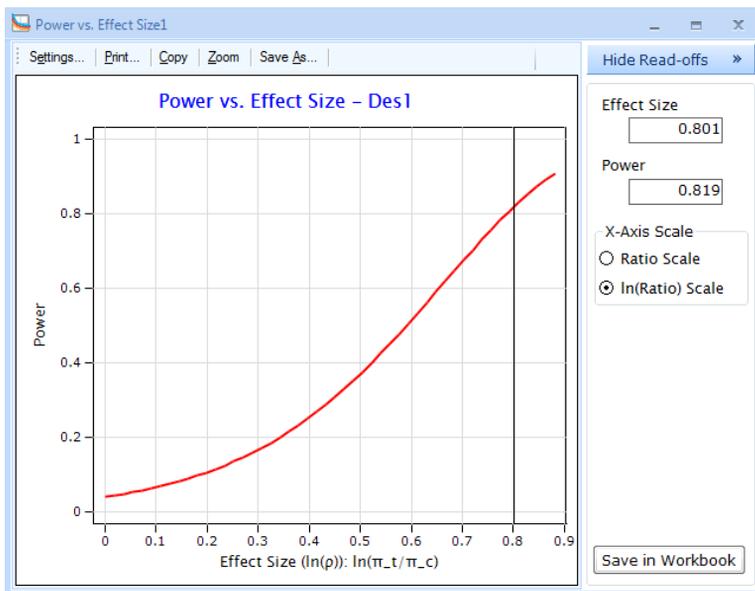
Boundaries chart and table, available in the **Library**.



The **Power vs. Treatment Effect** chart dynamically generates power under this design for all values of treatment effect (ratio or ln(ratio) scale). Here it is easy to see how as the ratio (treatment effect size) increases (H_1 : the new treatment is superior) the power

35 Binomial Superiority Two-Sample – Exact

of the study reaches the desired 0.9%. This is available in tabular form as well.



36 Binomial Non-Inferiority Two-Sample – Exact

In a non-inferiority trial, the goal is to establish that the response rate of an experimental treatment is **no worse than** that of an established control. A therapy that is demonstrated to be non-inferior to the current standard therapy might be an acceptable alternative if, for instance, it is easier to administer, cheaper, or less toxic. Non-inferiority trials are designed by specifying a non-inferiority margin, which is the acceptable amount by which the response rate on the experimental arm can be less than the response rate on the control arm. If the experimental response rate falls within this margin, the new treatment can claim to be non-inferior. This chapter presents the design of non-inferiority trials in which this margin is expressed as either the difference between or the ratio of two binomial proportions. The difference is examined in Section 36.1 and is followed by two formulations for the ratio in Section 36.2.

Caution: The methods presented in this chapter are computationally intensive and could consume several hours of computer time if the exact sample sizes are very large. Here are some guidelines:

1. Estimate the likely sample size under the Exact method by first determining the asymptotic sample size
2. If the exact sample size is likely to be larger than 1000, computing power is preferable to computing the sample size

36.1 Difference of Proportions

36.1.1 Trial Design

Let π_c and π_t denote the response rates for the control and experimental treatments, respectively. Let $\delta = \pi_t - \pi_c$. The null hypothesis is specified as

$$H_0: \delta = \delta_0$$

and is tested against one-sided alternative hypotheses. If the occurrence of a response denotes patient harm rather than benefit, then $\delta_0 > 0$ and the alternative hypothesis is

$$H_1: \delta < \delta_0$$

or equivalently as

$$H_1: \pi_c > \pi_t - \delta_0 .$$

Conversely, if the occurrence of a response denotes patient benefit rather than harm, then $\delta_0 < 0$ and the alternative hypothesis is

$$H_1: \delta > \delta_0$$

or equivalently as

$$H_1: \pi_c < \pi_t - \delta_0 .$$

36 Binomial Non-Inferiority Two-Sample – Exact

For any given π_c , the sample size is determined by the desired power at a specified value of $\delta = \delta_1$. A common choice is $\delta_1 = 0$ (or equivalently $\pi_t = \pi_c$) but East allows the study to be powered at any value of δ_1 which is consistent with the choice of H_1 .

Let $\hat{\pi}_t$ and $\hat{\pi}_c$ denote the estimates of π_t and π_c based on n_t and n_c observations from the experimental and control treatments, respectively. The test statistic is

$$Z = \frac{\hat{\delta} - \delta_0}{\text{se}(\hat{\delta})} \tag{36.1}$$

where

$$\hat{\delta} = \hat{\pi}_t - \hat{\pi}_c \tag{36.2}$$

and

$$\text{se}(\hat{\delta}) = \sqrt{\frac{\tilde{\pi}_t(1 - \tilde{\pi}_t)}{n_t} + \frac{\tilde{\pi}_c(1 - \tilde{\pi}_c)}{n_c}}. \tag{36.3}$$

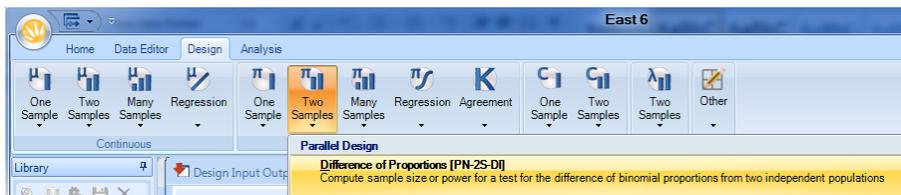
Here $\tilde{\pi}_t$ and $\tilde{\pi}_c$ are the restricted maximum likelihood estimates of π_t and π_c . For more details refer to Appendix V.

36.1.1 Trial Design

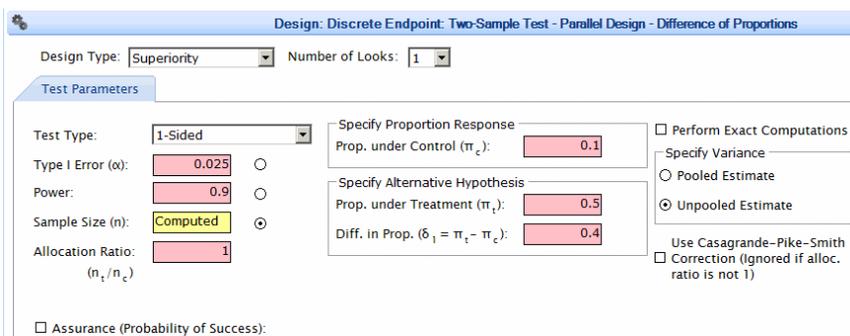
To evaluate the efficacy and safety of drug A vs. drug B in antiretroviral naive HIV-infected individuals, a phase3, 52 week double-blind randomized study is conducted. The primary response measure is the proportion of patients with HIV-RNA levels \leq 50 copies/mL. The study is a non-inferiority designed trial where a standard drug A is expected to have a response rate of 80% and a new experimental drug B is to be compared under a non-inferiority margin of 20% ($\delta_0 = 0.20$). For these studies, inferiority is assumed as the null hypothesis and is to be tested against the alternative of non-inferiority using a one-sided test. Therefore under the null hypothesis H_0 : $\pi_c = 0.8$ and $\pi_t = 0.60$. We will test this hypothesis against H_1 , that both response rates are equal to the null response rate of the control arm, i.e. $\delta_1 = 0$. Thus, under H_1 , we have $\pi_c = \pi_t = 0.8$. East will be used to conduct a one-sided $\alpha = 0.025$ level test with 90% power.

Single Look Design To illustrate this example, in **East** under the **Design** ribbon for **Discrete** data, click **Two Samples** and then choose **Parallel Design: Difference of**

Proportions:



This will launch the following input window:



Change **Design Type: Noninferiority** and keep **Number of Looks: 1**. In the **Design Parameters** dialog box, select the **Perform Exact Computations** checkbox and enter the following parameters:

Test Type: 1 sided (required)
 Type 1 Error (α): 0.025
 Power: 0.9
 Sample Size (n): Computed (select radio button)

Specify Proportion Response
 Prop. under Control (π_c): 0.8

Specify Null Hypothesis
 Prop. under Treatment (π_{t0}): 0.6
 Noninferiority margin (δ_0): -0.2 (will be calculated)

Specify Alternative Hypothesis
 Prop. under Treatment (π_{t1}): 0.8

36 Binomial Non-Inferiority Two-Sample – Exact

Diff. in Prop. ($\delta_1 = \pi_{t1} - \pi_c$): 0

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

Design Type: Noninferiority Number of Looks: 1

Test Parameters

Test Type: 1-Sided

Type I Error (α): 0.025

Power: 0.9

Sample Size (n): Computed

Allocation Ratio: 1 (n_1/n_2)

Specify Proportion Response
 Prop. under Control (π_c): 0.8

Specify Null Hypothesis
 Prop. under Treatment (π_{0t}): 0.6
 Noninferiority Margin (δ_0): -0.2
 ($\delta_0 = \pi_{0t} - \pi_c$)

Specify Alternative Hypothesis
 Prop. under Treatment (π_{1t}): 0.8
 Diff. in Prop. ($\delta_1 = \pi_{1t} - \pi_c$): 0

Perform Exact Computations

Compute

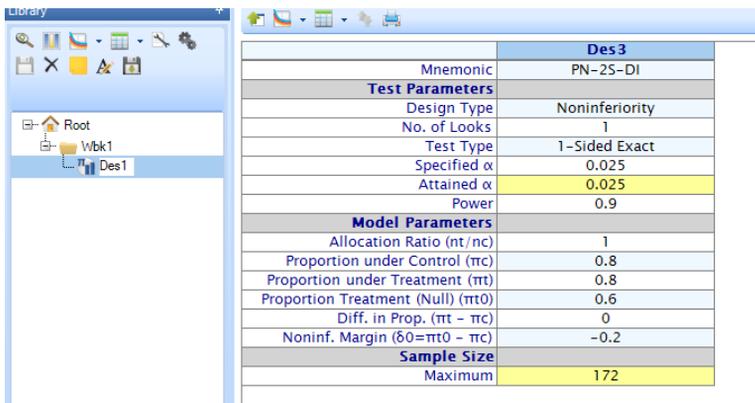
Click **Compute**. The sample size for this design is calculated and the results are shown as a row in the **Output Preview** window:

| Output Preview | | | | | | | | | | | | | |
|----------------|----------------|--------------|---------------|--------------------|-------------------|-------|-------------|-------|---------|------------------------|------------------------|------------|------------|
| ID | Design Type | No. of Looks | Test Type | Specified α | Attained α | Power | Sample Size | nt/nc | π_c | Prop. Treatment (Alt.) | Prop. Treatment (Null) | δ_1 | δ_0 |
| Des9 | Noninferiority | 1 | 1-Sided Exact | 0.025 | 0.025 | 0.9 | 172 | 1 | 0.8 | 0.8 | 0.6 | 0 | -0.2 |

This single look design requires a combined total of 172 patients in order to achieve 90% power.

As is standard in **East**, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the **Output Preview** toolbar. The design details will be displayed in the upper pane, labeled **Output Summary**. This can be saved to the **Library** by selecting **Des 1** and

clicking the  icon.



| Des3 | |
|---|----------------|
| Mnemonic | PN-2S-DI |
| Test Parameters | |
| Design Type | Noninferiority |
| No. of Looks | 1 |
| Test Type | 1-Sided Exact |
| Specified α | 0.025 |
| Attained α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (n_t/n_c) | 1 |
| Proportion under Control (π_c) | 0.8 |
| Proportion under Treatment (π_t) | 0.8 |
| Proportion Treatment (Null) (π_0) | 0.6 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0 |
| Noninf. Margin ($\delta = \pi_0 - \pi_c$) | -0.2 |
| Sample Size | |
| Maximum | 172 |

The design details can be displayed by clicking the  icon.



Design: Discrete Endpoint: Two-Sample Exact Test - Parallel Design - Difference of Proportions

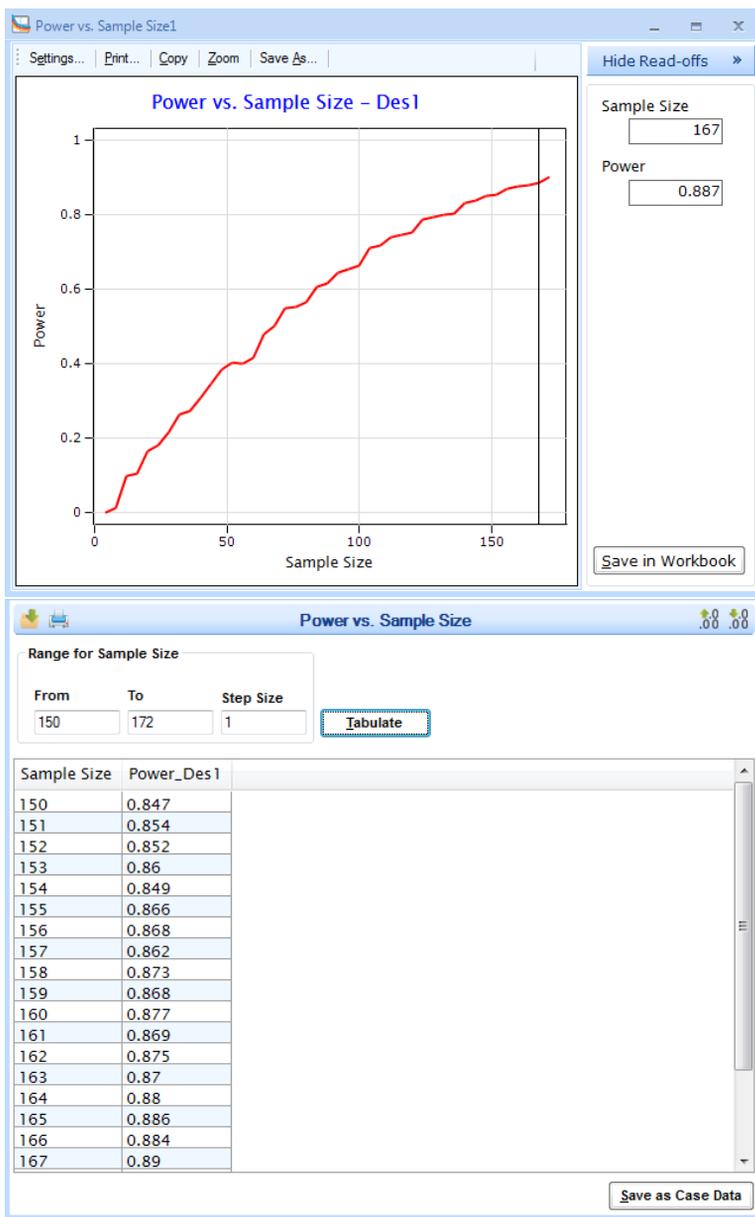
| Test Parameters | |
|------------------------------------|----------------|
| Design ID | Des3 |
| Design Type | Noninferiority |
| Number of Looks | 1 |
| Test Type | 1-Sided Exact |
| Specified α | 0.025 |
| Attained α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Prop. under Control (π_c) | 0.8 |
| Prop. under Treatment (π_t) | 0.8 |
| Prop. Treatment (Null) (π_0) | 0.6 |
| $\delta = \pi_t - \pi_c$ | |
| Under H0 | -0.2 |
| Under H1 | 0 |
| Allocation Ratio (n_t/n_c) | 1 |

| Sample Size Information | |
|-----------------------------|-----|
| Sample Size (n) | 172 |
| Treatment (n _t) | 86 |
| Control (n _c) | 86 |

| Critical Points | |
|---------------------|-------|
| Critical Point | 1.991 |
| Attained at π_c | 0.569 |

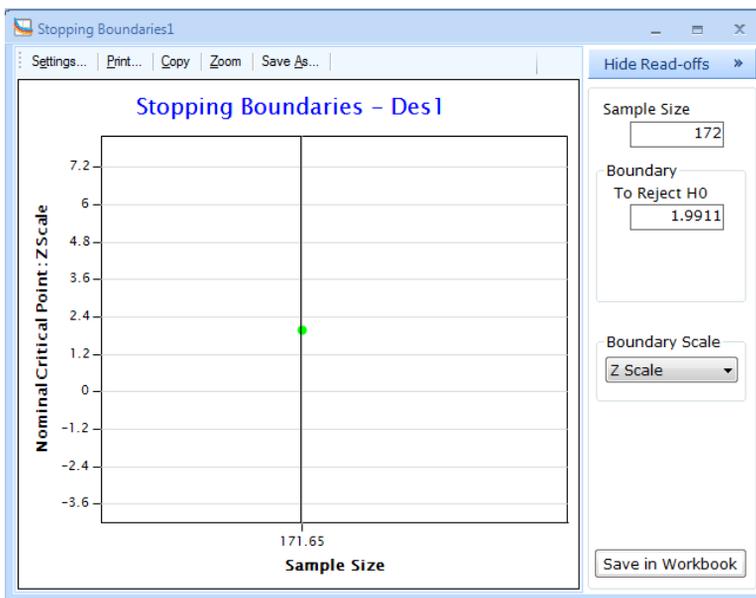
It is important to note that in this exact computation the displayed sample size may not be unique due to the discreteness of the distribution. This can be seen in the **Power Vs Sample Size** graph, which is a useful tool along with its corresponding table, and can be used to find all other sample sizes that guarantee the desired power. In this example, sample sizes ranging from approximately 168-175 result in power close to the required 0.9. These visual tools are available in the **Library** under the **Plots** and **Tables** menus.

36 Binomial Non-Inferiority Two-Sample – Exact



The critical point, or the efficacy boundary set for the rejection of H_0 is 1.991 (on the

Z scale) and (-0.056 on the δ scale). If the magnitude of the observed test statistic exceeds this boundary the null will be rejected in favor of declaring the new treatment to be non-inferior. This can also be seen in the **Stopping Boundaries** chart and table, available in the **Library**.



Stopping Boundaries

Boundary Scales : δ Scale

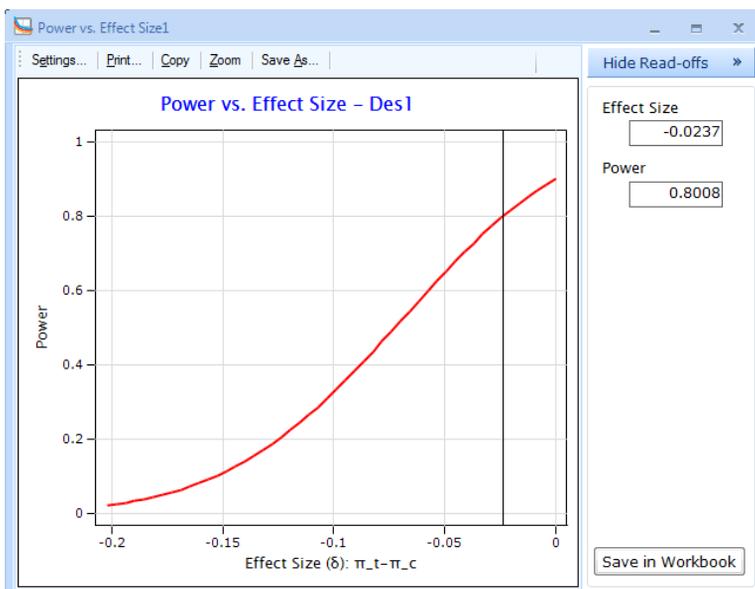
Des1

| Look # | Info. Fraction | Sample Size | Cum. α Spent | Boundaries | |
|--------|----------------|-------------|---------------------|------------|----------|
| | | | | Efficacy | Boundary |
| 1 | 1 | 172 | 0.025 | -0.056 | |

The **Power vs. Treatment Effect** chart dynamically generates power under this design for all values of treatment effect $\delta = \pi_t - \pi_c$. Here it is easy to see how as treatment effect size approaches zero (H_1 : no difference between the two treatments) the power

36 Binomial Non-Inferiority Two-Sample – Exact

of the study reaches the desired 90%. This is available in tabular form as well.



36.2 Ratio of Proportions

Let π_c and π_t denote the response rates for the control and the experimental treatments, respectively. Let the difference between the two arms be captured by the ratio

$$\rho = \frac{\pi_t}{\pi_c}.$$

The null hypothesis is specified as

$$H_0: \rho = \rho_0$$

and is tested against one-sided alternative hypotheses. If the occurrence of a response denotes patient benefit rather than harm, then $\rho_0 < 1$ and the alternative hypothesis is

$$H_1: \rho > \rho_0$$

or equivalently as

$$H_1: \pi_t > \rho_0 \pi_c.$$

Conversely, if the occurrence of a response denotes patient harm rather than benefit, then $\rho_0 > 1$ and the alternative hypothesis is

$$H_1: \rho < \rho_0$$

or equivalently as

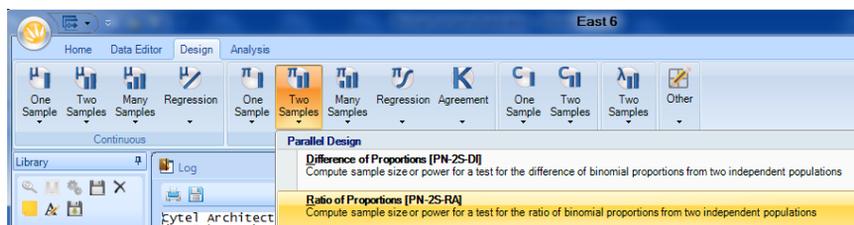
$$H_1: \pi_t < \rho_0 \pi_c .$$

For any given π_c , the sample size is determined by the desired power at a specified value of $\rho = \rho_1$. A common choice is $\rho_1 = 1$ (or equivalently $\pi_t = \pi_c$), but East permits you to power the study at any value of ρ_1 which is consistent with the choice of H_1 .

36.2.1 Trial Design

Suppose with a rare disease condition, the cure rate with an expensive treatment A is estimated to be 90%. The claim of non-inferiority for an inexpensive new treatment B can be held if it can be statistically proven that the ratio $\rho = \pi_t / \pi_c$ is at least 0.833. In other words, B is considered to be non-inferior to A as long as $\pi_t > 0.75$. Thus the null hypothesis $H_0: \rho = 0.833$ is tested against the one-sided alternative hypothesis $H_1: \rho > 0.833$. We want to determine the sample size required to have power of 80% when $\rho = 1$ using a one-sided test with a type-1 error rate of 0.05.

Single Look Design Powered at $\rho = 1$ Consider a one look study with equal sample sizes in the two groups. In **East** under the **Design** ribbon for **Discrete** data, click **Two Samples** and then choose **Parallel Design: Ratio of Proportions**:



36 Binomial Non-Inferiority Two-Sample – Exact

This will launch the following input window:

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Ratio of Proportions

Design Type: Superiority Number of Looks: 1 Include Options

Test Parameters

Test Type: 1-Sided

Type I Error (α): 0.025

Power: 0.9

Sample Size (n): Computed

Allocation Ratio: 1 (n_1/n_2)

Specify Proportion

Prop. under Control (π_c): 0.5

Specify Alternative Hypothesis

Prop. under Treatment (π_t): 0.625

Ratio of Proportions (ρ_1): 1.25 ($\rho_1 = \pi_t / \pi_c$)

Perform Exact Computations

Test Statistic

Pooled Estimate

Unpooled Estimate

Assurance (Probability of Success):

Compute

Change **Design Type: Noninferiority** and keep **Number of Looks: 1**. In the **Design Parameters** dialog box, select the **Perform Exact Computations** checkbox and keep the Test Statistic selected to Wald. Enter the following parameters:

Test Type: 1 sided (required)
 Type 1 Error (α): 0.05
 Power: 0.8
 Sample Size (n): Computed (select radio button)

Specify Proportion
 Prop. under Control (π_c): 0.9

Specify Null Hypothesis
 Prop. under Treatment (π_{t0}): 0.75
 Noninferiority margin (ρ_0): 0.833 (will be calculated)

Specify Alternative Hypothesis
 Prop. under Treatment (π_{t1}): 0.9
 Ratio of Proportions ($\rho_1 = \pi_{t1}/\pi_c$): 1

Design: Discrete Endpoint, Two-Sample Test - Parallel Design - Ratio of Proportions

Design Type: Noninferiority Number of Looks: 1

Test Parameters

Test Type: 1-Sided

Type I Error (α): 0.025

Power: 0.8

Sample Size (n): Computed

Allocation Ratio: 1
 (n_t/n_c)

Specify Proportion
 Prop. under Control (π_c): 0.9

Specify Null Hypothesis
 Prop. under Treatment (π_0): 0.75
 Noninferiority Margin (ρ_0): 0.833
 $\rho_0 = \pi_0 / \pi_c$

Specify Alternative Hypothesis
 Prop. under Treatment (π_1): 0.9
 Ratio of Proportions (ρ_1): 1
 $\rho_1 = \pi_1 / \pi_c$

Perform Exact Computations

Test Statistic

Wald

Score (Farrington Manning)

Compute

Click **Compute**. The sample size for this design is calculated and the results are shown as a row in the **Output Preview** window:

| Output Preview | | | | | | | | | | | | | | |
|----------------|----------------|--------------|---------------|--------------------|-------------------|-------|-------------|-------|-----|------------------------|------------------------|----------|----------|----------------|
| ID | Design Type | No. of Looks | Test Type | Specified α | Attained α | Power | Sample Size | nt/nc | ttc | Prop. Treatment (Alt.) | Prop. Treatment (Null) | ρ_1 | ρ_0 | Test Statistic |
| Des10 | Noninferiority | 1 | 1-Sided Exact | 0.05 | 0.05 | 0.823 | 120 | 1 | 0.9 | 0.9 | 0.75 | 1 | 0.833 | Wald |

The sample size required in order to achieve 80% power is 120 subjects. Note that because of the discreteness involved in performing exact computations, the attained power is 0.823, slightly larger than the specified value of 0.80.

As is standard in **East**, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the **Output Preview** toolbar. The design details will be displayed in the upper pane, labeled **Output Summary**. This can be saved to the **Library** by selecting **Des 1** and

36 Binomial Non-Inferiority Two-Sample – Exact

clicking the  icon.

| Des 10 | |
|--|----------------|
| Mnemonic | PN-2S-RA |
| Test Parameters | |
| Design Type | Noninferiority |
| No. of Looks | 1 |
| Test Type | 1-Sided Exact |
| Specified α | 0.05 |
| Attained α | 0.05 |
| Power | 0.823 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.9 |
| Proportion under Treatment (π_t) | 0.9 |
| Proportion Treatment (Null) (π_0) | 0.75 |
| Ratio of Proportions (π_t / π_c) | 1 |
| Noninf. Margin (π_0 / π_c) | 0.833 |
| Test Statistic | Wald |
| Sample Size | |
| Maximum | 120 |

Design details can be displayed by clicking the  icon.

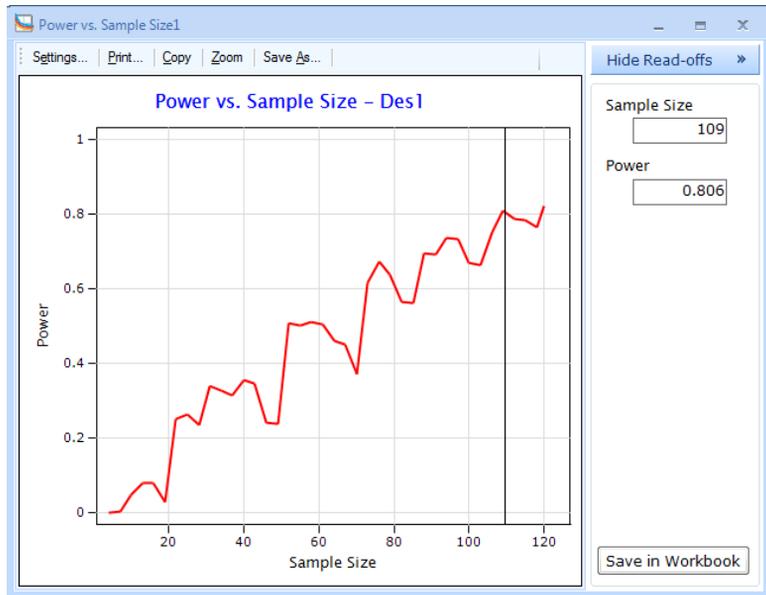
Design: Discrete Endpoint: Two-Sample Exact Test - Parallel Design - Ratio of Proportions

| Test Parameters | | Sample Size Information | |
|------------------------------------|----------------|-------------------------|-------|
| Design ID | Des10 | Sample Size (n) | 120 |
| Design Type | Noninferiority | Treatment (n_t) | 60 |
| Number of Looks | 1 | Control (n_c) | 60 |
| Test Type | 1-Sided Exact | Critical Points | |
| Specified α | 0.05 | Critical Point | 1.961 |
| Attained α | 0.05 | Attained at π_c | 0.999 |
| Power | 0.823 | | |
| Model Parameters | | | |
| Test Statistic | Wald | | |
| Prop. under Control (π_c) | 0.9 | | |
| Prop. under Treatment (π_t) | 0.9 | | |
| Prop. Treatment (Null) (π_0) | 0.75 | | |
| $\rho = \pi_t / \pi_c$ | | | |
| Under H0 | 0.833 | | |
| Under H1 | 1 | | |
| Allocation Ratio (n_t/n_c) | 1 | | |

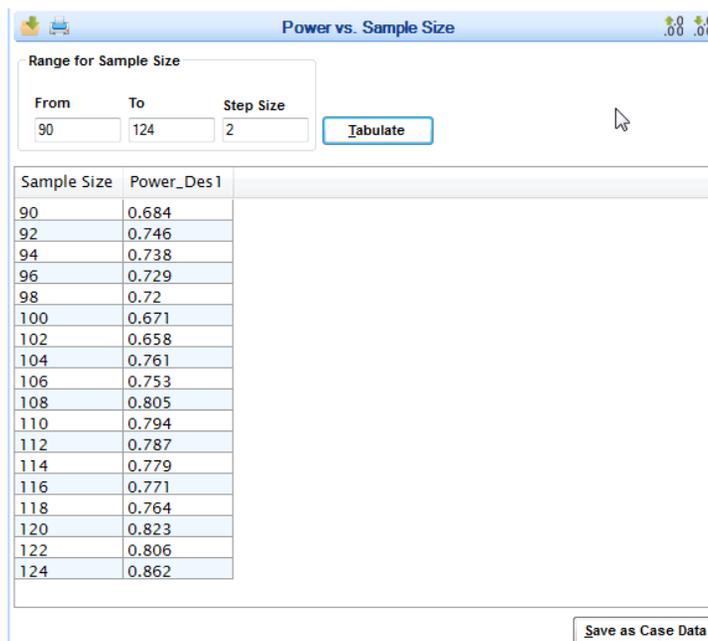
Sample sizes have been rounded.

It is important to note that in this exact computation the displayed sample size may not

be unique due to the discreteness of the distribution. This can be seen in the **Power Vs Sample Size** graph, which is a useful tool along with its corresponding table, and can be used to find all other sample sizes that guarantee the desired power. These visual tools are available in the **Library** under the **Plots** and **Tables** menus.

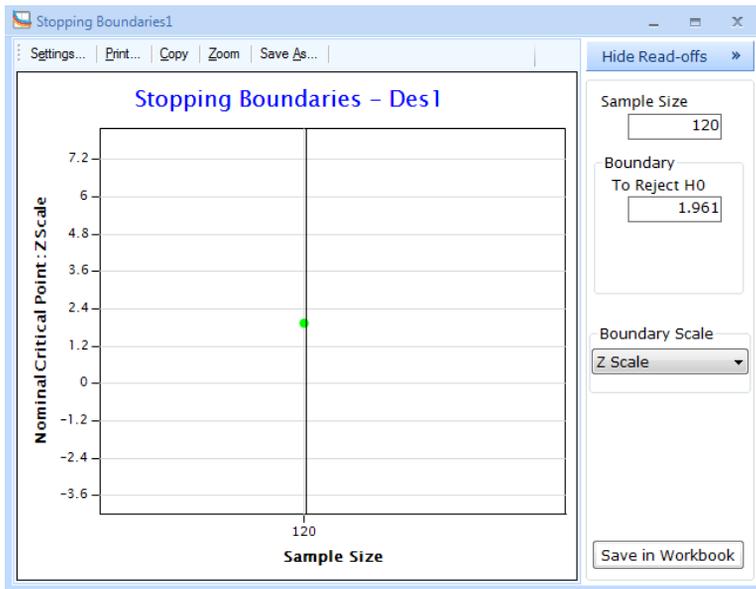


36 Binomial Non-Inferiority Two-Sample – Exact



The critical point, or the boundary set for the rejection of H_0 is 1.961 (on the Z scale), 0.076 (on the $\ln(\rho)$ scale) and 1.079 (on the ρ scale). If the observed test statistic exceeds this boundary the null will be rejected in favor of declaring the new treatment to be non-inferior. This can also be seen in the **Stopping Boundaries** chart and table,

available in the **Library**.



Stopping Boundaries

Boundary Scales :

Des1

| Look # | Info. Fraction | Sample Size | Cum. α Spent | Boundaries |
|--------|----------------|-------------|---------------------|-------------------|
| | | | | Efficacy Boundary |
| 1 | 1 | 120 | 0.05 | 1.079 |

The **Power vs. Treatment Effect** chart dynamically generates power under this design for all values of treatment effect (ratio or ln(ratio) scale). Here it is easy to see how as treatment effect size approaches zero (H_1 : no difference between the two

36 Binomial Non-Inferiority Two-Sample – Exact

treatments) the power of the study reaches the desired 0.8%. This is available in tabular form as well.



37 Binomial Equivalence Two-Sample – Exact

37.1 Equivalence Test

In some experimental situations, it is desired to show that the response rates for the control and the experimental treatments are "close", where "close" is defined prior to the collection of any data. It may be of interest to show that the rate of an adverse event associated with an aggressive therapy is similar to that of the established control. For example, the bleeding rate associated with thrombolytic therapy or cardiac outcomes with a new stent. Let π_c and π_t denote the response rates for the control and the experimental treatments, respectively and let

$$\delta = \pi_t - \pi_c. \quad (37.1)$$

The null hypothesis $H_0: |\pi_t - \pi_c| = \delta_0$ is tested against the two-sided alternative $H_1: |\pi_t - \pi_c| < \delta_0$, where $\delta_0 (> 0)$ defines equivalence. The theory is presented in Section V.4 of Appendix V.

Caution: The methods presented in this chapter are computationally intensive and could consume several hours of computer time if the exact sample sizes are very large.

Here are some guidelines:

1. Estimate the likely sample size under the Exact method by first determining the asymptotic sample size
2. If the exact sample size is likely to be larger than 1000, computing power is preferable to computing the sample size

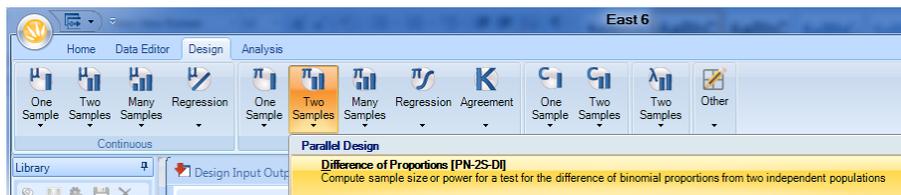
37.1.1 Trial Design

Burgess et al. (2005) describe a randomized controlled equivalence trial, in which the objective is to evaluate the efficacy and safety of a 4% dimeticone lotion for treatment of head lice infestation, relative to a standard treatment. The success rate of the standard treatment is estimated to be about 77.5%. Equivalence is defined as $\delta_0 = 0.20$. The sample size is to be determined with $\alpha = 0.025$ (two-sided) and power, i.e. probability of declaring equivalence, of $1 - \beta = 0.90$.

To illustrate this example, in **East** under the **Design** ribbon for **Discrete** data, click

37 Binomial Equivalence Two-Sample – Exact

Two Samples and then choose **Parallel Design: Difference of Proportions**:



This will launch the following input window:

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

Design Type: Number of Looks:

Test Parameters

Test Type:

Type I Error (α):

Power:

Sample Size (n):

Allocation Ratio: (n_t/n_c)

Assurance (Probability of Success):

Specify Proportion Response

Prop. under Control (π_c):

Specify Alternative Hypothesis

Prop. under Treatment (π_t):

Diff. in Prop. ($\delta_1 = \pi_t - \pi_c$):

Perform Exact Computations

Specify Variance

Pooled Estimate

Unpooled Estimate

Use Casagrande-Pike-Smith Correction (Ignored if alloc. ratio is not 1)

Change **Design Type: Equivalence** and in the **Design Parameters** dialog box, select the **Perform Exact Computations** checkbox. Enter the following parameters:

Test Type: 2 sided (required)
 Type 1 Error (α): 0.025
 Power: 0.9
 Sample Size (n): Computed (select radio button)

Specify Proportion Response
 Prop. under Control (π_c): 0.775
 Prop. under Treatment (π_{t0}): 0.775 (will be calculated)
 Expected Diff. ($\delta_1 = \pi_t - \pi_c$): 0
 Equivalence Margin (δ_0): 0.2

Design: Discrete Endpoint, Two-Sample Test - Parallel Design - Difference of Proportions

Design Type: Equivalence

Test Parameters

Test Type: 2-Sided

Type I Error (α): 0.025

Power: 0.9

Sample Size (n): Computed

Allocation Ratio: 1
 (n_1/n_2)

Specify Proportion Response

Prop. under Control (π_c): 0.775

Prop. under Treatment (π_t): 0.775

Expected Diff. ($\delta_j = \pi_t - \pi_c$): 0

Equivalence Margin (δ_e): 0.2

Perform Exact Computations

Compute

Click **Compute**. The sample size for this design is calculated and the results are shown as a row in the **Output Preview** window:

Output Preview

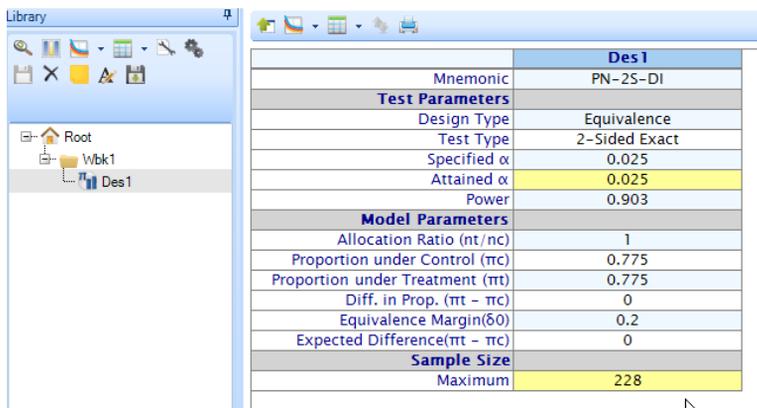
| ID | Design Type | Test Type | Specified α | Attained α | Power | nt/nc | Sample Size | π_c | Prop. Treatment (Alt.) | δ_1 | Equivalence Margin | Expected Difference |
|------|-------------|---------------|--------------------|-------------------|-------|-------|-------------|---------|------------------------|------------|--------------------|---------------------|
| Des1 | Equivalence | 2-Sided Exact | 0.025 | 0.025 | 0.903 | 1 | 228 | 0.775 | 0.775 | 0 | 0.2 | 0 |

This single look design requires a combined total of 228 patients in order to achieve 90% power.

As is standard in **East**, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the **Output Preview** toolbar. The design details will be displayed in the upper pane, labeled **Output Summary**. This can be saved to the **Library** by selecting **Des 1** and

37 Binomial Equivalence Two-Sample – Exact

clicking the  icon.



The screenshot shows a software interface with a 'Library' pane on the left and a parameter table on the right. The library pane shows a tree structure with 'Root', 'Wbk1', and 'Des1'. The parameter table is titled 'Des1' and contains the following data:

| Des1 | |
|--|---------------|
| Mnemonic | PN-2S-DI |
| Test Parameters | |
| Design Type | Equivalence |
| Test Type | 2-Sided Exact |
| Specified α | 0.025 |
| Attained α | 0.025 |
| Power | 0.903 |
| Model Parameters | |
| Allocation Ratio (n_t/n_c) | 1 |
| Proportion under Control (π_c) | 0.775 |
| Proportion under Treatment (π_t) | 0.775 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0 |
| Equivalence Margin(δ) | 0.2 |
| Expected Difference($\pi_t - \pi_c$) | 0 |
| Sample Size | |
| Maximum | 228 |

The design details, which include critical points, or the boundaries set for the rejection of H_0 , can be displayed by clicking the  icon.

Design: Discrete Endpoint: Two-Sample Exact Test - Parallel Design - Difference of Proportions

| Test Parameters | |
|-----------------------------------|---------------|
| Design ID | Des1 |
| Design Type | Equivalence |
| Test Type | 2-Sided Exact |
| Specified α | 0.025 |
| Attained α | 0.025 |
| Power | 0.903 |
| Model Parameters | |
| Prop. under Control (π_c) | 0.775 |
| Prop. under Treatment (π_t) | 0.775 |
| $\delta = \pi_t - \pi_c$ | |
| Under H0 | 0.2 |
| Under H1 | 0 |
| Allocation Ratio (n_t/n_c) | 1 |

 **Sample Size Information**

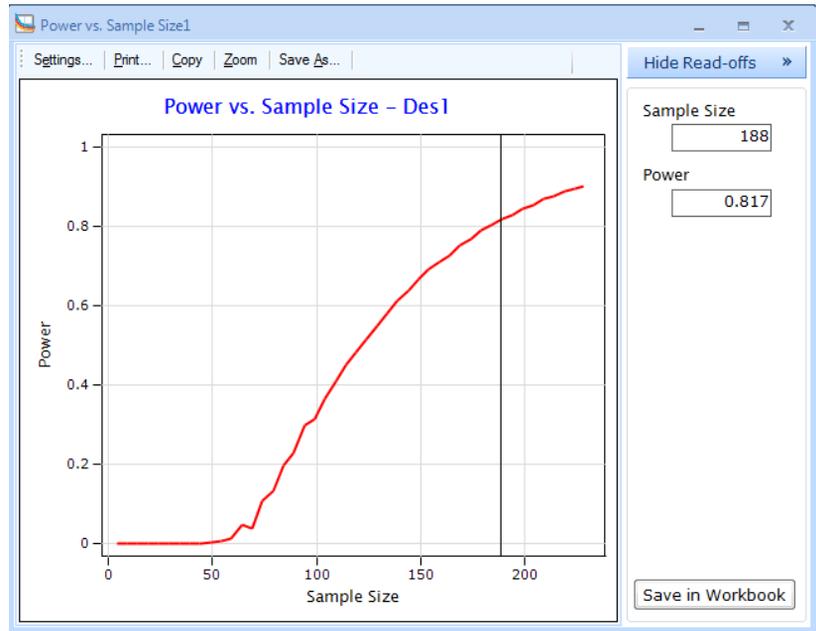
| | |
|-----------------------------|-----|
| Sample Size (n) | 228 |
| Treatment (n _t) | 114 |
| Control (n _c) | 114 |

 **Critical Points**

| | |
|----------------------|-------|
| Lower Critical Point | -1.96 |
| Upper Critical Point | 1.96 |

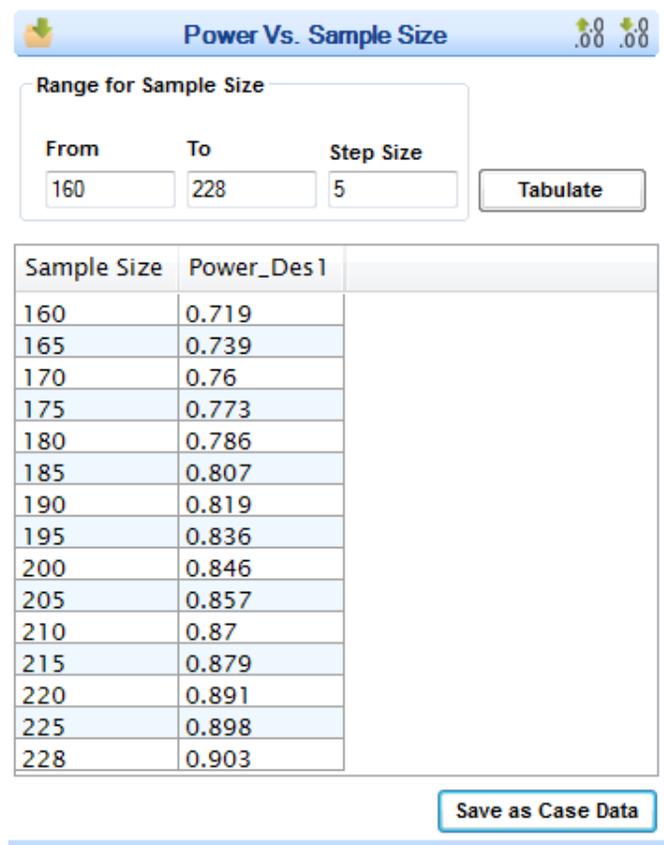
It is important to note that in this exact computation the displayed sample size may not be unique due to the discreteness of the distribution. This can be seen in the **Power Vs Sample Size** graph, which is a useful tool along with its corresponding table, and can be used to find all other sample sizes that guarantee the desired power. These visual

tools are available in the **Library** under the **Plots** and **Tables** menus.



In tabular form:

37 Binomial Equivalence Two-Sample – Exact



Suppose the expected value of the difference in treatment proportions δ_1 is 0.05 or 0.10. A recalculation of the design shows the required sample size will increase to 300

and 606 respectively:

| | Des 11 | Des 12 | Des 13 |
|--|---------------|---------------|---------------|
| Mnemonic | PN-2S-DI | PN-2S-DI | PN-2S-DI |
| Test Parameters | | | |
| Design Type | Equivalence | Equivalence | Equivalence |
| Test Type | 2-Sided Exact | 2-Sided Exact | 2-Sided Exact |
| Specified α | 0.025 | 0.025 | 0.025 |
| Attained α | 0.025 | 0.025 | 0.025 |
| Power | 0.903 | 0.9 | 0.901 |
| Model Parameters | | | |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Proportion under Control (π_c) | 0.775 | 0.775 | 0.775 |
| Proportion under Treatment (π_t) | 0.775 | 0.825 | 0.875 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0 | 0.05 | 0.1 |
| Equivalence Margin(δ_0) | 0.2 | 0.2 | 0.2 |
| Expected Difference($\pi_t - \pi_c$) | 0 | 0.05 | 0.1 |
| Sample Size | | | |
| Maximum | 228 | 300 | 606 |

38 *Binomial Simon's Two-Stage Design*

The purpose of a phase II trial is to determine if a new drug has sufficient efficacy against a specific disease or condition to either warrant further development within Phase II, or to advance onto a Phase III study. In a two-staged design, a fixed number of patients are recruited and treated initially, and if the protocol is considered effective the second stage will continue to enroll additional patients for further study regarding efficacy and safety.

This chapter presents an example for the widely used two-stage optimal and minimax designs developed by Simon (1989). In addition, East supports the framework of an admissible two-stage design, a graphical method geared to search for an alternative with more favorable features (Jung, et al. 2004). The underlying theory is examined in Appendix U.

38.1 *An Example*

During a Phase II study of an experimental drug, a company determined that a response rate of 10% or less is to be considered poor, whereas a response rate is 40% or more is to be considered promising or good. Requirements call for a two-stage study with the following hypotheses:

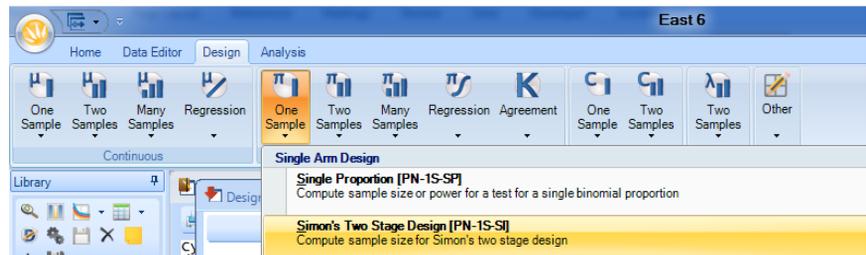
$$H_0 : \pi \leq 0.10$$

$$H_1 : \pi \geq 0.40$$

and design parameters $\alpha = 0.05$ and $1 - \beta = 0.90$.

38.1.1 *Trial Design*

To illustrate this example, in **East** under the **Design** ribbon for **Discrete** data, click **One Sample** and then choose **Single Arm Design: Simon's Two Stage Design**:



This will launch the following input window:

Design: Discrete Endpoint - One-Sample Test - Single Arm Design - Simon's Two Stage Design

Design Type: Minimax

Test Parameters

Test Type: 1-Sided

Type I Error (α): 0.025

Power: 0.9

Upper Limit for Sample Size: 100

Specify Proportion Response

Prop. Response under Null (π_0): 0.1

Prop. Response under Alternative (π_1): 0.4

Compute

Choose **Design Type: Optimal** and enter the following parameters in the **Design Parameters** dialog box:

- Test Type: 1 sided (required)
- Type 1 Error (α): 0.05
- Power: 0.9
- Upper Limit for Sample Size: 100
- Prop. Response under Null (π_0): 0.1
- Prop. Response under Alternative (π_1): 0.4

Design: Discrete Endpoint - One-Sample Test - Single Arm Design - Simon's Two Stage Design

Design Type: Optimal

Test Parameters

Test Type: 1-Sided

Type I Error (α): 0.05

Power: 0.9

Upper Limit for Sample Size: 100

Specify Proportion Response

Prop. Response under Null (π_0): 0.1

Prop. Response under Alternative (π_1): 0.4

Compute

38 Binomial Simon's Two-Stage Design

Click **Compute**. The design is calculated and the results are shown as a row in the **Output Preview** window:

| Output Preview | | | | | | | | | | | | | |
|----------------|-------------|--------------|---------------|--------------------|----------------|-------------------|-------|-------------|------------------|---------|---------|----------------|----------------|
| ID | Design Type | No. of Looks | Test Type | Specified α | SS Upper Limit | Attained α | Power | Sample Size | Expected SS (H0) | π_0 | π_1 | Look1 Futility | Look2 Futility |
| Des 1 | Optimal | 2 | 1-Sided Exact | 0.05 | 100 | 0.035 | 0.902 | 20 | 11.477 | 0.1 | 0.4 | 1 | 4 |

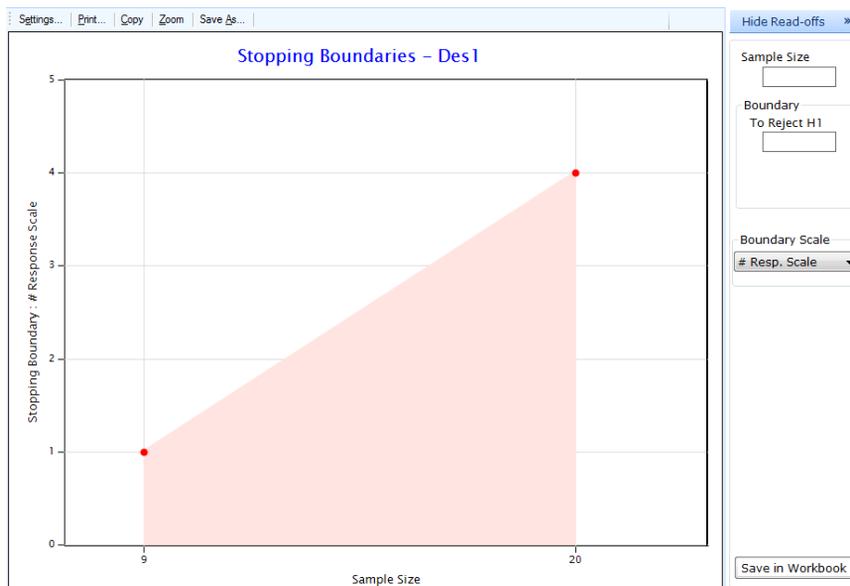
As is standard in **East**, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon. The design details will be displayed in the upper pane, labeled **Output Summary**. Note that because of the discreteness involved in performing exact computations, the attained type-1 error is less than the specified value of 0.05. Similarly, the attained power is slightly larger than the specified value. Save this design to the **Library** by selecting **Des 1** and clicking the  icon.

| Library | |
|---------|--|
| Root | |
| Wbk1 | |
| Des1 | |

| Des 1 | |
|---------------------------------------|---------------|
| Mnemonic | PN-1S-SI |
| Test Parameters | |
| Design Type | Optimal |
| No. of Looks | 2 |
| Test Type | 1-Sided Exact |
| Specified α | 0.05 |
| SS Upper Limit | 100 |
| Attained α | 0.035 |
| Power | 0.902 |
| Model Parameters | |
| Prop. Response under Null (π_0) | 0.1 |
| Prop. Response under Alt. (π_1) | 0.4 |
| Boundary Parameters | |
| Look1 Futility | 1 |
| Look2 Futility | 4 |
| Sample Size | |
| Maximum | 20 |
| Expected Under H0 | 11.477 |

Under the optimal design, the combined maximum sample size for both stages is computed to be 20. The boundary parameter for futility at the first look is 1, and at the second look it is 4. What this means can be further explained using the **Stopping**

Boundaries chart available under the **Plots**  menu.



The scale of the stopping boundaries can be displayed using either number of responses (**# Resp. Scale**) or **Proportion Scale**. The above graph uses the number of responses, which tells us that at the first look, when the cumulative sample size is 9, the trial could be stopped for futility if no more than one patient shows a favorable response to treatment. At the second stage, when all 20 patients are enrolled, the boundary response to reject H_1 is 4 or less. The **Stopping Boundaries** table under the **Tables**  menu also tells us that the probability of crossing the stopping

38 Binomial Simon's Two-Stage Design

boundary, thus warranting early termination, is 0.775.

Stopping Boundaries
0.00 0.00

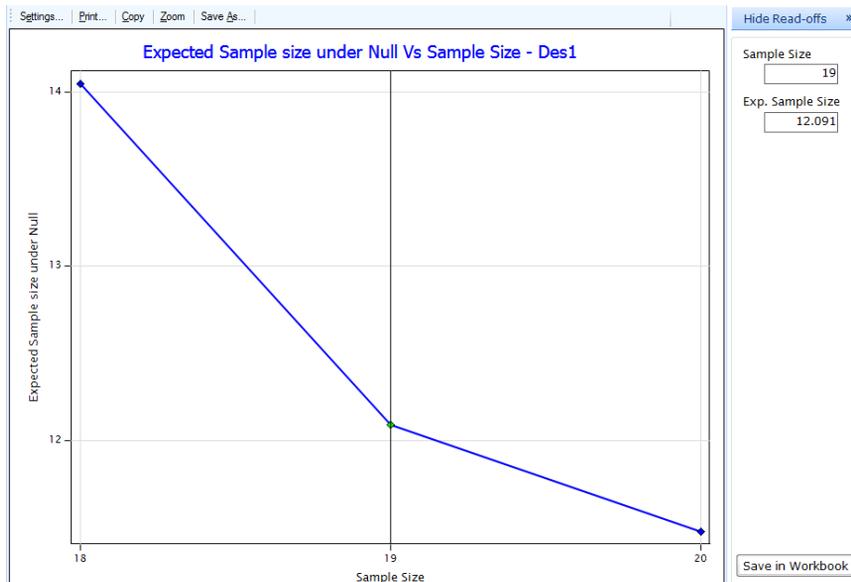
Boundary Scales : # Resp. Scale Settings...

Des1

| Look # | Sample Size | Boundaries | |
|--------|-------------|-------------------|----------|
| | | Futility Boundary | Futility |
| 1 | 9 | 1 | 0.775 |
| 2 | 20 | 4 | NA |

Results can be further analyzed using the **Expected Sample size (under Null) vs.**

Sample Size graph, which is also available in tabular form:



To generate a more sophisticated analysis of the design, select the  icon in the **Library**. In addition to details pertaining to the required optimal design, East also generates results for both minimax as well as admissible designs in regards to sample size, power and probability, and weights used.

Design: Discrete Endpoint: One-Sample Test - Single Arm Design - Simon's Two Stage Design

| Test Parameters | |
|---------------------------------------|---------------|
| Design ID | Des16 |
| Design Type | Optimal |
| Number of Looks | 2 |
| Test Type | 1-Sided Exact |
| Specified α | 0.05 |
| Attained α | 0.035 |
| Power | 0.902 |
| Model Parameters | |
| Prop. Response under Null (π_0) | 0.1 |
| Prop. Response under Alt. (π_1) | 0.4 |

Sample Size Information

| | Total |
|--------------------------------|--------|
| Sample Size (n) | |
| Maximum | 20 |
| Expected H0 (E _{N0}) | 11.477 |

Stopping Boundaries: Look by Look

| Look # | Sample Size (n) | Boundaries | Incr. Boundary Crossing Prob. |
|--------|-----------------|------------------|-------------------------------|
| | | Futility # Resp. | Futility |
| 1 | 9 | 1 | 0.775 |
| 2 | 20 | 4 | |

Output Details



38 Binomial Simon's Two-Stage Design

☰ Sample Size

| Design # | Total Sample Size (n) | Look1 Sample Size (n_1) | Futility Boundary | | Expected Sample Size H0 (EN0) |
|----------|-----------------------|-------------------------|-------------------|-------------|-------------------------------|
| | | | Look1 (r_1) | Look2 (r_2) | |
| Minimax | 18 | 12 | 1 | 4 | 14.046 |
| 2 | 19 | 6 | 0 | 4 | 12.091 |
| Optimal | 20 | 9 | 1 | 4 | 11.477 |

☰ Power and probability

| Design # | Total Sample Size (n) | Attained Alpha | Attained Power | Probability of Early Stopping |
|----------|-----------------------|----------------|----------------|-------------------------------|
| Minimax | 18 | 0.028 | 0.903 | 0.659 |
| 2 | 19 | 0.032 | 0.9 | 0.531 |
| Optimal | 20 | 0.035 | 0.902 | 0.775 |

☰ Weights

| Design # | Total Sample Size (n) | Interval for w | |
|----------|-----------------------|----------------|-------|
| | | Lower | Upper |
| Minimax | 18 | 0.662 | 1 |
| 2 | 19 | 0.381 | 0.662 |
| Optimal | 20 | 0 | 0.381 |

Note: Interval for w is the set of values w such that the design minimizes $w * n + (1 - w) * EN0$

For the optimal design the expected sample size under the null, which assumes the drug performs poorly, is 11.447, which can also be seen in the **Admissible Designs** table, available under the **Tables**  menu:

| Admissible Designs | | | | | |
|---------------------------------|-------------------|-------------------|-------------------|-------|-------------------------------|
| Boundary Scales : # Resp. Scale | | | | | |
| Des1 | | | | | |
| Design # | Total Sample Size | Look1 Sample Size | Futility Boundary | | Expected Sample Size H0 (EN0) |
| | | | Look1 | Look2 | |
| Minimax | 18 | 12 | 1 | 4 | 14.046 |
| 2 | 19 | 6 | 0 | 4 | 12.091 |
| Optimal | 20 | 9 | 1 | 4 | 11.477 |

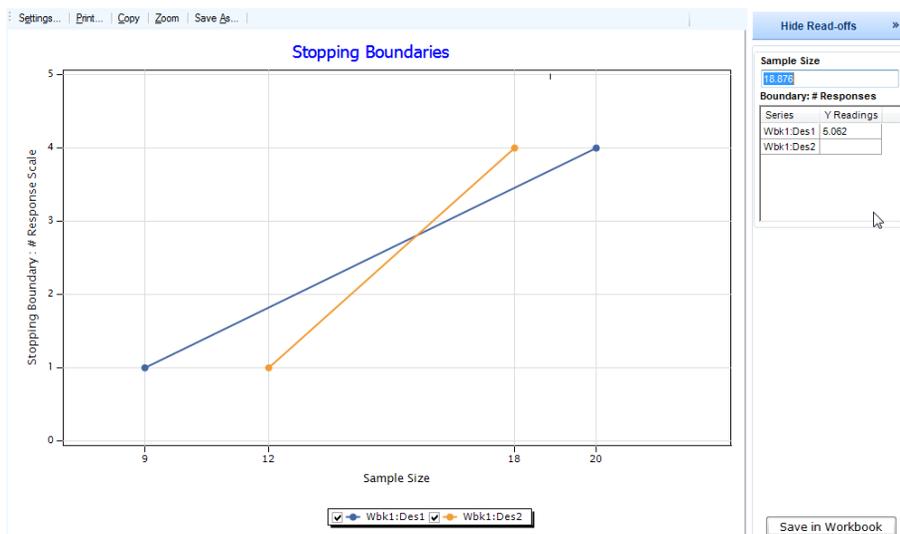
To regenerate the study using a minimax design, select the **Edit Design**  icon. Select **Design Type: Minimax**, leave all design parameters the same and click **Compute**. The cumulative maximum sample size for both stages using this design is 18. As with the optimal design, the first stage boundary response to reject H_1 is 1 or

less and the second stage boundary response to reject H_1 is 4 or less.

| Output Preview | | | | | | | | | | | | | |
|----------------|-------------|--------------|---------------|--------------------|-------------------|-------|-------------|---------|---------|------------------|----------------|----------------|----------------|
| ID | Design Type | No. of Looks | Test Type | Specified α | Attained α | Power | Sample Size | π_0 | π_1 | Expected SS (H0) | SS Upper Limit | Look1 Futility | Look2 Futility |
| Des1 | Optimal | 2 | 1-Sided Exact | 0.05 | 0.035 | 0.902 | 20 | 0.1 | 0.4 | 11.477 | 100 | 1 | 4 |
| Des2 | Minimax | 2 | 1-Sided Exact | 0.05 | 0.028 | 0.903 | 18 | 0.1 | 0.4 | 14.046 | 100 | 1 | 4 |

Save this design to the **Library** by selecting **Des 2** and clicking the  icon. Design details, graphs and tables can be attained as with the optimal design described above.

East provides the capability to visually compare stopping boundaries for both methods simultaneously using a compare plots graph. From the **Library** select both designs, click the  icon, and select **Stopping Boundaries**.



38 Binomial Simon's Two-Stage Design

These stopping boundaries can be compared in tabular format as well:

Boundary Scales : # Resp. Scale ▾

Wbk1:Des1

| Look # | Sample Size | Boundaries | Incremental Boundary Crossing Probabilities |
|--------|-------------|-------------------|---|
| | | Futility Boundary | Futility |
| | | 1 | 9 |
| 2 | 20 | 4 | NA |

Wbk1:Des2

| Look # | Sample Size | Boundaries | Incremental Boundary Crossing Probabilities |
|--------|-------------|-------------------|---|
| | | Futility Boundary | Futility |
| | | 1 | 12 |
| 2 | 18 | 4 | NA |

Although the two futility boundaries are the same for both designs, the cumulative sample size at both stages differ. We also see that the probability of early stopping for futility is higher under the optimal design (0.775) than with the minimax design (0.659). However the cumulative sample size at stage one for the optimal design is only 9 whereas the minimax design requires 12 subjects for the first stage. Referring to the design details generated for the optimal design above, we see that an admissible design (labeled **Design # 2**) requires a total sample size of 19. Here, the cumulative number of subjects required at the end of stage one is only 6 and offers a probability of early stopping of 0.531, less than both the optimal and minimax designs. It is also worthy to note that for the admissible design, the boundary parameter for futility at the first look is 0. This means that only one patient has to show a promising result for the study to proceed to a second stage, whereas at least two successes are required for both

the optimal and minimax designs to warrant a second stage.

Volume 5 Poisson and Negative Binomial Endpoints

39 Introduction to Volume 4 **785**

40 Count Data One-Sample **790**

41 Count Data Two-Samples **799**

39 *Introduction to Volume 4*

This volume describes various cases of clinical trials involving count data. This is often useful in medical research due to its nature of modeling events counted in terms of whole numbers, particularly events that may be considered rare. Typically, interest lies in the rate of occurrence of a particular event during a specific time interval or other unit of space.

Chapter 40 describes the design of tests involving count or Poisson response rates in which an observed response rate is compared to a fixed response rate, possibly derived from historical data.

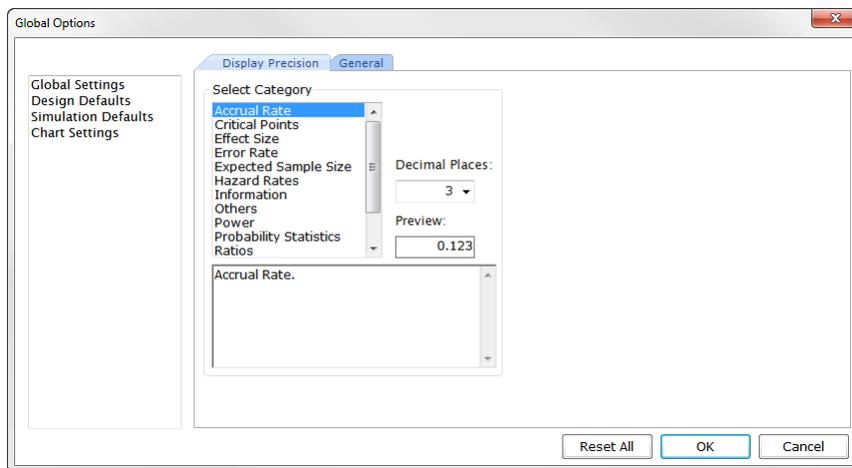
Chapter 41 deals with the comparison of independent samples from two populations in terms of the rate of occurrence of a particular outcome. It supports the design of clinical trials in which this comparison is based on the ratio of rates, assuming a Poisson or Negative Binomial distribution.

39 Introduction to Volume 4

39.1 Settings



Click the icon in the **Home** menu to adjust default values in East 6.

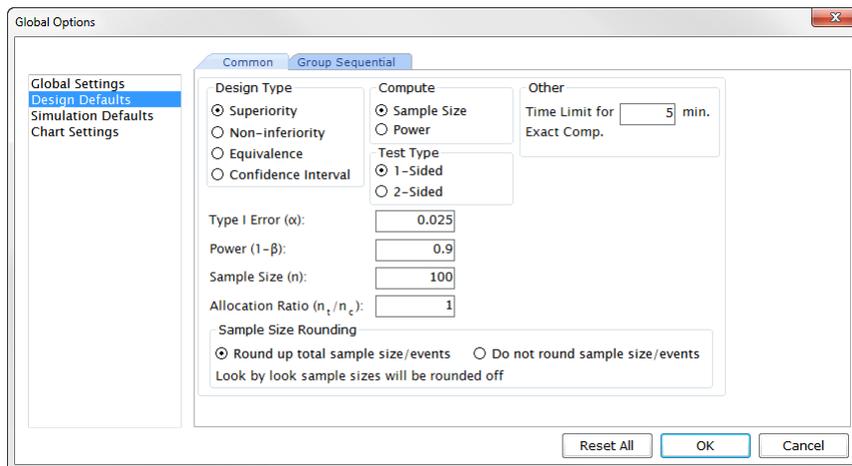


The options provided in the **Display Precision** tab are used to set the decimal places of numerical quantities. The settings indicated here will be applicable to all tests in East 6 under the **Design** and **Analysis** menus.

All these numerical quantities are grouped in different categories depending upon their usage. For example, all the average and expected sample sizes computed at simulation or design stage are grouped together under the category "Expected Sample Size". So to view any of these quantities with greater or lesser precision, select the corresponding category and change the decimal places to any value between 0 to 9.

The **General** tab has the provision of adjusting the paths for storing workbooks, files, and temporary files. These paths will remain throughout the current and future sessions even after East is closed. This is the place where we need to specify the installation directory of the R software in order to use the feature of R Integration in East 6.

The **Design Defaults** is where the user can change the settings for trial design:



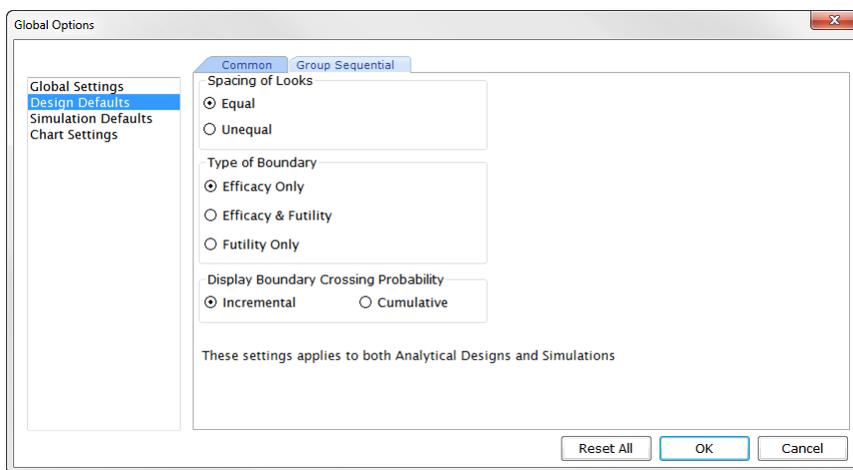
Under the **Common** tab, default values can be set for input design parameters.

You can set up the default choices for the design type, computation type, test type and the default values for type-I error, power, sample size and allocation ratio. When a new design is invoked, the input window will show these default choices.

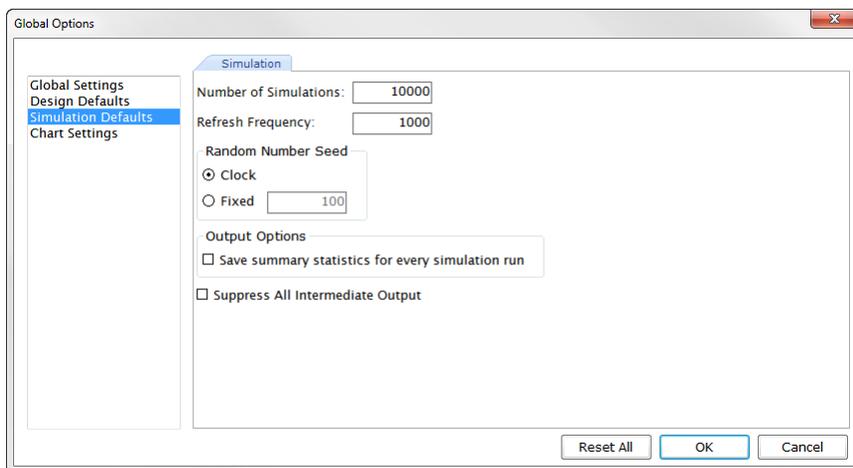
- **Time Limit for Exact Computation**
 This time limit is applicable only to exact designs and charts. Exact methods are computationally intensive and can easily consume several hours of computation time if the likely sample sizes are very large. You can set the maximum time available for any exact test in terms of minutes. If the time limit is reached, the test is terminated and no exact results are provided. Minimum and default value is 5 minutes.
- **Type I Error for MCP**
 If user has selected 2-sided test as default in global settings, then any MCP will use half of the alpha from settings as default since MCP is always a 1-sided test.
- **Sample Size Rounding**
 Notice that by default, East displays the integer sample size (events) by rounding up the actual number computed by the East algorithm. In this case, the look-by-look sample size is rounded off to the nearest integer. One can also see the original floating point sample size by selecting the option "Do not round sample size/events".

39 Introduction to Volume 4

Under the **Group Sequential** tab, defaults are set for boundary information. When a new design is invoked, input fields will contain these specified defaults. We can also set the option to view the Boundary Crossing Probabilities in the detailed output. It can be either Incremental or Cumulative.



Simulation Defaults is where we can change the settings for simulation:

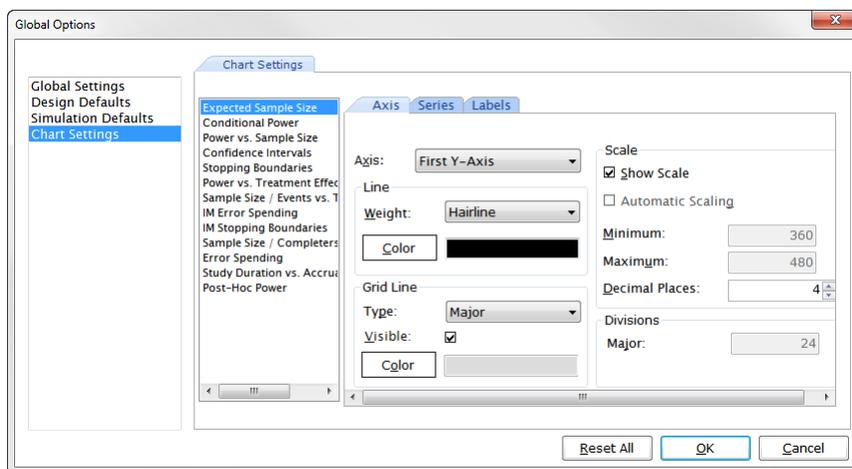


If the checkbox for "Save summary statistics for every simulation" is checked, then East simulations will by default save the per simulation summary data for all the

simulations in the form of a case data.

If the checkbox for "Suppress All Intermediate Output" is checked, the intermediate simulation output window will be always suppressed and you will be directed to the **Output Preview** area.

The **Chart Settings** allows defaults to be set for the following quantities on East6 charts:



We suggest that you do not alter the defaults until you are quite familiar with the software.

40

Count Data One-Sample

This chapter deals with the design of tests involving count or Poisson response rates. Here, independent outcomes or events under examination can be counted in terms of whole numbers, and typically are considered rare. In other words, a basic assumption of the Poisson distribution is that the probability of an event occurring is proportional to the length of time under consideration. The longer the time interval, the more likely the event will occur. Therefore, in this context interest lies in the rate of occurrence of a particular event during a specified period. Section 40.1 focuses on designs in which an observed Poisson response rate is compared to a fixed response rate, possibly derived from historical data.

40.1 Single Poisson Rate

Data following a Poisson distribution are non-negative integers, and the probability that an outcome occurs exactly k times can be calculated as:

$$P(k) = \frac{e^{-\lambda} \lambda^k}{k!}, k = 0, 1, 2, \dots \text{ where } \lambda \text{ is the average rate of occurrence.}$$

When comparing a new protocol or treatment to a well-established control, a preliminary single-sample study may result in valuable information prior to a full-scale investigation. In experimental situations it may be of interest to determine whether the response rate λ differs from a fixed value λ_0 . Specifically we wish to test the null hypothesis $H_0: \lambda = \lambda_0$ against the two sided alternative hypothesis $H_1: \lambda \neq \lambda_0$ or against one sided alternatives of the form $H_1: \lambda > \lambda_0$ or $H_1: \lambda < \lambda_0$. The sample size, or power, is determined for a specified value of λ which is consistent with the alternative hypothesis, denoted λ_1 .

40.1.1 Trial Design

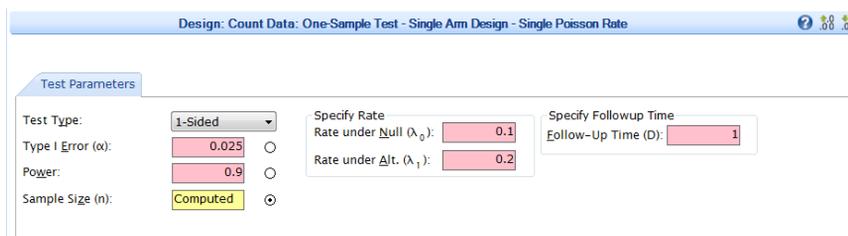
Consider the design of a single-arm clinical trial in which we wish to determine if the positive response rate of a new acute pain therapy is at least 30% per single treatment cycle. Thus, it is desired to test the null hypothesis $H_0: \lambda = 0.2$ against the one-sided alternative hypothesis $H_1: \lambda \geq 0.3$. The trial will be designed such that a one sided $\alpha = 0.05$ test achieves 80% power at $\lambda = \lambda_1 = 0.3$.

In the **Design** tab under the **Count** group choose **One Sample** and then **Single Poisson**

Rate.



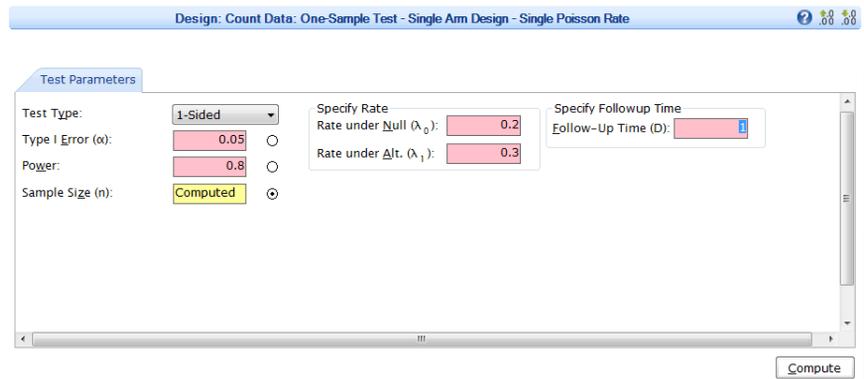
This will launch the following input window:



Enter the following design parameters:

- Test Type: 1 sided
- Type 1 Error (α): 0.05
- Power: 0.8
- Sample Size (n): Computed (select radio button)
- Rate under Null (λ_0): 0.2
- Rate under Alt. (λ_1): 0.3
- Follow-up Time (D): 1

40 Count Data One-Sample



Click **Compute**. The design is shown as a row in the **Output Preview** window:

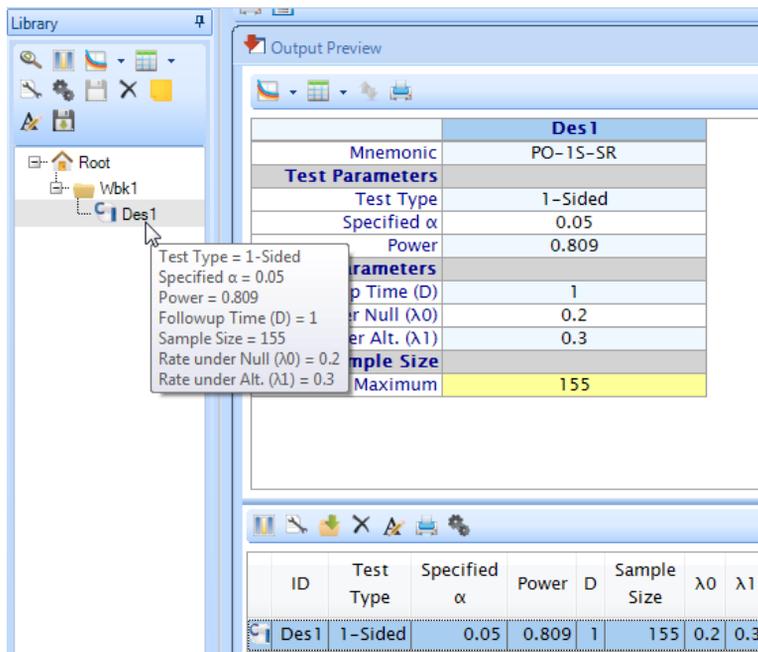
| ID | Test Type | Specified α | Power | D | Sample Size | λ_0 | λ_1 |
|-------|-----------|--------------------|-------|---|-------------|-------------|-------------|
| Des 1 | 1-Sided | 0.05 | 0.809 | 1 | 155 | 0.2 | 0.3 |

The sample size required in order to achieve the desired 80% power is 155 subjects. As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview toolbar. The design details are displayed labeled **Output Summary**.

| Des1 | |
|---------------------------------|----------|
| Mnemonic: | PO-1S-SR |
| Test Parameters | |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.809 |
| Model Parameters | |
| Follow-Up Time (D) | 1 |
| Rate under Null (λ_0) | 0.2 |
| Rate under Alt. (λ_1) | 0.3 |
| Sample Size | |
| Maximum | 155 |

In the **Output Preview** toolbar, click  icon to save this design **Des1** to workbook **Wbk1** in the **Library**. An alternative method to view design details is to hover the cursor over the node **Des1** in the **Library**. A tooltip will appear that summarizes the

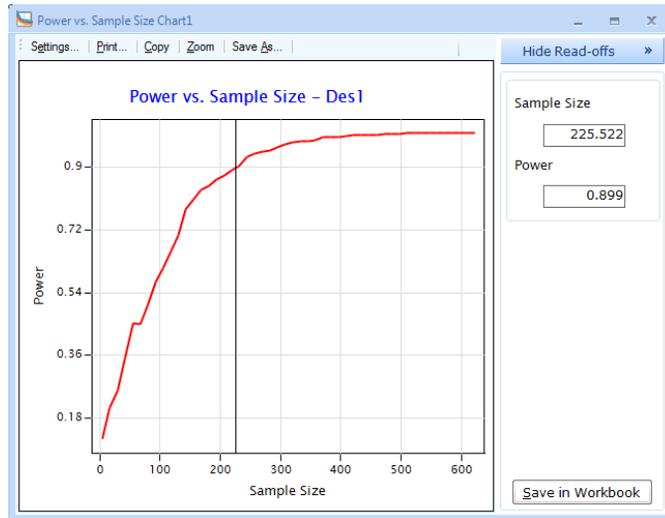
input parameters of the design.



Click  icon on the Library toolbar, and then click **Power vs. Sample Size**. The power curve for this design will be displayed. You can save this chart to the Library by clicking **Save in Workbook**. Alternatively, you can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...** or **Export** into a

40 Count Data One-Sample

PowerPoint presentation.



Close the Power vs. Sample Size chart. To view a summary of all characteristics of this design, select **Des1** in the **Library**, and click  icon.

Design: Count Data: One-Sample Test - Single Arm Design - Single Poisson Rate

| Test Parameters | |
|--------------------|----------------|
| Design ID | Des1 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.809 |
| Sample Size (n) | To be Computed |

| Model Parameters | |
|---------------------------------|-----|
| Rate Under Null (λ_0) | 0.2 |
| Rate Under Alt. (λ_1) | 0.3 |
| Follow-Up Time (D) | 1 |

Sample Size Information
Sample Size (n) 155

Critical Points
Critical Point 1.645

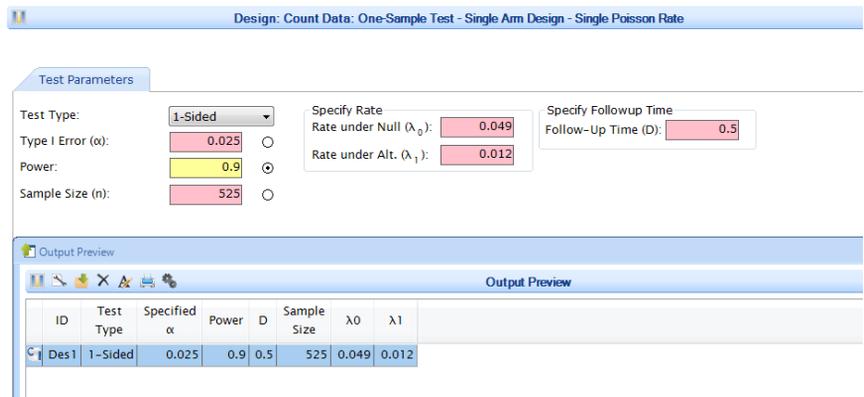
Summary
A total sample size of 155 is required in a study to achieve 0.8 power at 0.05 level of significance when average response rates under null and alternative are 0.2 and 0.3 respectively. Here, the subjects are followed up to 1 units of time.

In addition to the **Power vs. Sample** size chart and table, East also provides the efficacy boundary in the **Stopping Boundaries** chart and table.

Alternatively, East allows the computation of either the **Type-1 error** (α) or **Power** for a given sample size. Using the **Design Input/Output** window as described above, simply enter the desired sample size and click **Compute** to calculate the resulting power of the test.

Power vs Sample Size: Sawtooth paradigm Consider the following design which uses East to compute power assuming a one sample, single Poisson rate.

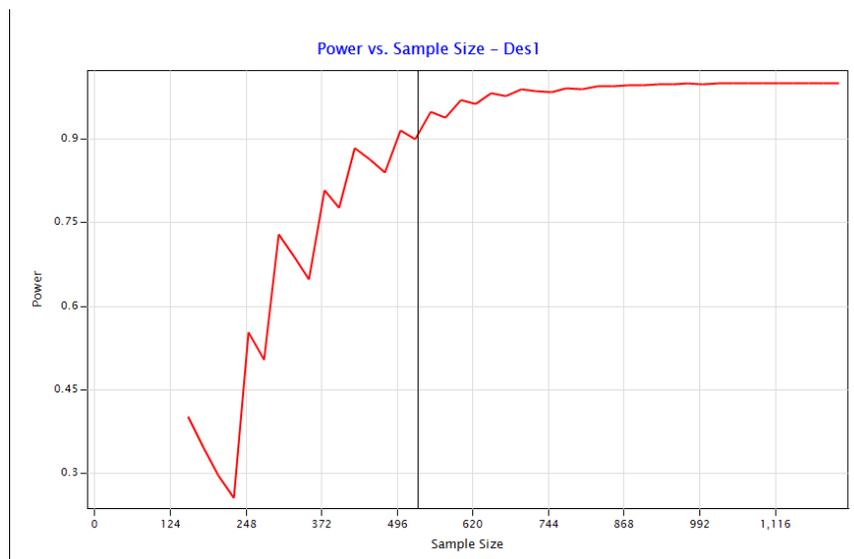
Test Type: 1 sided
 Type 1 Error (α): 0.025
 Power: Computed
 Sample Size (n): 525
 Rate under Null (λ_0): 0.049
 Rate under Alt. (λ_1): 0.012
 Follow-up Time (D): 0.5



Save the design to a workbook, and then generate the **Power vs. Sample Size** graph to obtain the power chart. The resulting curve is commonly described in the literature as a

40 Count Data One-Sample

sawtooth chart.



This chart illustrates that it is possible to have a design where different sample sizes could obtain the same power. As with the binomial distribution, the Poisson distribution is discrete. For power and sample size computations for discrete data, the so called "Saw tooth" phenomena occurs.

The data can also be displayed in a chart form by selecting the  icon in the

Library, and can be printed or saved as case data.

| Range for Sample Size | | |
|-----------------------|-----|-----------|
| From | To | Step Size |
| 525 | 600 | 5 |

Tabulate

| Sample Size | Power_Des 1 |
|-------------|-------------|
| 525 | 0.9 |
| 530 | 0.897 |
| 535 | 0.955 |
| 540 | 0.953 |
| 545 | 0.951 |
| 550 | 0.949 |
| 555 | 0.947 |
| 560 | 0.945 |
| 565 | 0.943 |
| 570 | 0.941 |
| 575 | 0.938 |
| 580 | 0.936 |
| 585 | 0.934 |
| 590 | 0.972 |
| 595 | 0.97 |
| 600 | 0.969 |

It is important to note that for designs with the same power, the attained significance level may vary. For example, the sample sizes of 565 and 580 seem to have a similar power of about 0.94. Upon computing two new designs based on the above design with sample sizes of 565 and 580 respectively, it is apparent that the attained significance levels are different. The design with a lower sample size of 565 pays a higher penalty in terms of type-1 error ($\alpha = 0.03$) than the plan with a larger sample

40 Count Data One-Sample

size of 580 ($\alpha = 0.016$).

| | ID | Test Type | Specified α | Power | D | Sample Size | λ_0 | λ_1 | Attained α |
|---|------|-----------|--------------------|-------|-----|-------------|-------------|-------------|-------------------|
|  | Des1 | 1-Sided | 0.025 | 0.9 | 0.5 | 525 | 0.049 | 0.012 | |
|  | Des2 | 1-Sided | NA | 0.943 | 0.5 | 565 | 0.049 | 0.012 | 0.03 |
|  | Des3 | 1-Sided | NA | 0.936 | 0.5 | 580 | 0.049 | 0.012 | 0.016 |

41

Count Data Two-Samples

Often in experiments based on count data, the aim is to compare independent samples from two populations in terms of the rate of occurrence of a particular outcome. In medical research, outcomes such as the number of times a patient responds to a therapy, develops a certain side effect, or requires specialized care, are of interest. Or perhaps a therapy is being evaluated to determine the number of times it must be applied until an acceptable response rate is observed. East supports the design of clinical trials in which this comparison is based on the ratio of rates, assuming a Poisson or Negative Binomial distribution. These two cases are presented in Sections 41.1 and 41.2, respectively.

41.1 Poisson - Ratio of Rates

41.1.1 Trial Design

41.1.2 Example - Coronary Heart Disease

Let λ_c and λ_t denote the Poisson rates for the control and treatment arms, respectively, and let $\rho_1 = \lambda_t/\lambda_c$. We want to test the null hypothesis that $\rho_1 = 1$ against one or two-sided alternatives. The sample size, or power, is determined to be consistent with the alternative hypothesis, that is $H_1 : \lambda_t \neq \lambda_c$, $H_1 : \lambda_t > \lambda_c$, or $H_1 : \lambda_t < \lambda_c$.

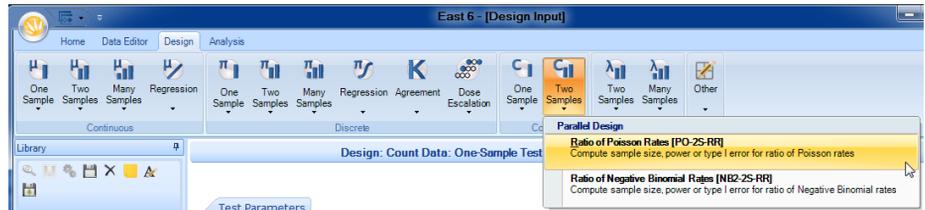
41.1.1 Trial Design

Suppose investigators are preparing design objectives for a prospective randomized trial of a standard treatment (control arm) vs. a new combination of medications (therapy arm) to present at a clinical trials workshop. The endpoint of interest is the number of abnormal ECGs (electrocardiogram) within seven days. The investigators were interested in comparing the therapy arm to the control arm with a two sided test conducted at the 0.025 level of significance. It can be assumed that the rate of abnormal ECGs in the control arm is 30%, thus $\lambda_t = \lambda_c = 0.3$ under H_0 . The investigators wish to determine the sample size to attain power of 80% if there is a 25% decline in the event rate, that is $\lambda_t/\lambda_c = 0.75$. It is important to note that the power of the test depends on λ_c and λ_t , not just the ratio, so different values of the pair (λ_c, λ_t) with the same ratio will yield different solutions.

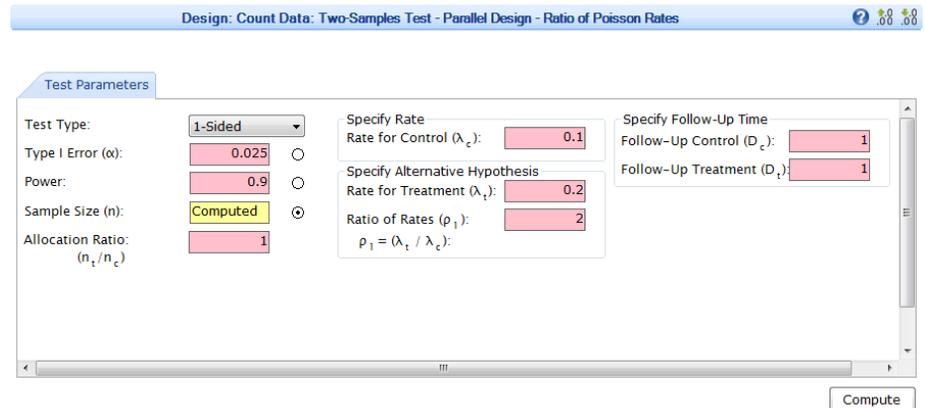
We will now design a study that compares the control arm to the combination therapy arm. In the **Design** tab under the **Count** group choose **Two Samples** and then **Parallel**

41 Count Data Two-Samples

Design - Ratio of Poisson Rates.



This will launch the following input window:



Enter the following design parameters:

- Test Type: 2-sided
- Type 1 Error (α): 0.05
- Power: 0.8
- Sample Size (n): Computed (select radio button)
- Allocation Ratio (n_t/n_c): 1
- Rate for Control (λ_c): 0.3
- Rate for Treatment (λ_t): 0.225 (will be automatically calculated)
- Ratio of Rates $\rho_1 = (\lambda_t/\lambda_c)$: 0.75
- Follow-up Control (D_c): 7
- Follow-up Treatment (D_t): 7

Design: Count Data: Two-Samples Test - Parallel Design - Ratio of Poisson Rates

Test Parameters

Test Type: 2-Sided

Type I Error (α): 0.05

Power: 0.8

Sample Size (n): Computed

Allocation Ratio: 1
(n_t / n_c)

Specify Rate
 Rate for Control (λ_c): 0.3

Specify Alternative Hypothesis
 Rate for Treatment (λ_t): 0.225

Ratio of Rates (ρ_1): 0.75
 $\rho_1 = \lambda_t / \lambda_c$

Specify Follow-Up Time
 Follow-Up Control (D_c): 7

Follow-Up Treatment (D_t): 7

Compute

The Allocation Ratio ($n_t : n_c$) describes the ratio of patients to each arm. For example, an allocation ratio of 3:1 indicates that 75% of the patients are randomized to the treatment arm as opposed to 25% to the control. Here we assume the same number of patients in both arms. Click **Compute**. The design is shown as a row in the **Output Preview** window:

Output Preview

| ID | Test Type | Specified α | Power | Sample Size | λ_c | λ_t | ρ_1 | D_c | D_t | n_t / n_c |
|-------|-----------|--------------------|-------|-------------|-------------|-------------|----------|-------|-------|-------------|
| Des 1 | 2-Sided | 0.05 | 0.8 | 211 | 0.3 | 0.225 | 0.75 | 7 | 7 | 1 |

The sample size required in order to achieve the desired 80% power is 211 subjects. As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the

41 Count Data Two-Samples

Output Preview toolbar. The design details are displayed, labeled **Output Summary**.

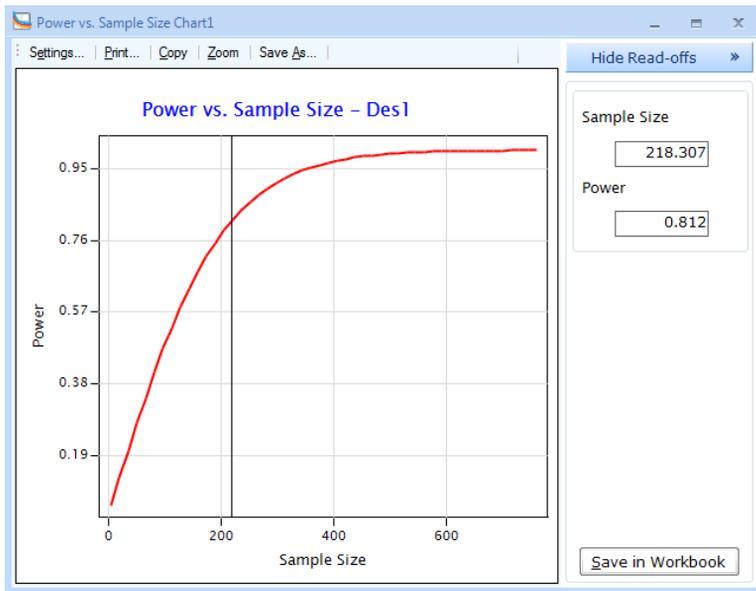
| Des 1 | |
|------------------------------------|----------|
| Mnemonic | PO-25-RR |
| Test Parameters | |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.8 |
| Model Parameters | |
| Rate for Control (λ_c) | 0.3 |
| Rate for Treatment (λ_t) | 0.225 |
| Ratio of Rates (p_1) | 0.75 |
| Follow-Up Control (Dc) | 7 |
| Follow-Up Treatment (Dt) | 7 |
| Allocation Ratio (nt/nc) | 1 |
| Sample Size | |
| Maximum | 211 |

In the **Output Preview** toolbar, click  icon to save this design **Des1** to workbook **Wbk1** in the **Library**. An alternative method to view design details is to hover the cursor over the node **Des1** in the **Library**. A tooltip will appear that summarizes the input parameters of the design.

| Des 1 | |
|------------------------------------|----------|
| Mnemonic | PO-25-RR |
| Test Parameters | |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.8 |
| Model Parameters | |
| Rate for Control (λ_c) | 0.3 |
| Rate for Treatment (λ_t) | 0.225 |
| Ratio of Rates (p_1) | 0.75 |
| Follow-Up Control (Dc) | 7 |
| Follow-Up Treatment (Dt) | 7 |
| Allocation Ratio (nt/nc) | 1 |
| Sample Size | |
| Maximum | 211 |

With the design **Des1** selected in the Library, click  icon on the Library toolbar, and then click **Power vs. Sample Size**. The power curve for this design will be displayed. You can save this chart to the Library by clicking **Save in Workbook**. Alternatively, you can export the chart in one of several image formats (e.g., Bitmap or

JPEG) by clicking **Save As...** or **Export** into a PowerPoint presentation.



Close the Power vs. Sample Size chart. To view all computed characteristics of this

41 Count Data Two-Samples

design, select **Des1** in the **Library**, and click  icon.

Design: Count Data: Two-Samples Test - Parallel Design - Ratio of Poisson Rates

| Test Parameters | |
|--------------------|----------------|
| Design ID | Des1 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.8 |
| Sample Size (n) | To be Computed |

| Model Parameters | |
|--------------------------------------|-------|
| Rate Under Control (λ_c) | 0.3 |
| Rate Under Treatment (λ_t) | 0.225 |
| $\rho_1 = \lambda_t / \lambda_c$ | |
| Under H0 | 1 |
| Under H1 | 0.75 |
| Follow-Up Control (D_c) | 7 |
| Follow-Up Treatment (D_t) | 7 |
| Allocation Ratio (n_t/n_c) | 1 |

Sample Size Information

| | |
|-----------------|-----|
| Sample Size (n) | 211 |
| Treatment (n_t) | 105 |
| Control (n_c) | 106 |

Critical Points

| | |
|----------------------|-------|
| Lower Critical Point | -1.96 |
| Upper Critical Point | 1.96 |

Summary

A total sample size of 211 is required in a study to achieve 0.8 power at 0.05 level of significance when response rates for Control group and Treatment group are 0.3 and 0.225 respectively. Here, the subjects from the two groups are followed up to 7 and 7 units of time.

In addition to the **Power vs. Sample** size chart and table, East also provides the efficacy boundary in the **Stopping Boundaries** chart and table.

Alternatively, East allows the computation of either the **Type-1 error** (α) or **Power** for a given sample size. Using the **Design Input Output** window as described above, simply enter the desired sample size and click **Compute** to calculate the resulting power of the test.

41.1.2 Example - Coronary Heart Disease

The following example is presented in the paper by Gu, et al. (2008) which references a prospective study reported by Stampfer and Willett (1985) examining the relationship between post-menopausal hormone use and coronary heart disease (CHD). Researchers were interested if the group using hormone replacement therapy exhibited less coronary heart disease. The study did show strong evidence that the incidence rate of CHD in the group who did not use hormonal therapy was higher than that in the group who did use post-menopausal hormones. The authors then determined the sample size necessary for the two groups when what they referred to as the ratio of sampling frames is 2, known as the allocation ratio in East. The study assumed an observation time of 2 years, and that the incidence rate of CHD for those using the

hormone therapy is 0.0005. The following excerpt from the paper presents the required sample sizes for the participants using hormone therapy in order to achieve 90% power at $\alpha = 0.05$, for multiple different test procedures:

Table 6 The values of $\lambda_1 = t_1\gamma_1$ and the required sample sizes for the example with $1 - \beta = 0.9$, $\alpha = 0.05$, $R = 1$, $R' = 4$, $\gamma_1 = 0.0005$.

| Test Procedure | λ_1 | Required sample size |
|------------------------------|-------------|----------------------|
| $p_1^{(A)}$ | 8.53 | 8527 |
| $p_2^{(A)}$ | 6.86 | 6860 |
| $p_3^{(A)}$ | 6.66 | 6655 |
| $p_4^{(A)}$ | 6.66 | 6655 |
| $p_5^{(A)}$ | 8.63 | 8627 |
| $p_j^{(P)}, j = 1, \dots, 5$ | 6.59 | 6590 |
| $p^{(C)}$ | 7.26 | 7260 |
| $p^{(M)}$ | 6.58 | 6580 |
| $p^{(L)}$ | 6.37 | 6370 |

The epidemiologist would like to know the required sample size for the two groups when the ratio of sampling frames is 2 ($d = 2$). Suppose also that the observation time domain is 2 years, from the previous study that the incidence rate of CHD for those using the hormone is 0.0005 ($\gamma_1 = 0.0005$), and R' is 4. From formulae (6)–(12), the values of $\lambda_1 = t_1\gamma_1$ and the required sample sizes for the hormone using group to achieve 90% power at $\alpha = 0.05$ for different test procedures are presented in Table 6. The smallest are for $p_3^{(A)}$ and $p_4^{(A)}$.

It is first necessary to determine the difference in notation between the referenced paper and that used by East:

| Gu et al. (2008) | East |
|------------------|----------------------|
| γ_1 | λ_t |
| γ_0 | λ_c |
| $R' = 4$ | $1/\rho_1 = 0.25$ |
| D | Allocation Ratio = 2 |

Once again in the **Design** tab under the **Count** group choose **Two Samples** and then **Parallel Design - Ratio of Poisson Rates**. Enter the following design parameters:

- Test Type: 1-sided
- Type 1 Error (α): 0.05
- Power: 0.9
- Sample Size (n): Computed (select radio button)
- Allocation Ratio (n_t/n_c): 2

41 Count Data Two-Samples

Rate for Control (λ_c): 0.002
 Rate for Treatment (λ_t): 0.0005
 Ratio of Rates $\rho_1 = (\lambda_t/\lambda_c)$: 0.25
 Follow-up Control (D_c): 2
 Follow-up Treatment (D_t): 2

Design: Count Data: Two-Samples Test - Parallel Design - Ratio of Poisson Rates

Test Parameters

| | | | | | | | |
|---------------------------------|----------|--------------------------------|--------------------------------------|--------|--------------------------------|------------------------------|---|
| Test Type: | 1-Sided | Specify Rate | Rate for Control (λ_c): | 0.002 | Specify Follow-Up Time | Follow-Up Control (D_c): | 2 |
| Type I Error (α): | 0.05 | Specify Alternative Hypothesis | Rate for Treatment (λ_t): | 0.0005 | Follow-Up Treatment (D_t): | 2 | |
| Power: | 0.9 | Ratio of Rates (ρ_1): | $\rho_1 = (\lambda_t / \lambda_c)$: | 0.25 | | | |
| Sample Size (n): | Computed | | | | | | |
| Allocation Ratio: (n_t/n_c) | 2 | | | | | | |

Compute

The Allocation Ratio ($n_t : n_c$) describes the ratio of patients to each arm. For example, an allocation ratio of 2:1 indicates that two-thirds of the patients are randomized to the treatment arm as opposed to one-third to the control. **Compute** the design to produce the following output:

Design: Count Data: Two-Samples Test - Parallel Design - Ratio of Poisson Rates

| Test Parameters | |
|--------------------|----------------|
| Design ID | Des2 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Sample Size (n) | To be Computed |

| Model Parameters | |
|--------------------------------------|-------|
| Rate Under Control (λ_c) | 0.002 |
| Rate Under Treatment (λ_t) | 5E-4 |
| $\rho_1 = \lambda_t / \lambda_c$ | |
| Under H0 | 1 |
| Under H1 | 0.25 |
| Follow-Up Control (D_c) | 2 |
| Follow-Up Treatment (D_t) | 2 |
| Allocation Ratio (n_t/n_c) | 2 |

| Sample Size Information | |
|-----------------------------|-------|
| Sample Size (n) | 10027 |
| Treatment (n _t) | 6685 |
| Control (n _c) | 3342 |

| Critical Points | |
|-----------------|--------|
| Critical Point | -1.645 |

Summary
 A total sample size of 10027 is required in a study to achieve 0.9 power at 0.05 level of significance when response rates for Control group and Treatment group are 0.002 and 0 respectively. Here, the subjects from the two groups are followed up to 2 and 2 units of time.

Table 6 in the referenced paper shows the number of subjects required for the treatment group. The East results show that the total number of subjects required for the entire study is 10027. Given that the allocation ratio is 2, the number of subjects required for the control group is $10027/3=3342$ and the treatment group is therefore 6685. This falls in the range of the sample sizes presented in the referenced paper (and close to the minimum of 6655), which again calculates these sizes using a number of different methods.

41.2 Negative Binomial Ratio of Rates

In experiments where the data follows a binomial distribution, the number of successful outcomes for a fixed number of trials is of importance when determining the sample size to adequately power a study. Suppose instead that it is of interest to observe a fixed number of successful outcomes (or failures), but the overall number of trials necessary to achieve this is unknown. In this case, the data is said to follow a Negative Binomial Distribution. There are two underlying parameters of interest. As with the Poisson distribution, λ denotes the average rate of response for a given outcome. In addition, a **shape** parameter γ specifies the desired number of observed "successes". As with the Poisson distribution, the Negative Binomial distribution can be useful when designing a trial where one must wait for a particular event. Here, we are waiting for a specific number of successful outcomes to occur. A Poisson regression analysis assumes a common rate of events for all subjects within a stratum, as well as equal mean and variance (equidispersion). With over dispersed count data, estimates of standard error from these models can be invalid, leading to difficulties in planning a clinical trial. Increased variability resulting from over dispersed data requires a larger sample size in order to maintain power. To address this issue of allowing variability between patients, East provides valid sample size and power calculations for count data using a negative binomial model, resulting in a better evaluation of study design and increased likelihood of trial success.

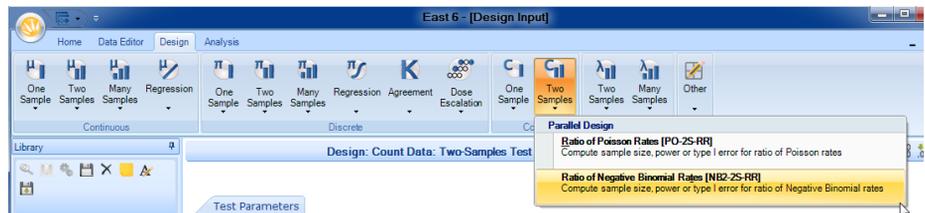
41.2.1 Trial Design

Suppose that a hypothetical manufacturer of robotic prostheses, those that require several components to fully function, has an order to produce a large quantity of artificial limbs. According to historical data, about 20% of the current limbs fail the rigorous quality control test and therefore cannot be shipped to patients. For each order, the manufacturer must produce more than requested; in fact they must continue to produce the limbs until the desired quantity passes quality control. Given that there is a high cost in producing these prosthetic limbs, it is of great interest reduce the number of those that fail the test.

41 Count Data Two-Samples

The company plans to introduce a new feature to the current model, the goal being the probability of failure is reduced to 10%. It is safe to assume that the enhancement will not cause a decline in the original success rate. In this scenario, we wish to test the null hypothesis $H_0: \lambda_c = \lambda_t = 0.2$ against the one sided alternative of the form $H_1: \lambda_c > \lambda_t$. Quality control investigators wish to conduct a one-sided test at the $\alpha = 0.05$ significance level to determine the sample size required obtain 90% power to observe a 50% decline in the event rate, i.e. $\lambda_t/\lambda_c = 0.5$. It is important to note that the power of the test depends on λ_c and λ_t , not just the ratio, so different values of the pair (λ_c, λ_t) with the same ratio will have different solutions. The same holds true for the shape parameter. Different values of (γ_c, γ_t) will result in different sample sizes or power calculations. East allows user specific shape parameters for both the treatment and control groups, however for this example assume that the desired number of successful outcomes for both groups is 10.

The following illustrates the design of a two-arm study comparing the control arm, which the current model of the prosthesis, to the treatment arm, which is the enhanced model. In the **Design** tab under the **Count** group choose **Two Samples** and then **Parallel Design - Ratio of Negative Binomial Rates**.



This will launch the following input window:

The screenshot shows a software window titled "Design: Count Data: Two-Samples Test - Parallel Design - Ratio of Negative Binomial Rates". The "Test Parameters" tab is active. The parameters are as follows:

- Test Type: 1-Sided (dropdown menu)
- Type I Error (α): 0.025 (input field)
- Power: 0.9 (input field)
- Sample Size (n): Computed (input field)
- Allocation Ratio: 1 (input field)
- Specify Rate: Rate for Control (λ_c): 0.2 (input field)
- Specify Alternative Hypothesis: Rate for Treatment (λ_t): 0.1 (input field)
- Ratio of Rates (ρ): 0.5 (input field)
- Specify Follow-Up Time: Follow-Up Time (D): 1 (input field)
- Specify Shape Parameter: Shape Control (γ_c): 1 (input field)
- Shape Treatment (γ_t): 1 (input field)

A "Compute" button is located at the bottom right of the window.

Enter the following design parameters:

- Test Type: 1 sided
- Type 1 Error (α): 0.05
- Power: 0.9
- Sample Size (n): Computed (select radio button)
- Allocation Ratio (n_t/n_c): 1
- Rate for Control (λ_c): 0.2
- Rate for Treatment (λ_t): 0.1
- Ratio of Rates $\rho = (\lambda_t/\lambda_c)$: 0.5
- Follow-up Time (D): 1
- Shape Control (γ_c): 10
- Shape Treatment (γ_t): 10

41 Count Data Two-Samples

Design: Count Data: Two-Samples Test - Parallel Design - Ratio of Negative Binomial Rates

Test Parameters

Test Type: 1-Sided

Type I Error (α): 0.05

Power: 0.9

Sample Size (n): Computed

Allocation Ratio: 1
(n_t / n_c)

Specify Rate
Rate for Control (λ_c): 0.2

Specify Alternative Hypothesis
Rate for Treatment (λ_t): 0.1
Ratio of Rates (ρ_1): 0.5
 $\rho_1 = \lambda_t / \lambda_c$

Specify Follow-Up Time
Follow-Up Time (D): 1

Specify Shape Parameter
Shape Control (γ_c): 10
Shape Treatment (γ_t): 10

Compute

The Allocation Ratio ($n_t : n_c$) describes the ratio of patients to each arm. For example, an allocation ratio of 3:1 indicates that 75% of the patients are randomized to the treatment arm as opposed to 25% to the control. Here we assume the same number of patients in both arms. Click **Compute**. The design is shown as a row in the **Output Preview** window:

| ID | Test Type | Specified α | Power | Sample Size | λ_c | λ_t | ρ_1 | D | γ_c | γ_t | n_t/n_c |
|-------|-----------|--------------------|-------|-------------|-------------|-------------|----------|---|------------|------------|-----------|
| Des 1 | 1-Sided | 0.05 | 0.9 | 1248 | 0.2 | 0.1 | 0.5 | 1 | 10 | 10 | 1 |

The sample size required in order to achieve the desired 90% power is 1248 subjects. As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview toolbar. The design details will be displayed, labeled **Output**

Summary.

| Des 1 | |
|------------------------------------|-----------|
| Mnemonic | NB2-2S-RR |
| Test Parameters | |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Rate for Control (λ_c) | 0.2 |
| Rate for Treatment (λ_t) | 0.1 |
| Ratio of Rates (ρ_1) | 0.5 |
| Follow-Up Time (D) | 1 |
| Shape Control (γ_c) | 10 |
| Shape Treatment (γ_t) | 10 |
| Allocation Ratio (n_t/n_c) | 1 |
| Sample Size | |
| Maximum | 1248 |

In the **Output Preview** toolbar, click icon to save this design **Des1** to workbook **Wbk1** in the **Library**. An alternative method to view design details is to hover the cursor over the node **Des1** in the **Library**. A tooltip will appear that summarizes the input parameters of the design.

Design Input
Design: Count Data: Two-Samples Test - Parallel Design - Ratio of Negative Binomial Rates

Test Parameters

- Test Type: 1-Sided
- Type I Error (α): 0.05
- Power: 0.9
- Sample Size (n): 1248
- Allocation Ratio (n_t/n_c): 1

Specify Rate

- Rate for Control (λ_c): 0.2
- Rate for Treatment (λ_t): 0.1
- Ratio of Rates (ρ_1): 0.5
- $\rho_1 = \lambda_t / \lambda_c$

Specify Shape Parameter

- Shape Control (γ_c): 10
- Shape Treatment (γ_t): 10

Specify Follow-Up Time

- Follow-Up Time (D): 1

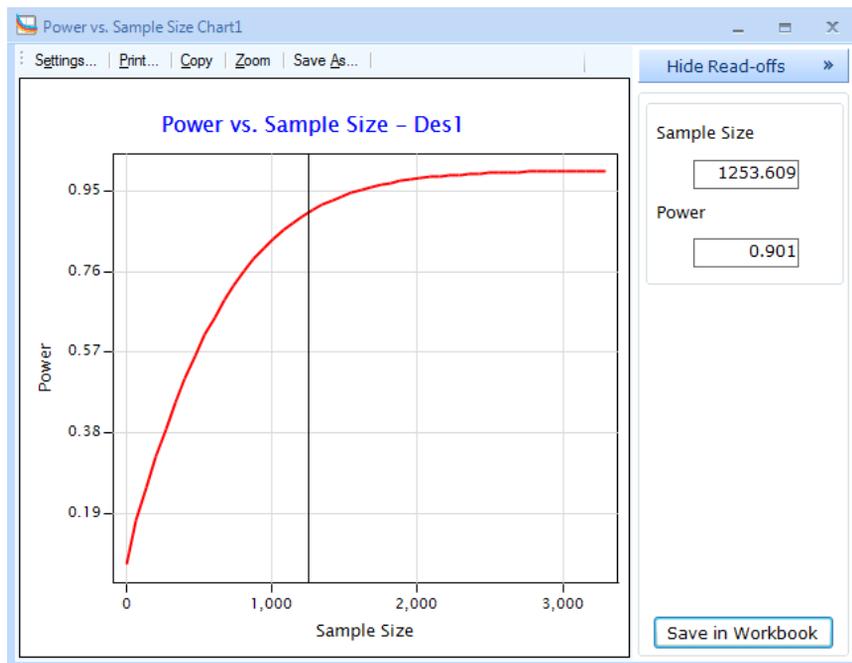
Output Preview

| ID | Test Type | Specified α | Power | Sample Size | λ_c | λ_t | ρ_1 | D | γ_c | γ_t | n_t/n_c |
|------|-----------|--------------------|-------|-------------|-------------|-------------|----------|---|------------|------------|-----------|
| Des1 | 1-Sided | 0.05 | 0.9 | 1248 | 0.2 | 0.1 | 0.5 | 1 | 10 | 10 | 1 |

With the design **Des1** selected in the Library, click icon on the Library toolbar, and then click **Power vs. Sample Size**. The power curve for this design will be displayed. You can save this chart to the Library by clicking **Save in Workbook**. Alternatively, you can export the chart in one of several image formats (e.g., Bitmap or

41 Count Data Two-Samples

JPEG) by clicking **Save As...** or **Export** into a PowerPoint presentation.



Close the Power vs. Sample Size chart. To view all computed characteristics of this

design, select **Des1** in the **Library**, and click  icon.

Design: Count Data: Two-Samples Test - Parallel Design - Ratio of Negative Binomial Rates

| Test Parameters | |
|--------------------|----------------|
| Design ID | Des1 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Sample Size (n) | To be Computed |

| Model Parameters | |
|--------------------------------------|-----|
| Rate Under Control (λ_c) | 0.2 |
| Rate Under Treatment (λ_t) | 0.1 |
| $p_1 = \lambda_t / \lambda_c$ | |
| Under H0 | 1 |
| Under H1 | 0.5 |
| Follow-Up Time (D) | 1 |
| Shape Control (v_c) | 10 |
| Shape Treatment (v_t) | 10 |
| Allocation Ratio (n_t/n_c) | 1 |

Sample Size Information

| | |
|-----------------------------|------|
| Sample Size (n) | 1248 |
| Treatment (n _t) | 624 |
| Control (n _c) | 624 |

Critical Points

| | |
|----------------|--------|
| Critical Point | -1.645 |
|----------------|--------|

Summary
 A total sample size of 1248 is required in a study to achieve 0.9 power at 0.05 level of significance when response rates for Control group and Treatment group are 0.2 and 0.1 respectively and the shape parameters for the two groups are 10 and 10 respectively. Here, the subjects from the two groups are followed up to 1 units of time.

In addition to the **Power vs. Sample** size chart and table, East also provides the efficacy boundary in the **Stopping Boundaries** chart and table.

For a specific desired sample size, East allows the computation of either the **Type-1 error** (α) or **Power** for a test. Using the **Design Input Output** window and methods as described above, simply enter the desired sample size and click **Compute** to calculate the resulting power of the test.

In addition to this example, consider the following illustration of the benefit of using the negative binomial model in clinical trials. In real life settings, the variance of count data observed between patients is typically higher than the observed mean. The negative binomial model accommodates between subject heterogeneity according to a Gamma distribution. For example:

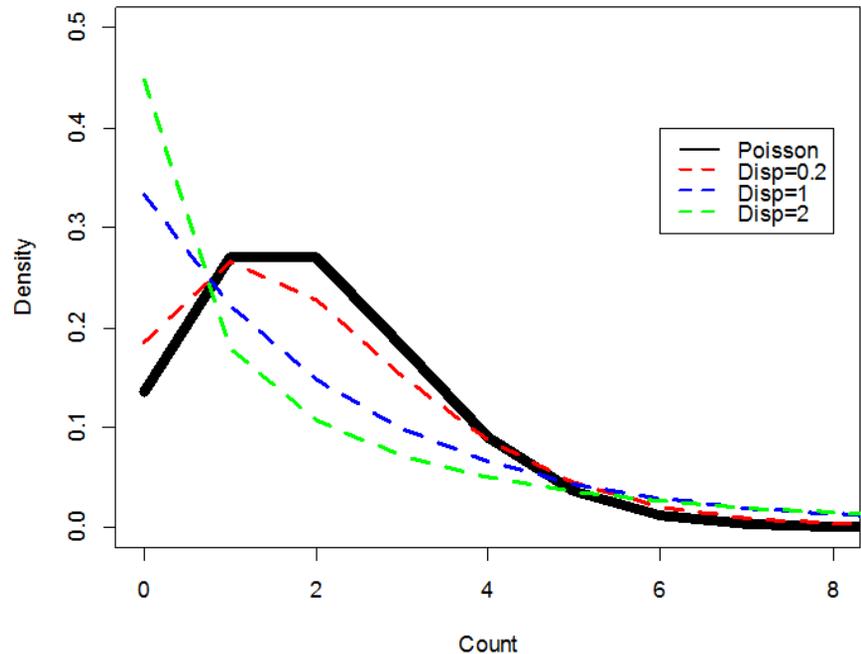
Poisson: $Y \sim Poisson(\lambda)$

Negative Binomial: $Y_i \sim Poisson(\lambda k_i)$ where $k_i \sim Gamma(k)$

In the case of no overdispersion ($k = 0$) the negative binomial model reduces to the

41 Count Data Two-Samples

Poisson model. In the figure below, the Poisson and negative binomial models are displayed under various values of the dispersion parameter.



Assuming the above parameterization, the variance of the negative binomial model is $\lambda + k\lambda^2$. The inflation in variance is thus linear by the factor $1 + k * \lambda$ and dependent on the mean. Depending on the distributional assumption used and its impact on the variance, sample size and power can vary widely.

In multiple sclerosis (MS) patients, magnetic resonance imaging (MRI) is used as a marker of efficacy by means of serial counts of lesions appearing on the brain. Exacerbations rates as a primary endpoint are frequently used in MS as well as in chronic obstructive pulmonary disease (COPD) and asthma (Keene *et al.* 2007). Poisson regression could be considered, however this model would not address variability between patients, resulting in over dispersion. The negative binomial model offers an alternative approach.

TRISTAN (Keene *et al.* 2007) was a double-blind, randomized study for COPD comparing the effects of the salmeterol/fluticasone propionate combination product

(SFC) to salmeterol alone, fluticasone proprionate alone and placebo. Although the primary end-point was pre-bronchodilator FEV1, the number of exacerbations was an important secondary endpoint.

Suppose we are to design a new trial to be observed over a period of 1 to 2 years. The primary objective is the reduction of the rate of exacerbations, defined as a worsening of COPD symptoms that require treatment with antibiotics, cortisone or both, with the combination product SFC versus placebo. Based on the TRISTAN results, we aim to reduce the incidence of events by 33%. Suppose the exacerbation rate is 1.5 per year, and can expect to detect a rate of 1.0 in the combination group. Assume a 2-sided test with a 5% significance level and 90% power. Using a Poisson model, a total of 214 patients are needed to be enrolled in the study.

For the TRISTAN data, the estimate of the overdispersion parameter was 0.46 (95% CI: 0.34-0.60). Using a negative binomial model with overdispersion of 0.33, 0.66, 1

41 Count Data Two-Samples

and 2, the increase in sample size ranged from 298 to 725, respectively.

Design: Count Data: Two-Samples Test - Parallel Design - Ratio of Negative Binomial Rates

Test Parameters

Test Type: 2-Sided

Type I Error (α): 0.05

Power: 0.9

Sample Size (n): Computed

Allocation Ratio: 1
(n_1/n_2)

Specify Rate
Rate for Control (λ_c): 1.5

Specify Alternative Hypothesis
Rate for Treatment (λ_t): 1
Ratio of Rates (ρ_1): 0.667
 $\rho_1 = (\lambda_t / \lambda_c)$

Specify Follow-Up Time
Follow-Up Time (D): 1

Specify Shape Parameter
Shape Control (γ_c): 0.66
Shape Treatment (γ_t): 0.66

Compute

Output Preview

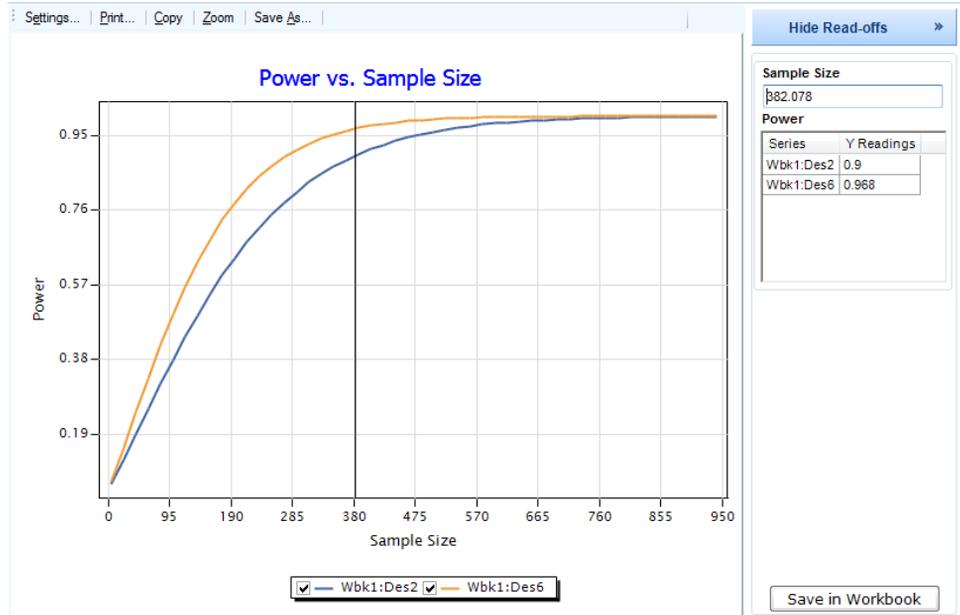
Output Summary

| | Des 1 | Des 2 | Des 3 | Des 4 |
|------------------------------------|-----------|-----------|-----------|-----------|
| Mnemonic | NB2-25-RR | NB2-25-RR | NB2-25-RR | NB2-25-RR |
| Test Parameters | | | | |
| Test Type | 2-Sided | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 | 0.05 |
| Power | 0.9 | 0.9 | 0.9 | 0.9 |
| Model Parameters | | | | |
| Rate for Control (λ_c) | 1.5 | 1.5 | 1.5 | 1.5 |
| Rate for Treatment (λ_t) | 1 | 1 | 1 | 1 |
| Ratio of Rates (ρ_1) | 0.667 | 0.667 | 0.667 | 0.667 |
| Follow-Up Time (D) | 1 | 1 | 1 | 1 |
| Shape Control (γ_c) | 0.33 | 0.66 | 1 | 2 |
| Shape Treatment (γ_t) | 0.33 | 0.66 | 1 | 2 |
| Allocation Ratio (n_t/n_c) | 1 | 1 | 1 | 1 |
| Sample Size | | | | |
| Maximum | 298 | 382 | 469 | 725 |

Exacerbation rates are calculated as number of exacerbations divided by the length of time in treatment in years. EAST can be used to illustrate the impact of a one versus two year study by changing the follow-up duration.

For 382 patients and a shape parameter of 0.66, power is increased from 90% to 97%

when follow-up time is doubled :



The number of patients required for a two year study powered at 90% is 277, whereas 382 patients would be required to achieve the same power for a study period of one

41 Count Data Two-Samples

year.

| Output Summary: Wbk1 | | |
|------------------------------------|-----------|-----------|
| | Wbk1:Des2 | Wbk1:Des8 |
| Mnemonic | NB2-2S-RR | NB2-2S-RR |
| Test Parameters | | |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.9 | 0.902 |
| Model Parameters | | |
| Rate for Control (λ_c) | 1.5 | 1.5 |
| Rate for Treatment (λ_t) | 1 | 1 |
| Ratio of Rates (ρ_1) | 0.667 | 0.667 |
| Follow-Up Time (D) | 1 | 2 |
| Shape Control (γ_c) | 0.66 | 0.66 |
| Shape Treatment (γ_t) | 0.66 | 0.66 |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Sample Size | | |
| Maximum | 382 | 277 |

Negative binomial models are increasing in popularity for medical research, and as the industry standard for trial design, East continues to evolve by incorporating sample size methods for count data. These models allow the count to vary around the mean for groups of patients instead of the population means. Additionally, increased variability does lead to a larger test population; consequently the balance between power, sample size and duration of observation needs to be evaluated.

Volume 6 *Time to Event Endpoints*

- 42 *Introduction to Volume 6* 820
- 43 *Tutorial: Survival Endpoint* 826
- 44 *Superiority Trials with Variable Follow-Up* 865
- 45 *Superiority Trials with Fixed Follow-Up* 908
- 46 *Non-Inferiority Trials Given Accrual Duration and Accrual Rates* 934
- 47 *Non-Inferiority Trials with Fixed Follow-Up* 950
- 48 *Superiority Trials Given Accrual Duration and Study Duration* 966
- 49 *Non Inferiority Trials Given Accrual Duration and Study Duration* 984
- 50 *A Note on Specifying Dropout parameters in Survival Studies* 994
- 51 *Multiple Comparison Procedures for Survival Data* 999

42 *Introduction to Volume 6*

The chapters in this volume deal with clinical trials where the endpoint of interest is the time from entry into the study until a specific event –for example, death, tumour recurrence, or heart attack – occurs. Such trials are also referred to as survival trials, time-to-event trials, or time-to-failure trials. Long-term mortality trials in oncology, cardiology or HIV usually select time-to-event as the primary endpoint. The group sequential methodology is particularly appropriate for such trials because of the potential to shorten the study duration and thereby bring beneficial new therapies to patients sooner than would be possible by a conventional single-look design. In contrast to studies involving normal and binomial endpoints, the statistical power of a time-to-event study is determined, not by the number of individuals accrued, but rather by the number of events observed. Accruing only as many individuals as the number of events required to satisfy power considerations implies having to wait until all individuals have reached the event. This will probably make the trial extend over an unacceptably long period of time. Therefore one usually accrues a larger number of patients than the number of events required, so that the study may be completed within a reasonable amount of time. East allows the user a high degree of flexibility in this respect.

This volume contains Chapters 42 through 47. Chapter 42 is the present chapter. It describes the contents of the remaining chapters of Volume 6.

Chapter 43 introduces you to East on the Architect platform, using an example clinical trial to compare survival in two groups.

In Chapter 44 we discuss the Randomized Aldactone Evaluation Study (RALES) for decreasing mortality in patients with severe heart failure (Pitt et al., 1999). The chapter illustrates how East may be used to design and monitor a group sequential two-sample superiority trial with a time-to-event endpoint. We begin with the simple case of a constant enrollment rate, exponential survival and no drop-outs. The example is then extended to cover non-uniform enrollment, non-constant hazard rates for survival, and differential drop-out rates between the treatment and control arms. The role of simulation in providing additional insights is discussed. Simulations in presence of non-proportional hazard rates, stratification variables are explained. The trial was designed so that every subject who had not dropped out or reached the stated endpoint would be followed until the trial was terminated. This is an example of a **variable follow-up design**, because subjects who are enrolled at the beginning of the enrollment phase are followed for a longer time than subjects who are enrolled later.

In contrast to Chapter 44, Chapter 45 deals with the **fixed follow-up design**. Here we

design a trial in which each subject can only be followed for a maximum of one year and then goes off study. We use East to design such a trial basing the design parameters on the PASSION and TYPHOON trials – two recently published studies of drug eluting stents (Spaulding et al., 2006; Laarman et al., 2006). The impact of variable accrual patterns and drop-outs is also taken into account.

Chapter 46 shows how to use East to design a non-inferiority trial with a time-to-event endpoint. The setting is a clinical trial to demonstrate the non-inferiority of Xeloda to 5-FU+LV in patients with metastatic colorectal cancer (Rothman et al., 2003). Part of the discussion in this chapter is about the choice of the non-inferiority margin.

Chapter 47 will illustrate through a worked example how to design, monitor and simulate a two-sample non-inferiority trial with a time-to-event endpoint in which each subject who has not dropped out or experienced the event is followed for a fixed duration only. This implies that each subject who does not drop-out or experience the event within a given time interval, as measured from the time of randomization, will be administratively censored at the end of that interval. In East we refer to such designs as **fixed follow-up designs**.

Chapters 48 and 49 handle the trade-off between patient accruals and study duration in a different way from the previous chapters. In contrast to publicly funded trials, which usually lack the resources to exert control over the accrual rate of a trial, industry trials are often run with a fixed timeframe as the constraint. Industry sponsors would rather adjust the patient recruitment rate by opening and closing investigator sites than delay the end of a study and therefore their entire drug development program, time to market, and revenue. Chapters 48 and 49 illustrate how to design superiority and non-inferiority trials in East given a fixed accrual period and fixed study duration. Additionally, these design options provide the users with many useful graphs that chart the relationship between power, sample size, number of events, accrual duration, and study duration.

Also note that Chapter 44 contains a section that guides the user through the powerful survival simulation tool available in East.

Chapter 50 is a note which gives details on specifying dropout parameters for survival studies in East with the help of an example.

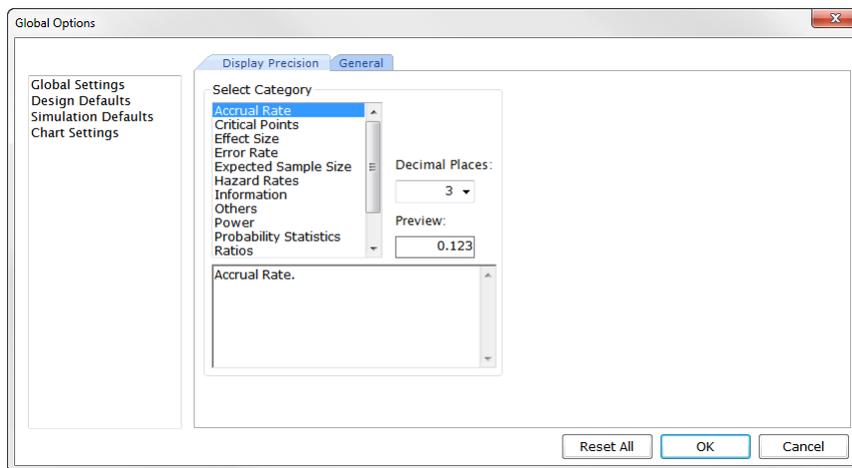
A unified formula for calculating the expected number of events $d(l)$ in a time-to-event trial can be found in the Appendix D.

42 Introduction to Volume 6

42.1 Settings



Click the **Global Options** icon in the **Home** menu to adjust default values in East 6.

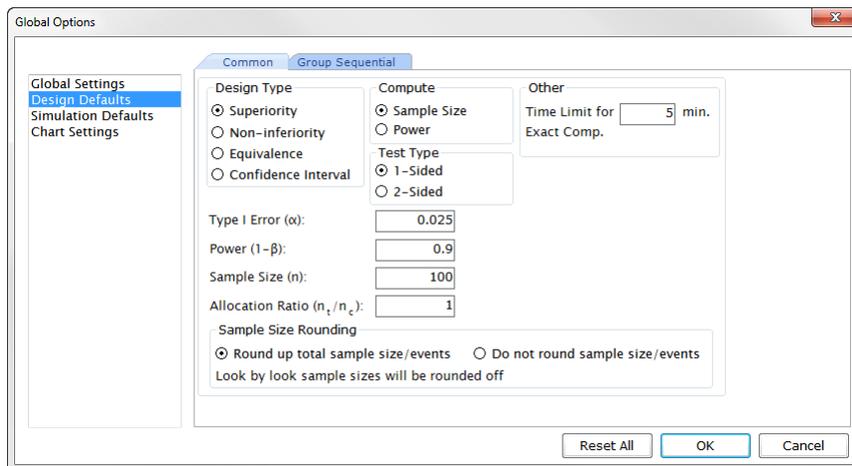


The options provided in the **Display Precision** tab are used to set the decimal places of numerical quantities. The settings indicated here will be applicable to all tests in East 6 under the **Design** and **Analysis** menus.

All these numerical quantities are grouped in different categories depending upon their usage. For example, all the average and expected sample sizes computed at simulation or design stage are grouped together under the category "Expected Sample Size". So to view any of these quantities with greater or lesser precision, select the corresponding category and change the decimal places to any value between 0 to 9.

The **General** tab has the provision of adjusting the paths for storing workbooks, files, and temporary files. These paths will remain throughout the current and future sessions even after East is closed. This is the place where we need to specify the installation directory of the R software in order to use the feature of R Integration in East 6.

The **Design Defaults** is where the user can change the settings for trial design:



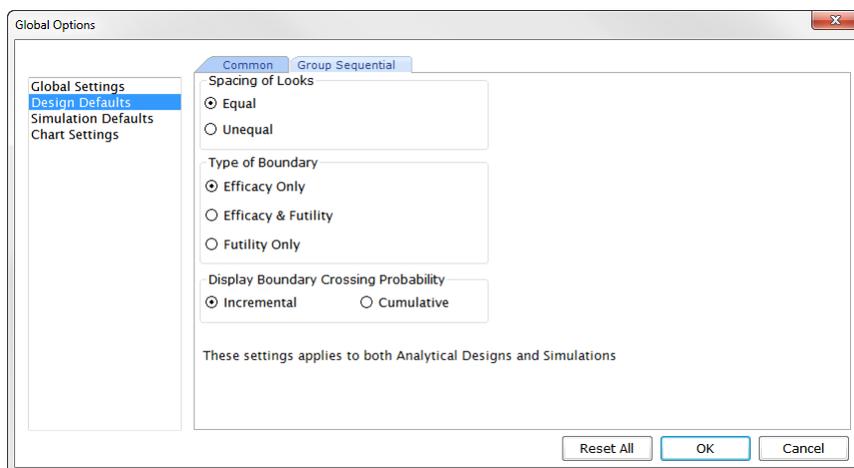
Under the **Common** tab, default values can be set for input design parameters.

You can set up the default choices for the design type, computation type, test type and the default values for type-I error, power, sample size and allocation ratio. When a new design is invoked, the input window will show these default choices.

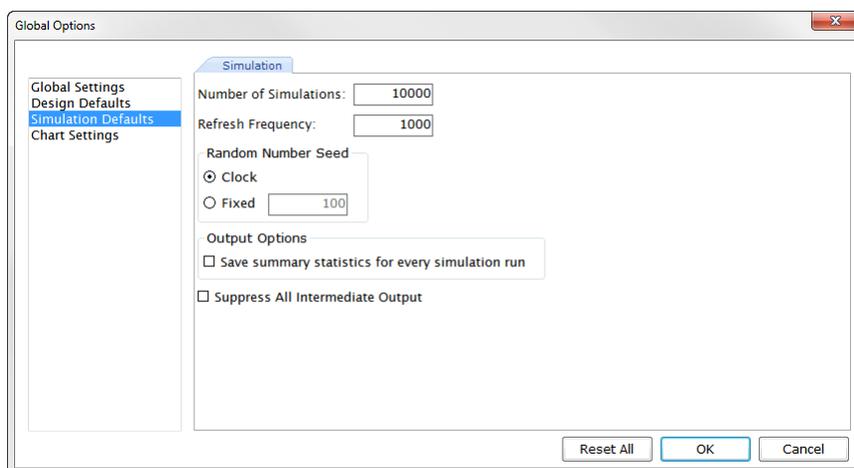
- **Time Limit for Exact Computation**
 This time limit is applicable only to exact designs and charts. Exact methods are computationally intensive and can easily consume several hours of computation time if the likely sample sizes are very large. You can set the maximum time available for any exact test in terms of minutes. If the time limit is reached, the test is terminated and no exact results are provided. Minimum and default value is 5 minutes.
- **Type I Error for MCP**
 If user has selected 2-sided test as default in global settings, then any MCP will use half of the alpha from settings as default since MCP is always a 1-sided test.
- **Sample Size Rounding**
 Notice that by default, East displays the integer sample size (events) by rounding up the actual number computed by the East algorithm. In this case, the look-by-look sample size is rounded off to the nearest integer. One can also see the original floating point sample size by selecting the option "Do not round sample size/events".

42 Introduction to Volume 6

Under the **Group Sequential** tab, defaults are set for boundary information. When a new design is invoked, input fields will contain these specified defaults. We can also set the option to view the Boundary Crossing Probabilities in the detailed output. It can be either Incremental or Cumulative.



Simulation Defaults is where we can change the settings for simulation:

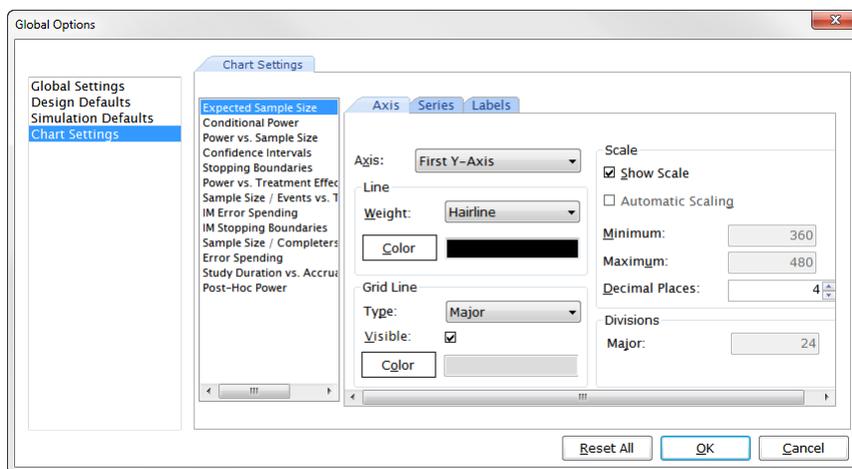


If the checkbox for "Save summary statistics for every simulation" is checked, then East simulations will by default save the per simulation summary data for all the

simulations in the form of a case data.

If the checkbox for "Suppress All Intermediate Output" is checked, the intermediate simulation output window will be always suppressed and you will be directed to the **Output Preview** area.

The **Chart Settings** allows defaults to be set for the following quantities on East6 charts:



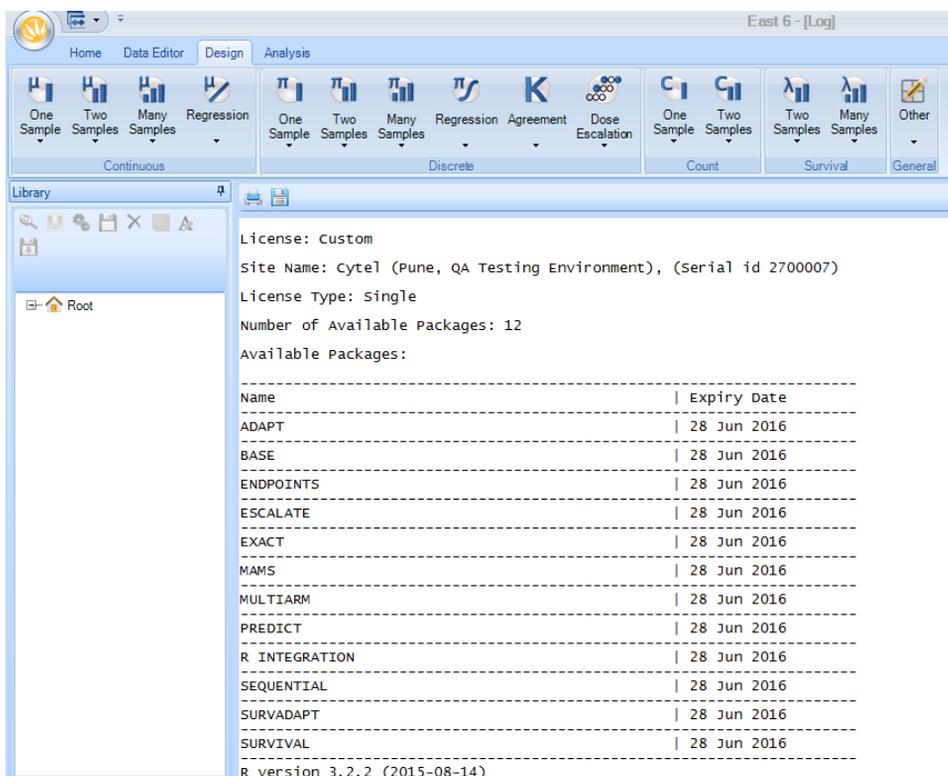
We suggest that you do not alter the defaults until you are quite familiar with the software.

43 *Tutorial: Survival Endpoint*

This tutorial introduces you to East 6, using examples for designing a clinical trial to compare survival in two groups. It is suggested that you go through the tutorial while you are at the computer, with East 6 running in it.

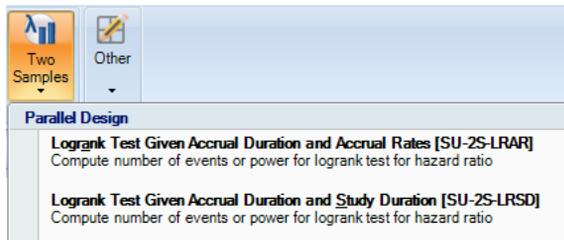
43.1 *A Quick Feel of the Software*

When you open East 6, the screen will look as shown below.

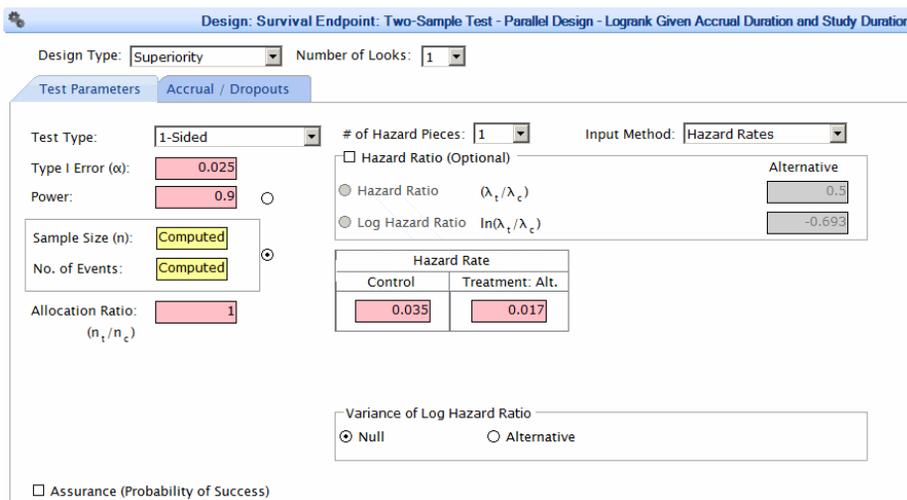


In the tabs bar at the top of the ribbon, Design tab is already selected. Each tab has its own ribbon. All the commands buttons under Design tab are displayed in its ribbon, with suggestive icons. These commands have been grouped under the categories of Continuous, Discrete, Count, Survival and General. For this tutorial, let us explore the command **Two Samples** under **Survival** category. In East, we use the terms 'time to event' and 'survival' interchangeably. Click on **Two Samples**. You will see a list of

action items, which are dialog box launchers.



Click on **Logrank Test Given Accrual Duration and Study Duration**. You will get the following dialog box in the work area.



This dialog box is for computing Sample Size (n) and Number of Events. All the default input specifications under the tab Design Parameters are on display: Design Type=Superiority, Number of Looks=1, Test Type=1-Sided, Type-1 Error (α)=0.025, Power ($1-\beta$)=0.9, Allocation Ratio (n_t/n_c)=1, # of Hazard Pieces=1, Input Method=Hazard Rates, Hazard Ratio (λ_t/λ_c)=0.5, Log Hazard Ratio $\ln(\lambda_t/\lambda_c)$ =-0.693, Hazard Rate (Control)=0.0347, Hazard Rate (Treatment)=0.0173, and Variance of Log-Hazard Ratio=Null. There are two radio buttons in this dialog box, one at the side of Power ($1-\beta$) box and the second at the side of the combined boxes for Sample Size (n) and Number of Events. By default, the latter radio button is

43 Tutorial: Survival Endpoint

selected indicating that the items against this radio button are to be computed using all other inputs. Similarly, if the first radio button is selected, then Power will be computed using all other inputs.

Now click on the tab Accrual/Dropout and you will see the following dialog box.

The default specifications in this dialog box are: Subjects are followed=Until End of Study, Accrual Duration=22, Study Duration=38, # of Accrual Periods=1, and no Dropouts. Now accept all the default specifications that are displayed for this single look design and be ready to compute the Sample Size (n) and the Number of Events for the design. Click Compute.

At the end of the computation, you will see the results appearing at the bottom of the screen, in the Output Preview pane, as shown below.

| Output Preview | | | | | | | | | | | | | | |
|----------------|-------------|--------------|-----------|--------------------|-------|---------|-------------|------------------|------------------|----------------|------------------|------------------|--------------------|----------------------------|
| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt / nc | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events | Exp. Events (H0) | Exp. Events (H1) | Comm. Accr. (Dur.) | Exp. Accrual Duration (H0) |
| Des1 | Superiority | 1 | 1-Sided | 0.025 | 0.902 | 1 | 182 | 182 | 182 | 88 | 88 | 88 | 22 | 22 |

| Exp. Accrual Duration (H1) | Hazard Ratio (Alt.) | Study Duration | Exp. Study Duration (H0) | Exp. Study Duration (H1) | Var (Log HR) | No. of Accrual Periods |
|----------------------------|---------------------|----------------|--------------------------|--------------------------|--------------|------------------------|
| 22 | 0.5 | 38 | 30.758 | 37.959 | Null | 1 |

This single row of output preview contains relevant details of all the inputs and the computed results for events and accruals. The maximum value for events is 88 and the committed accrual is 182 subjects. Since this is a fixed-look design, the expected events are same as the maximum required. Click anywhere in this row, and then click on the icon to get a detailed display in the upper pane of the screen as shown

below.

| | Des2 |
|---|--------------------|
| Mnemonic | SU-25-LRSD |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.902 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.5 |
| Var (Log HR) | Null |
| Accrual & Dropout Parameters | |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 182 |
| Expected Under H0 | 182 |
| Expected Under H1 | 182 |
| Events | |
| Maximum | 88 |
| Expected Under H0 | 88 |
| Expected Under H1 | 88 |
| Study Duration | |
| Maximum | 38 |
| Expected Under H0 | 30.758 |
| Expected Under H1 | 37.959 |
| Accrual Duration | |
| Maximum | 22 |
| Expected Under H0 | 22 |
| Expected Under H1 | 22 |

The contents of this output, displayed in the upper pane, are the same as what is contained in the output preview row for Design1 shown in the lower pane, but the upper pane display is easier to read and comprehend. The title of the upper pane display is **Output Summary**. This is because, you can choose more than one design in the Output Preview pane and the display in the upper pane will show the details of all the selected designs in juxtaposed columns.

43 Tutorial: Survival Endpoint

The discussion so far gives you a quick feel of the software for computing the required events and sample size for a single look survival design. We have not discussed about all the icons in the output preview pane or the library pane or the hidden Help pane in the screen. We will describe them taking an example for a group sequential design in the next section.

43.2 Group Sequential Design for a Survival Superiority Trial

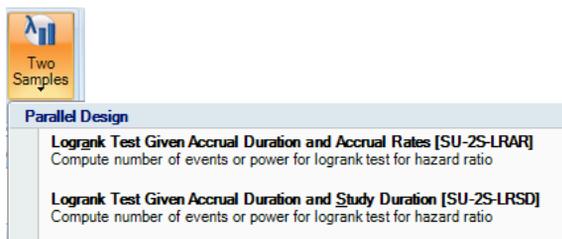
- 43.2.1 Background Information on the study
- 43.2.2 Creating the design in East
- 43.2.3 Design Outputs
- 43.2.4 East icons explained
- 43.2.5 Saving created designs
- 43.2.6 Displaying Detailed Output
- 43.2.7 Comparing Multiple Designs
- 43.2.8 Events vs. Time plot
- 43.2.9 Simulation
- 43.2.10 Interim Monitoring

43.2.1 Background Information on the study

The randomized aldactone evaluation study (RALES) was a double-blind multicenter clinical trial of aldeosterone-receptor blocker vs. placebo published in New England Journal of Medicine (vol 341, 10, pages 709-717, 1999). This trial was open to patients with severe heart failure due to systolic left ventricular dysfunction. The Primary endpoint was all-causes mortality. The anticipated accrual rate was 960 patients/year. The mortality rate for the placebo group was 38%. The investigators wanted 90% power to detect a 17% reduction in the mortality hazard rate for the Aldactone group (from 0.38 to 0.3154) with $\alpha = 0.05$, 2-sided test. Six DMC meetings were planned. The dropout rate in both the groups is expected to be 5% each year. The patient accrual period is planned to be 1.7 years and the total study duration to be 6 years.

43.2.2 Creating the design in East

For our purpose, let us create our own design from the basic details of this study. Now start afresh East. On the Design tab, click on **Two Samples** under **Survival** category. You will see a list of action items, which are dialog box launchers.



Click on the second option **Logrank Test Given Accrual Duration and Study**

Duration. You will get the following dialog box in the work area.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

Design Type: Superiority Number of Looks: 1

Test Parameters Accrual / Dropouts

Test Type: 1-Sided # of Hazard Pieces: 1 Input Method: Hazard Rates

Type I Error (α): 0.025

Power: 0.9

Sample Size (n): Computed

No. of Events: Computed

Allocation Ratio: 1 (n_1/n_2)

Hazard Ratio (Optional)

Hazard Ratio (λ_1/λ_2) Alternative: 0.5

Log Hazard Ratio $\ln(\lambda_1/\lambda_2)$ Alternative: -0.693

| Hazard Rate | |
|-------------|-----------------|
| Control | Treatment: Alt. |
| 0.035 | 0.017 |

Variance of Log Hazard Ratio

Null Alternative

Assurance (Probability of Success)

All the specifications you see in this dialog box are default values, which you will have to modify for the study under consideration.

Now, let the Design Type be Superiority.

Design Type: Superiority

Superiority

Noninferiority

Next, enter 6 in the Number of Looks box. You can see the range of choices for the

43 Tutorial: Survival Endpoint

number of looks is from 1 to 20.

Number of Looks:

1 ▲

2

3

4

5

6

7

8

9

10 ▼

Immediately after this selection, you will see a new tab **Boundary Info** added to the input dialog box. We will look into this tab, after you complete the filling of current tab **Design Parameters**.

Next, choose 2-Sided in the Test Type box.

Test Type:

2-Sided ▼

1-Sided

2-Sided

2-Sided (Asymmetric)

Next, enter 0.05 in the Type-1 Error (α) box, and 0.9 in the Power box.

Type I Error (α):

Power:

Next enter the specifications for survival parameters. Keep **# of Hazard Pieces** as 1. Click on the check box against Hazard Ratio and choose Hazard Rates as the Input Method. Enter 0.83 as the Hazard Ratio and 0.38 as the Hazard Rate (Control). East computes and displays the Hazard Rate (Treatment) as 0.3154. Keep the default choice of Null for Variance of Log-Hazard Ratio. Now the dialog box will look as shown

below.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

Design Type: Superiority Number of Looks: 6

Test Parameters Boundary **Accrual / Dropouts**

Test Type: 2-Sided # of Hazard Pieces: 1 Input Method: Hazard Rates

Type I Error (α): 0.05

Power: 0.9

Sample Size (n): Computed

No. of Events: Computed

Allocation Ratio: 1 (n₁, n₂)

Hazard Ratio (Optional) Alternative: 0.83

Hazard Ratio θ_1 / λ_2

Log Hazard Ratio $\ln \theta_1 / \lambda_2$ Alternative: -0.186

| Hazard Rate | |
|-------------|-----------------|
| Control | Treatment: Alt. |
| 0.38 | 0.315 |

Variance of Log Hazard Ratio

Null Alternative

Assurance (Probability of Success)

Next click the tab **Accrual/Dropout** . Keep the specification ‘Until End of Study’ for **Subjects are followed**. Enter 1.7 as **Accrual Duration** and 6 as **Study Duration**. Keep **# of Accrual Periods** as 1. Change the **# of Pieces** for dropouts to 1. Choose ‘Prob. of Dropout’ as the **Input Method** for entering information on dropouts. Enter 0.05 as probability of dropout at end of 1 year for both the groups. Now the dialog box will appear as shown below.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

Design Type: Superiority Number of Looks: 6

Test Parameters Boundary **Accrual / Dropouts**

Subjects are followed: Until End of Study

Accrual Info

Accrual Duration: 1.7 Study Duration: 6

of Accrual Periods: 1

| Period # | By Time | Cum. % Accrued |
|----------|---------|----------------|
| 1 | 1.700 | 100.000 |

Piecewise Dropout Information

of Pieces: 1 Input Method: Hazard Rates

Hazard Rate (Control): 0.05

Hazard Rate (Treatment): 0.05

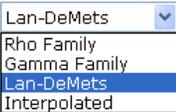
Now click on the **Boundary** tab. In the dialog box of this tab, you can specify stopping boundaries for efficacy or futility or both. For this trial, let us consider only Efficacy boundaries only. Choose ‘Spending Functions’ as the Efficacy Boundary Family.

Boundary Family: Spending Functions

- None
- Spending Functions**
- Haybittle Peto (p-value)
- Wang-Tsiatis

43 Tutorial: Survival Endpoint

Choose 'Lan-DeMets' in the Spending Function box.

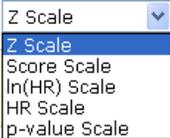
Spending Function: 

Choose 'OF' in the Parameter box.

Parameter: 

Next, click the radio button near 'Equal' for Spacing of Looks.

Choose 'Z Scale' in the Efficacy Boundary Scale box.

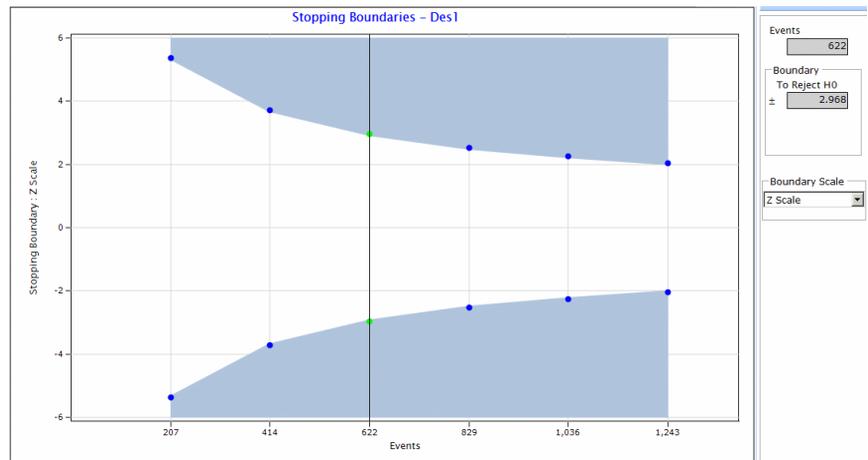
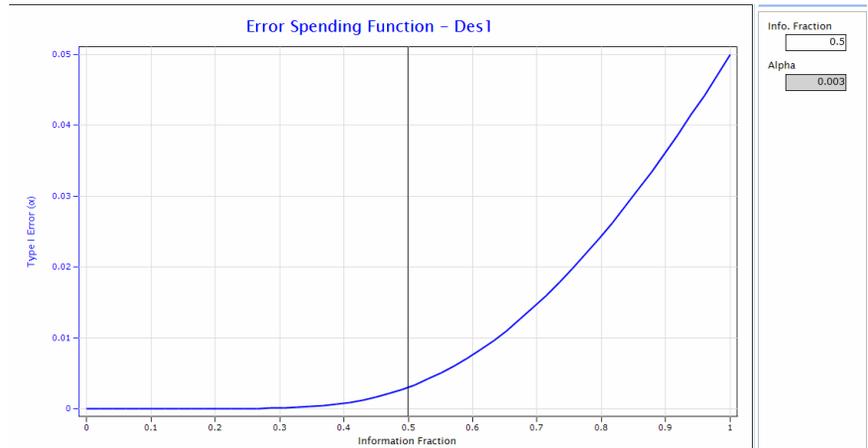
Efficacy Boundary: 

In the table below of look-wise details, the columns - Info Fraction, Cumulative Alpha Spent, and the upper and lower efficacy boundaries are computed and displayed as shown here. Scroll a little bit to see the sixth look details.

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|---------|
| | | | Upper | Lower |
| 1 | 0.167 | 0.0000 | 5.3666 | -5.3666 |
| 2 | 0.333 | 0.0002 | 3.7103 | -3.7103 |
| 3 | 0.500 | 0.0031 | 2.9697 | -2.9697 |
| 4 | 0.667 | 0.0121 | 2.5387 | -2.5387 |
| 5 | 0.833 | 0.0282 | 2.2522 | -2.2522 |

The two icons  and  represent buttons for Error Spending Function chart and Stopping Boundaries chart respectively. Click these two buttons one by one to see

the following charts.



43 Tutorial: Survival Endpoint

43.2.3 Design Outputs

Now you have completed specifying all the inputs required for a group sequential trial design and you are ready to compute the required events and sample size or accruals for the trial. Click on the **Compute** button. After the computation is over, East will show in the Output Preview pane the following results:

| Output Preview | | | | | | | | | | | | | | |
|----------------|-------------|--------------|-----------|--------------------|-------|-------|-------------|------------------|------------------|----------------|------------------|------------------|--------------------|----------------------------|
| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events | Exp. Events (H0) | Exp. Events (H1) | Comm. Accr. (Dur.) | Exp. Accrual Duration (H0) |
| Des1 | Superiority | 6 | 2-Sided | 0.05 | 0.9 | 1 | 1642 | 1641.997 | 1641.986 | 1243 | 1233.984 | 903.595 | 1.7 | 1.7 |

| Exp. Accrual Duration (H1) | Hazard Ratio (Alt.) | Study Duration | Exp. Study Duration (H0) | Exp. Study Duration (H1) | Var (Log HR) | No. of Accrual Periods | Spacing of Looks | Efficacy Boundary |
|----------------------------|---------------------|----------------|--------------------------|--------------------------|--------------|------------------------|------------------|-------------------|
| 1.7 | 0.83 | 6 | 5.359 | 3.729 | Null | 1 | Equal | LD (OF) |

This single row of output preview contains relevant details of all the inputs and the computed results for events and accruals. The maximum required Events is computed as 1243 and the Committed Accrual to be 1646 subjects. The expected Events under H0 and H1 are estimated to be 1234 and 904 respectively. The expected Study Duration under H0 and H1 are 5.359 and 3.729 respectively.

Click anywhere in this Output Preview row and then click on  icon to get a

summary in the upper pane of the screen as shown below.

| | Des 1 |
|---|--------------------|
| Mnemonic | SU-2S-LRSD |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 6 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.83 |
| Var (Log HR) | Null |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Accrual & Dropout Parameters | |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 1 |
| Sample Size | |
| Maximum | 1642 |
| Expected Under H0 | 1641.997 |
| Expected Under H1 | 1641.986 |
| Events | |
| Maximum | 1243 |
| Expected Under H0 | 1233.984 |
| Expected Under H1 | 903.595 |
| Study Duration | |
| Maximum | 6 |
| Expected Under H0 | 5.359 |
| Expected Under H1 | 3.729 |
| Accrual Duration | |
| Maximum | 1.7 |
| Expected Under H0 | 1.7 |
| Expected Under H1 | 1.7 |

43.2.4 East icons explained

In the 'Output Preview' pane, you see the following icons in the upper row.



The functions of the above icons are as indicated below. The tooltips also will indicate their functions.

43 Tutorial: Survival Endpoint

-  Output Summary (The output summary of selected design(s) will appear in the upper pane)
-  Edit Design (The input dialog box of a selected design will appear in the upper pane)
-  Save in Workbook (Save one or more selected designs in a workbook)
-  Delete (Delete one or more selected designs)
-  Rename (Rename Design names)
-  Print (Print selected designs)
-  Display Precision (Local Settings)
-  Filter (Filter and select designs according to specified conditions)
-  Show/Hide Columns (Show/Hide Columns of the designs in the Output Preview panel)

The following icons can be seen at the right end of Output Preview pane and Output Summary or Input/Output window respectively. Their functions are:

-  Maximize Output Preview Pane
-  Minimize Output Preview Pane

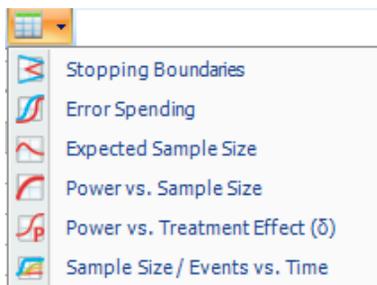
You may also notice a row of icons at the top of Output Summary window as shown below.



The first icon is for Plot (Plots of a selected design will appear in a pop-up window).



The second icon is for Show Tables (The data for different plots can be displayed in tabular form in pop-up windows).



If you have multiple designs in the output summary window, the third icon becomes active and can be used to move the order of those columns in the Output Summary.

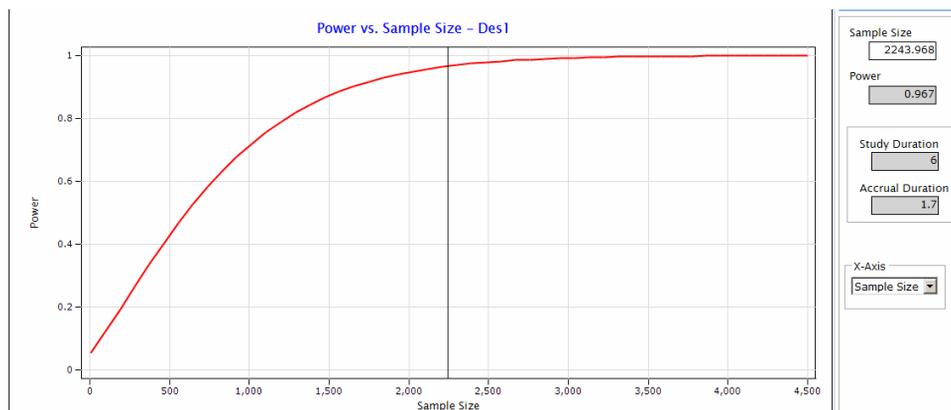


The fourth icon is to print the Output Summary window.

As an example, if you click Power vs. Sample Size under Plot icon, you will get the

43 Tutorial: Survival Endpoint

following chart.



If you want to see the data underlying the above chart, click Show Table icon and click

Power vs. Sample Size. You will see the following table in a pop-up window.

| Range for No. of Events | | | Input |
|-------------------------|------|-----------|-----------------|
| From | To | Step Size | No. of Events |
| 4 | 3200 | 65.2245 | ▼ |
| | | | Tabulate |

| No. of Events | Power_Des1 |
|---------------|------------|
| 4 | 0.0538 |
| 69.224 | 0.1183 |
| 134.449 | 0.1853 |
| 199.673 | 0.2529 |
| 264.898 | 0.3194 |
| 330.122 | 0.3838 |
| 395.347 | 0.4452 |
| 460.571 | 0.503 |
| 525.796 | 0.5568 |
| 591.02 | 0.6065 |
| 656.245 | 0.652 |
| 721.469 | 0.6934 |
| 786.694 | 0.7308 |
| 851.918 | 0.7644 |
| 917.143 | 0.7944 |

You can customize the format of the above table and also save it as case data in a workbook. You may experiment with all the above icon / buttons to understand their functions.

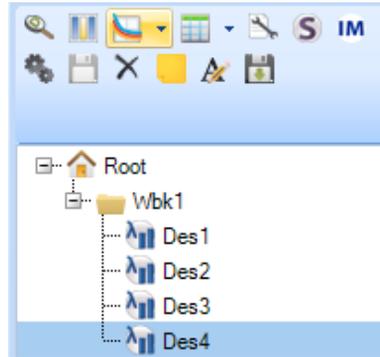
43.2.5 Saving created Designs in the library and hard disk

In the Output Preview pane, select one or more design rows and click the  icon,

The selected design(s) will then get added as a node(s) in the current workbook, as

43 Tutorial: Survival Endpoint

shown below.



The above action only adds the design to the workbook node in the library and it is not saved in the hard disk. For saving in the hard disk, you may either save the entire workbook or only the design by right-clicking on the desired item and choosing save or save as options.

Here in the library also, you see rows of icons.



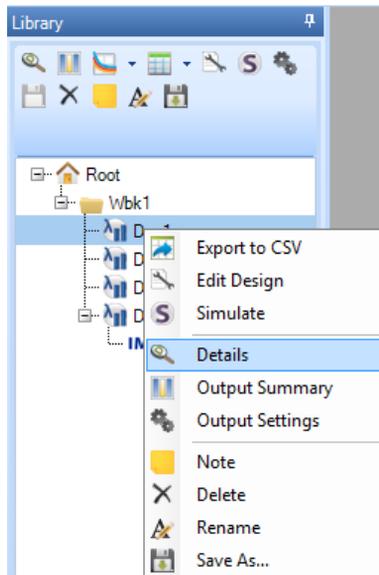
Some of these icons you have already seen. The functions of other icons are:

-  Details (Details of a selected design will appear on the upper pane in the work area)
-  Output Settings (Output Settings can be changed here)
-  Simulate (Start the simulation process for any selected design node)
-  Interim Monitoring (Start the Interim Monitoring process for any selected design)

43.2.6 Displaying Detailed Output

Select the design from the Library and click the  icon or Right-click on the Des1

node in the library and click Details.



You will see the detailed output of the design displayed in the Work area.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

| Test Parameters | |
|---------------------------------------|-------------|
| Design ID | Des1 |
| Design Type | Superiority |
| Number of Looks | 6 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| HR = λ_1/λ_0 | |
| Under H0 | 1 |
| Under H1 | 0.83 |
| $\delta = \ln(\text{HR})$ | -0.186 |
| Hazard Rate Control (λ_0) | 0.38 |
| Hazard Rate Treatment (λ_1) | 0.315 |
| Var (Log HR) | Null |
| Allocation Ratio (n_1/n_0) | 1 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Accrual / Dropouts Parameters | |
| Cum % Accrued | 100 |
| Accrual Duration | 1.7 |
| Max Study Duration | 6 |
| Dropout | Yes |
| Dropout Hazard Rate Control | 0.05 |
| Dropout Hazard Rate Treatment | 0.05 |

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|-------------------------|-------------|---------------|----------|
| Sample Size (n) | | | |
| Maximum | 820 | 822 | 1642 |
| Expected H1 | 820.993 | 820.993 | 1641.986 |
| Expected H0 | 820.998 | 820.998 | 1641.997 |
| Events (s) | | | |
| Maximum | 644 | 599 | 1243 |
| Expected H1 | 510.403 | 457.134 | 903.595 |
| Expected H0 | 618.836 | 618.836 | 1233.984 |
| Dropouts (d) | | | |
| Maximum | 85 | 95 | 180 |
| Expected H1 | 62.794 | 68.022 | 130.816 |
| Expected H0 | 81.402 | 81.419 | 162.821 |
| Maximum Information (I) | | | 310.75 |

Accrual and Study Duration

| | Accrual Duration | Study Duration |
|-------------|------------------|----------------|
| Maximum | 1.7 | 5.993 |
| Expected H1 | 1.7 | 3.729 |
| Expected H0 | 1.7 | 5.359 |

Stopping Boundaries: Look by Look

43 Tutorial: Survival Endpoint

able Follow-Up Design: All subjects e followed until failure, drop out or end of udy.

ample sizes and events have been unded.

| Look # | Info. Fraction (s/s_max) | Events (s) | Cumulative α Spent | Boundaries Efficacy Z | |
|--------|--------------------------|------------|---------------------------|-----------------------|--------|
| | | | | Upper | Lower |
| 1 | 0.167 | 207 | 7.926E-8 | 5.369 | -5.369 |
| 2 | 0.333 | 414 | 2.057E-4 | 3.712 | -3.712 |
| 3 | 0.5 | 622 | 0.003 | 2.968 | -2.968 |
| 4 | 0.667 | 829 | 0.012 | 2.538 | -2.538 |
| 5 | 0.833 | 1036 | 0.028 | 2.252 | -2.252 |
| 6 | 1 | 1243 | 0.05 | 2.045 | -2.045 |

Events, Sample Size, Dropouts, Pipeline and Analysis Times: Look by Look (Under H0)

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (s) | Dropouts (d) | Pipeline (n-s-d) | Analysis Time | Boundary Crossing Probability (Incremental) Efficacy | |
|--------|--------------------------|-----------------|------------|--------------|------------------|---------------|--|----------|
| | | | | | | | Upper | Lower |
| | | | | | | | 1 | 0.167 |
| 2 | 0.333 | 1626 | 414 | 55 | 1157 | 1.683 | 1.028E-4 | 1.028E-4 |
| 3 | 0.5 | 1642 | 622 | 82 | 938 | 2.203 | 0.001 | 0.001 |
| 4 | 0.667 | 1642 | 829 | 110 | 703 | 2.871 | 0.005 | 0.005 |
| 5 | 0.833 | 1642 | 1036 | 137 | 469 | 3.812 | 0.008 | 0.008 |
| 6 | 1 | 1642 | 1243 | 164 | 235 | 6.418 | 0.011 | 0.011 |

Events, Sample Size, Dropouts, Pipeline and Analysis Times: Look by Look (Under H1)

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (s) | Dropouts (d) | Pipeline (n-s-d) | Analysis Time | Boundary Crossing Probability (Incremental) Efficacy | |
|--------|--------------------------|-----------------|------------|--------------|------------------|---------------|--|-------|
| | | | | | | | Upper | Lower |
| | | | | | | | 1 | 0.167 |
| 2 | 0.333 | 1642 | 414 | 60 | 1168 | 1.755 | 1.026E-8 | 0.035 |
| 3 | 0.5 | 1642 | 622 | 90 | 930 | 2.329 | 5.886E-8 | 0.226 |
| 4 | 0.667 | 1642 | 829 | 120 | 693 | 3.071 | 7.835E-8 | 0.303 |
| 5 | 0.833 | 1642 | 1036 | 150 | 456 | 4.13 | 5.829E-8 | 0.218 |
| 6 | 1 | 1642 | 1243 | 180 | 219 | 5.993 | 3.309E-8 | 0.119 |

43.2.7 Comparing Multiple Designs

Click on Des1 row and then click Edit icon  . You will get the input dialog box in the upper pane. Change the Power value to 0.8 and then click Compute.

You will see now Des2 is created and a row added to Output Preview pane as shown below.

| Output Preview | | | | | | | | | | | | | | |
|----------------|-------------|--------------|-----------|--------------------|-------|-------|-------------|------------------|------------------|----------------|------------------|------------------|--------------------|----------------------------|
| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events | Exp. Events (H0) | Exp. Events (H1) | Comm. Accr. (Dur.) | Exp. Accrual Duration (H0) |
| Des1 | Superiority | 6 | 2-Sided | 0.05 | 0.9 | 1 | 1642 | 1641.997 | 1641.986 | 1243 | 1233.984 | 903.595 | 1.7 | 1.7 |
| Des2 | Superiority | 6 | 2-Sided | 0.05 | 0.8 | 1 | 1230 | 1229.997 | 1229.995 | 931 | 924.247 | 736.316 | 1.7 | 1.7 |

Click on Des1 row and then keeping Ctrl key pressed, click on Des2 row. Now both the rows will be selected. Next, click the Output Summary icon  .

Now you will see the output details of these two designs displayed in the upper pane Compare Designs in juxtaposed columns, as shown below.

| | Des 1 | Des 2 |
|---|--------------------|--------------------|
| Power | 0.9 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Hazard Ratio (Alt.) | 0.83 | 0.83 |
| Var (Log HR) | Null | Null |
| Boundary Parameters | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) |
| Accrual & Dropout Parameters | | |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 |
| Sample Size | | |
| Maximum | 1642 | 1230 |
| Expected Under H0 | 1641.997 | 1229.997 |
| Expected Under H1 | 1641.986 | 1229.995 |
| Events | | |
| Maximum | 1243 | 931 |
| Expected Under H0 | 1233.984 | 924.247 |
| Expected Under H1 | 903.595 | 736.316 |
| Study Duration | | |
| Maximum | 6 | 6 |
| Expected Under H0 | 5.359 | 5.358 |
| Expected Under H1 | 3.729 | 4.207 |
| Accrual Duration | | |
| Maximum | 1.7 | 1.7 |
| Expected Under H0 | 1.7 | 1.7 |
| Expected Under H1 | 1.7 | 1.7 |

In a similar way, East allows the user to easily create multiple designs by specifying a range of values for certain parameters in the design window. For example, in a survival trial the **Logrank Test given Accrual Duration and Study Duration** design allows the input of multiple key parameters at once to simultaneously create a number of different designs. For example, suppose in a multi-look study the user wants to generate designs for all combinations of the following parameter values: **Power** = 0.8 and 0.9, and **Hazard Ratio - Alternative** = 0.6, 0.7, 0.8 and 0.9. The number of combinations is 2 x 4 = 8. East creates all permutations using only a single specification under the **Design Parameters** tab in the design window. As shown below, the values for **Power** are entered as a list of comma separated values, while the alternative hazard ratios are entered as a colon separated range of values, 0.6 to 0.9 in

43 Tutorial: Survival Endpoint

steps of 0.1.

East computes all 8 designs and displays them in the **Output Preview** window:

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events | Exp. Events (H0) | Exp. Events (H1) | Comm. Accr. (Dur.) | Exp. Accrual Duration (H0) |
|-------|-------------|--------------|-----------|--------------------|-------|-------|-------------|------------------|------------------|----------------|------------------|------------------|--------------------|----------------------------|
| Des3 | Superiority | 6 | 2-Sided | 0.05 | 0.801 | 1 | 125 | 124.489 | 111.053 | 124 | 123.101 | 98.054 | 22 | 21.91 |
| Des4 | Superiority | 6 | 2-Sided | 0.05 | 0.8 | 1 | 255 | 253.992 | 225.034 | 254 | 252.158 | 200.903 | 22 | 21.913 |
| Des5 | Superiority | 6 | 2-Sided | 0.05 | 0.8 | 1 | 650 | 647.451 | 569.871 | 649 | 644.291 | 513.311 | 22 | 21.914 |
| Des6 | Superiority | 6 | 2-Sided | 0.05 | 0.8 | 1 | 2912 | 2900.636 | 2539.053 | 2910 | 2888.9 | 2301.893 | 22 | 21.914 |
| Des7 | Superiority | 6 | 2-Sided | 0.05 | 0.901 | 1 | 167 | 166.326 | 141.229 | 166 | 164.797 | 120.548 | 22 | 21.911 |
| Des8 | Superiority | 6 | 2-Sided | 0.05 | 0.901 | 1 | 341 | 339.649 | 285.526 | 340 | 337.535 | 246.997 | 22 | 21.913 |
| Des9 | Superiority | 6 | 2-Sided | 0.05 | 0.9 | 1 | 868 | 864.598 | 720.944 | 867 | 860.714 | 630.127 | 22 | 21.914 |
| Des10 | Superiority | 6 | 2-Sided | 0.05 | 0.9 | 1 | 3890 | 3874.833 | 3208.566 | 3888 | 3859.809 | 2826.252 | 22 | 21.914 |

East provides the capability to analyze multiple designs in ways that make comparisons between the designs visually simple and efficient. To illustrate this, a selection of a few of the above designs can be viewed simultaneously in both the **Output Summary** section as well as in the various tables and plots. The following is a subsection of the designs computed from the above example with differing values for number of looks, power and hazard ratio. Designs are displayed side by side, allowing

details to be easily compared:

| | Des3 | Des4 | Des5 | Des6 |
|---|--------------------|--------------------|--------------------|--------------------|
| Mnemonic | SU-2S-LRSD | SU-2S-LRSD | SU-2S-LRSD | SU-2S-LRSD |
| Test Parameters | | | | |
| Design Type | Superiority | Superiority | Superiority | Superiority |
| No. of Looks | 6 | 6 | 6 | 6 |
| Test Type | 2-Sided | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 | 0.05 |
| Power | 0.801 | 0.8 | 0.8 | 0.8 |
| Model Parameters | | | | |
| Allocation Ratio (N1:nc) | 1 | 1 | 1 | 1 |
| Hazard Ratio (Alt.) | 0.6 | 0.7 | 0.8 | 0.9 |
| Var (Log HR) | Null | Null | Null | Null |
| Boundary Parameters | | | | |
| Spacing of Looks | Equal | Equal | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) | LD (OF) |
| Accrual & Dropout Parameters | | | | |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 | 1 | 1 | 1 |
| No. of Dropout Pieces | 0 | 0 | 0 | 0 |
| Sample Size | | | | |
| Maximum | 125 | 255 | 650 | 2912 |
| Expected Under H0 | 124.489 | 253.992 | 647.451 | 2900.636 |
| Expected Under H1 | 111.053 | 225.034 | 569.871 | 2539.053 |
| Events | | | | |
| Maximum | 124 | 254 | 649 | 2910 |
| Expected Under H0 | 123.101 | 252.158 | 644.291 | 2888.9 |
| Expected Under H1 | 98.054 | 200.903 | 513.311 | 2301.893 |
| Study Duration | | | | |
| Maximum | 38 | 38 | 38 | 38 |
| Expected Under H0 | 28.829 | 30.651 | 33.045 | 35.107 |
| Expected Under H1 | 23.644 | 23.712 | 24.103 | 24.354 |
| Accrual Duration | | | | |
| Maximum | 22 | 22 | 22 | 22 |
| Expected Under H0 | 21.91 | 21.913 | 21.914 | 21.914 |
| Expected Under H1 | 19.545 | 19.415 | 19.288 | 19.182 |

In addition East allows multiple designs to be viewed simultaneously either graphically or in tabular format: Notice that all the four designs in the Output Summary window are selected. Following figures compare these four designs in different formats.

43 Tutorial: Survival Endpoint

Stopping Boundaries (table)

Stopping Boundaries

Boundary Scales : **Z Scale**

Des3

| Look # | Info. Fraction | Events | Cum. α Spent | Boundaries | | Sample Size | | Analysis Time | | Incremental Boundary Crossing Probabilities | | | |
|--------|----------------|--------|---------------------|-------------------|--------|-------------|----------|---------------|----------|---|-------|---------------------------------------|-------|
| | | | | Efficacy Boundary | | Under H0 | Under H1 | Under H0 | Under H1 | Under H0: $(\lambda_t/\lambda_{t-1})$ | | Under H1: $(\lambda_t/\lambda_{t-1})$ | |
| | | | | Upper | Lower | | | | | Efficacy | | Efficacy | |
| 1 | 0.169 | 21 | 0 | 5.322 | -5.322 | 35 | 38 | 6.065 | 6.613 | 0 | 0 | 0 | 0 |
| 2 | 0.331 | 41 | 0 | 3.727 | -3.727 | 56 | 60 | 9.784 | 10.5 | 0 | 0 | 0 | 0.018 |
| 3 | 0.5 | 62 | 0.003 | 2.969 | -2.969 | 77 | 82 | 13.528 | 14.332 | 0.001 | 0.001 | 0 | 0.152 |
| 4 | 0.669 | 83 | 0.012 | 2.532 | -2.532 | 98 | 103 | 17.236 | 18.08 | 0.005 | 0.005 | 0 | 0.255 |
| 5 | 0.831 | 103 | 0.028 | 2.258 | -2.258 | 118 | 123 | 20.759 | 21.621 | 0.008 | 0.008 | 0 | 0.218 |
| 6 | 1 | 124 | 0.05 | 2.044 | -2.044 | 125 | 125 | 29.117 | 33.474 | 0.011 | 0.011 | 0 | 0.158 |

Des4

| Look # | Info. Fraction | Events | Cum. α Spent | Boundaries | | Sample Size | | Analysis Time | | Incremental Boundary Crossing Probabilities | | | |
|--------|----------------|--------|---------------------|-------------------|--------|-------------|----------|---------------|----------|---|-------|---------------------------------------|-------|
| | | | | Efficacy Boundary | | Under H0 | Under H1 | Under H0 | Under H1 | Under H0: $(\lambda_t/\lambda_{t-1})$ | | Under H1: $(\lambda_t/\lambda_{t-1})$ | |
| | | | | Upper | Lower | | | | | Efficacy | | Efficacy | |
| 1 | 0.165 | 42 | 0 | 5.389 | -5.389 | 70 | 74 | 5.984 | 6.355 | 0 | 0 | 0 | 0 |
| 2 | 0.335 | 85 | 0 | 3.702 | -3.702 | 115 | 121 | 9.904 | 10.385 | 0 | 0 | 0 | 0.02 |
| 3 | 0.5 | 127 | 0.003 | 2.97 | -2.97 | 158 | 164 | 13.573 | 14.102 | 0.001 | 0.001 | 0 | 0.15 |
| 4 | 0.665 | 169 | 0.012 | 2.542 | -2.542 | 200 | 206 | 17.208 | 17.758 | 0.004 | 0.004 | 0 | 0.248 |
| 5 | 0.835 | 212 | 0.028 | 2.249 | -2.249 | 243 | 249 | 20.921 | 21.479 | 0.008 | 0.008 | 0 | 0.229 |
| 6 | 1 | 254 | 0.05 | 2.045 | -2.045 | 255 | 255 | 30.993 | 34.179 | 0.011 | 0.011 | 0 | 0.153 |

Expected Sample Size (table)

Expected Sample Size / Completers

Expected Events vs. Effect Size (ln(δ))

Range for Effect Size (ln(δ))

From: 0.5108 To: 0.5108 Step Size: 0.0209

Output: Events

Tabulate

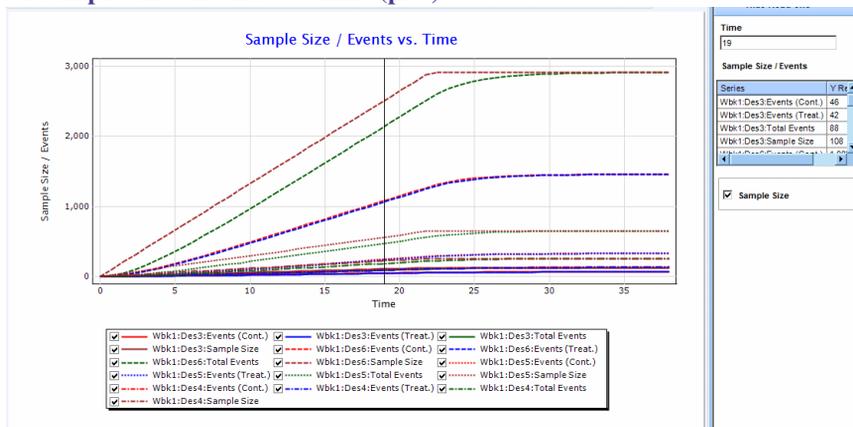
| Effect Size (ln(δ)) | Des3 | Des4 | Des5 | Des6 |
|------------------------------|----------|----------|----------|-----------|
| -0.511 | 98.0543 | 157.7905 | 274.1463 | 678.073 |
| -0.49 | 100.1039 | 163.0549 | 283.6335 | 721.9617 |
| -0.469 | 102.1151 | 168.567 | 293.964 | 766.1805 |
| -0.448 | 104.0743 | 174.3004 | 305.2224 | 808.5423 |
| -0.427 | 105.9685 | 180.2172 | 317.5214 | 847.2909 |
| -0.407 | 107.7856 | 186.2673 | 331.0004 | 881.51 |
| -0.386 | 109.5148 | 192.3892 | 345.8192 | 911.4633 |
| -0.365 | 111.1467 | 198.5118 | 362.1425 | 938.7922 |
| -0.344 | 112.6736 | 204.5574 | 380.1161 | 966.462 |
| -0.323 | 114.09 | 210.445 | 399.8297 | 998.3382 |
| -0.302 | 115.3919 | 216.0944 | 421.2725 | 1038.3976 |
| -0.281 | 116.5778 | 221.4313 | 444.2846 | 1089.8507 |
| -0.261 | 117.6478 | 226.391 | 468.5196 | 1154.7003 |
| -0.24 | 118.6039 | 230.9218 | 493.4314 | 1234.1664 |
| -0.219 | 119.4494 | 234.9881 | 518.2999 | 1329.8755 |
| -0.198 | 120.1892 | 238.5707 | 542.2995 | 1445.1624 |
| -0.177 | 120.829 | 241.6675 | 564.602 | 1585.6098 |
| -0.156 | 121.375 | 244.2909 | 584.4919 | 1757.687 |
| -0.136 | 121.834 | 246.4662 | 601.4662 | 1964.1053 |
| -0.115 | 122.2126 | 248.2271 | 615.2911 | 2196.0739 |
| -0.094 | 122.5171 | 249.6124 | 626.0035 | 2428.6701 |

Save as Case Data

Power vs. Sample Size (plot)



Total Sample Size / Events vs. Time (plot)



This capability allows the user to explore a greater space of possibilities when determining the best choice of study design.

43.2.8 Events vs. Time plot

For survival studies, East provides a variety of charts and plots to visually validate and

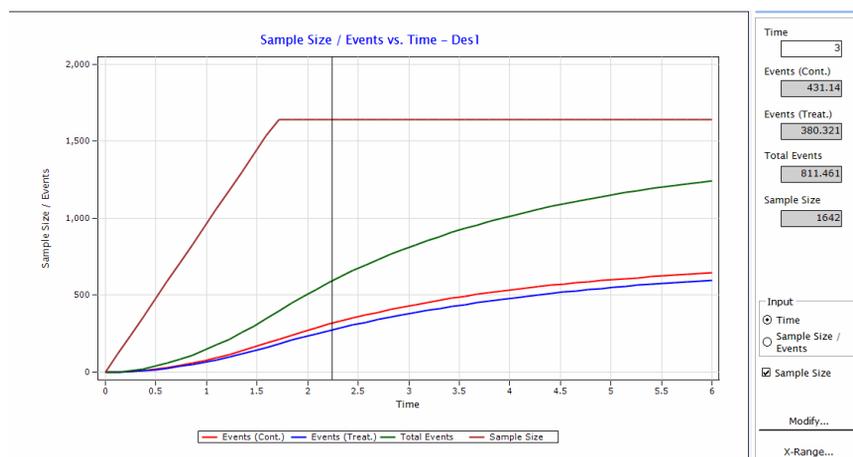
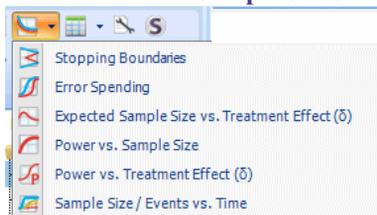
43 Tutorial: Survival Endpoint

analyze the design. For example, the **Sample Size / Events vs. Time** plot allows the user to see the rate of increase in the number of events (control and treatment) over time (accrual duration, study duration). An additional feature of this particular chart is that a user can easily update key input parameters to determine how multiple different scenarios can directly impact a study. This provides significant benefits during the design phase, as the user can quickly examine how a variety of input values affect a study before the potentially lengthy task of simulation is employed.

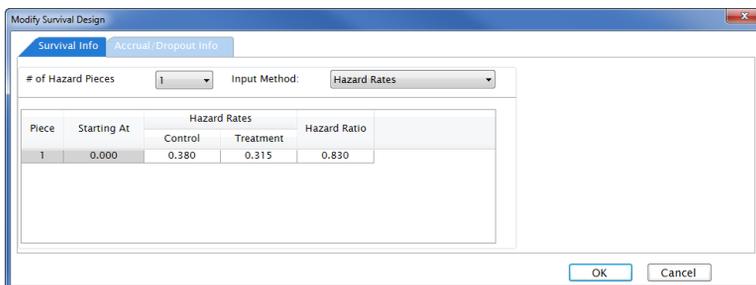
To illustrate this feature what follows is the example from the RALES study. For study details, refer to subsection **Background Information on the study** of this tutorial.

Currently there are ten designs in the **Output Preview** area. Select Des1 from them and save it to the current workbook. You may delete the remaining ones at this point.

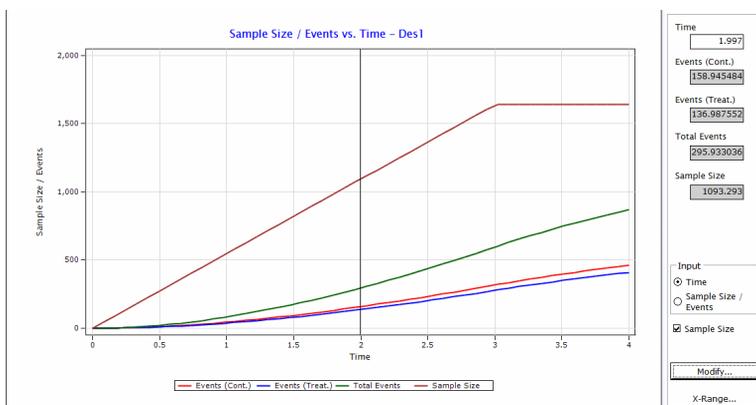
To view the **Sample Size / Events vs. Time** plot, select the corresponding node in the **Library** and under the **Charts** icon choose **Sample Size / Events vs. Time**:



Survival parameters for this design can be edited directly through this chart by clicking the **Modify** button. The **Modify Survival Design** window is then displayed for the user to update design parameters:



To illustrate the benefit of the modification feature, suppose at design time there is potential flexibility in the accrual and duration times for the study. To see how this may affect the number of subsequent events, modify the design to change the **Accrual Duration** to 3 and **Study Duration** to 4. Re-create the plot to view the effect of these new values on the shape and magnitude of the curves by clicking **OK**:



Similar steps can be taken to observe the effect of changing other parameter values on the number of events necessary to adequately power a study.

43 Tutorial: Survival Endpoint

43.2.9 Simulation

In the library, right-click on the node **Des1** and click **Simulate**. You will be presented with the following Simulation sheet.

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

Number of Looks: 6

Test Parameters | Response Generation | Accrual / Dropouts | Simulation Controls

Trial Type: Superiority
 Test Type: 2-Sided
 Fix at Each Look: Total No. of Events
 Total No. of Events: 1243

Test Statistic: Logrank

| Look # | Info. Fraction | Cum. α Spent | | Efficacy Z | |
|--------|----------------|---------------------|-------|------------|--------|
| | | Upper | Lower | Upper | Lower |
| 1 | 0.167 | 0.000 | 0.000 | 5.369 | -5.369 |
| 2 | 0.333 | 0.000 | 0.000 | 3.712 | -3.712 |
| 3 | 0.500 | 0.002 | 0.002 | 2.968 | -2.968 |
| 4 | 0.667 | 0.006 | 0.006 | 2.538 | -2.538 |
| 5 | 0.833 | 0.014 | 0.014 | 2.252 | -2.252 |

This sheet has four tabs - Test Parameters, Response Generation, Accrual/Dropout, and Simulation Controls. Additionally, you can click **Include Options** and add some more tabs like Site, Randomization, User Defined R Function and Stratification. The first three tabs essentially contain the details of the parameters of the design. In the Simulation Control tab, you can specify the number of simulations to carry out and specify the file for storing simulation data. Let us first carry out 1000 simulations to check whether the design can reach the specified power of 90%. The Response Generation tab, by default, shows the hazard rates for control and treatment. We will use these values in our simulation.

Survival Information

of Hazard Pieces: 1 | Input Method: Hazard Rates

| Hazard Rate | | Hazard Ratio |
|-------------|-----------|--------------|
| Control | Treatment | |
| 0.38 | 0.315 | 0.83 |

In the Simulation Control tab, specify the number of simulations as 1000. Use the

Random number seed as Fixed 12345.

Let us keep the values in other tabs as they are and click **Simulate**. The progress of simulation process will appear in a temporary window as shown below.

Simulations in progress

Input Summary: Sample Size = 1642; Total Events = 1243; HR = 0.83

| Look No. | Look Position (Events) | Efficacy Boundary | | Latest Simulated Test Stat. | Average Sample Size | Average Look Time | Stopping For (Incr.) | | | Total Simulations | |
|--------------|------------------------|-------------------|--------|-----------------------------|---------------------|-------------------|----------------------|----------------|----------------|-------------------|----------------|
| | | Upper | Lower | | | | Upper Efficacy | Lower Efficacy | No Efficacy | Count (Incr.) | % (Incr.) |
| 1 | 207.000 | 5.369 | -5.369 | -0.635 | 1159.589 | 1.200 | 0 | 0 | 0 | 0 | 0.000 |
| 2 | 414.000 | 3.712 | -3.712 | -1.365 | 1639.860 | 1.754 | 0 | 40 | 0 | 40 | 4.000 |
| 3 | 622.000 | 2.968 | -2.968 | -2.140 | 1642.000 | 2.329 | 0 | 214 | 0 | 214 | 21.400 |
| 4 | 829.000 | 2.538 | -2.538 | -2.555 | 1642.000 | 3.068 | 0 | 314 | 0 | 314 | 31.400 |
| 5 | 1036.000 | 2.252 | -2.252 | | 1642.000 | 4.122 | 0 | 222 | 0 | 222 | 22.200 |
| 6 | 1243.000 | 2.045 | -2.045 | | 1642.000 | 5.034 | 0 | 303 | 0 | 303 | 30.300 |
| Total | | | | | 1641.954 | 3.703 | | 892.000 | 108.000 | 1000 | 100.000 |
| | | | | | % | | | 0.000 | 89.200 | 10.800 | |

Stopping Boundaries

Stopping Decisions

Simulations Complete Waiting for User Action

Close

Simulation Seed = 12345 Elapsed Time = 00:00:13 1000/1000

This is the intermediate window showing the complete picture of simulations. Close this window after viewing it. You can see the complete simulation output in the details view. A new row, with the ID as Sim1, will be added in Output Preview.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events | Exp. Events (H0) | Exp. Events (H1) |
|------|-------------|--------------|-----------|--------------------|-------|-------|------------------|-------------------|-------------|------------------|------------------|----------------|------------------|------------------|
| Des1 | Superiority | 6 | 2-Sided | 0.05 | 0.9 | 1 | Equal | LD (OF) | 1642 | 1641.997 | 1641.986 | 1243 | 1233.984 | 903.595 |
| Sim1 | Superiority | 6 | 2-Sided | | 0.892 | | User Specified | User Specified | 1642 | | | 1243 | | |

43 Tutorial: Survival Endpoint

Click on Sim1 row and click the Output Summary icon  . You will see Simulation Output summary appearing in the upper pane. It shows that the simulated power as 0.892, indicating that in 892 out of 1000 simulations the boundary was crossed.

| | Sim1 |
|---|--------------------|
| Mnemonic | SU-2S-LRSD |
| Test Parameters | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Power | 0.892 |
| No. of Looks | 6 |
| Test Statistic | Logrank |
| Model Parameters | |
| No. of Hazard Pieces | 1 |
| Hazard Ratio | 0.83 |
| Boundary Parameters | |
| Spacing of Looks | User Specified |
| Efficacy Boundary | User Specified |
| Accrual & Dropout Parameters | |
| Subjects Followed-up | Until End of Study |
| Sample Size | |
| Maximum | 1642 |
| Events | |
| Maximum | 1243 |
| Simulation Results (Overall) | |
| Average Study Duration | 3.703 |
| Average Sample Size | 1641.954 |

You can save Sim1 as a node in the workbook. If you right-click on this node and then click Details, you will see the complete details of simulation appearing in the work

area. Here is a part of it.

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

| Test Parameters | |
|---|---------------------|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 6 |
| Test Type | 2-Sided |
| Sample Size (n) | 1642 |
| Fix at Each Look | Total No. of Events |
| Test Statistic | Logrank |
| Average Events | 900.996 |
| Total Accrual Duration | 1.7 |
| Avg. Power at Termination | 0.892 |
| Response Generation Parameters | |
| HR = λ_1/λ_2 | 0.83 |
| Hazard Rate - Control (λ_1) | 0.38 |
| Hazard Rate - Treatment (λ_2) | 0.315 |
| Accrual / Dropouts Parameters | |
| Sample Size | 1642 |
| Subject are followed | Until End of Study |
| Accrual Duration | 1.7 |
| Accr. % | 100 |
| Dropout Hazard Rate Control | 0.05 |
| Dropout Hazard Rate Treatment | 0.05 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |

☰ Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 1169.589 | 112.127 | 94.873 | 15.022 | 15.155 | 1.2 | 0.515 |
| 2 | 1639.86 | 222.829 | 191.171 | 29.69 | 30.313 | 1.754 | 0.727 |
| 3 | 1642 | 332.042 | 289.958 | 44.124 | 45.839 | 2.329 | 1.092 |
| 4 | 1642 | 436.823 | 392.177 | 58.43 | 61.303 | 3.068 | 1.455 |
| 5 | 1642 | 538.252 | 497.748 | 72.87 | 76.685 | 4.122 | 1.818 |
| 6 | 1642 | 636.59 | 606.41 | 86.971 | 92.014 | 5.974 | 2.184 |
| Average | 1641.954 | 477.412 | 423.584 | 63.16 | 67.312 | 3.703 | 1.583 |

☰ Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Events | Boundaries | | Stopping For | | Total Simulations | |
|--------|--------|------------|--------|----------------|----------------|-------------------|---------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 207 | 5.369 | -5.369 | 0 | 0 | 0 | 0.000% |
| 2 | 414 | 3.712 | -3.712 | 0 | 40 | 40 | 4.000% |
| 3 | 622 | 2.968 | -2.968 | 0 | 214 | 214 | 21.400% |
| 4 | 829 | 2.538 | -2.538 | 0 | 314 | 314 | 31.400% |
| 5 | 1036 | 2.252 | -2.252 | 0 | 222 | 222 | 22.200% |
| 6 | 1243 | 2.045 | -2.045 | 0 | 102 | 210 | 21.000% |
| Total | | | | 0 | 892 | 1000 | |
| % | | | | 0.000% | 89.200% | | |

43.2.10 Interim Monitoring

Click Des1 node under workbook wbk1 and click the **IM** icon. Alternatively, you can right-click the Des1 node and select the item **Interim Monitoring**. In either case, you will see the IM dashboard appearing as shown below.

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of HR | Est. of δ | Std. Error of Est. of δ | Efficacy | | 95% RCI for HR | | 95% RCI for δ | | Repeat p-value | CP | Predictive Power |
|--------|----------------------|-------------------|----------------|------------|------------------|--------------------------------|----------|-------|----------------|-------|----------------------|-------|----------------|----|------------------|
| | | | | | | | Upper | Lower | Upper | Lower | Upper | Lower | | | |
| 1 | | | | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | | | | |

Click the "Enter Interim Data" button to enter the Look # 1 data.

Stopping Boundaries

Events Efficacy Upper Efficacy Lower

Conditional Power

| | |
|-------|-------|
| HR | CP |
| 0.83 | 0.9 |
| 0.846 | 0.829 |
| 0.865 | 0.709 |
| 0.885 | 0.56 |
| 0.906 | 0.403 |
| 0.927 | 0.26 |
| 0.948 | 0.151 |
| 0.97 | 0.082 |
| 0.992 | 0.052 |
| 1 | 0.05 |

Error Spending Function

Info. Fraction α

Conf. Intervals for HR

| | | | | |
|----------------|-----------|-----------|----------------|----------------|
| Info. Fraction | RCI Upper | RCI Lower | Naive CI Upper | Naive CI Lower |
| | | | | |

43 Tutorial: Survival Endpoint

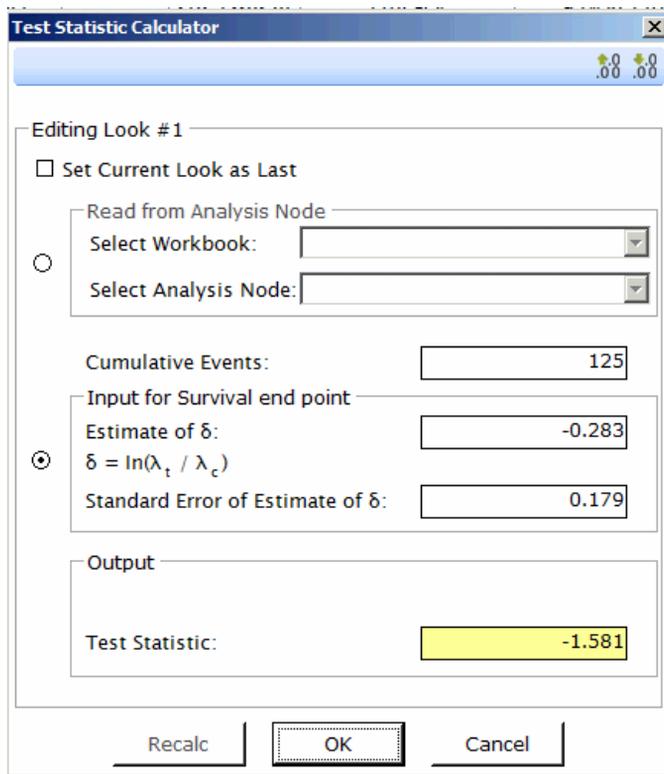
In the top row, you see a few icons. For now, we will discuss only the first icon [Enter Interim Data](#) which represents Test Statistic Calculator. Using this calculator, you will enter the details of interim look data analysis results into the IM dashboard.

Suppose we have the following data used by the Data Monitoring Committee during the first 5 looks of interim monitoring.

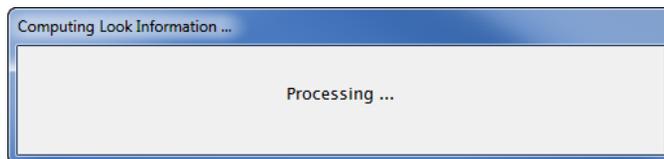
| Date | Total Deaths | $\hat{\delta}$ | SE($\hat{\delta}$) | Z-Statistic |
|--------|--------------|----------------|----------------------|-------------|
| Aug 96 | 125 | -0.283 | 0.179 | -1.581 |
| Mar 97 | 299 | -0.195 | 0.116 | -1.681 |
| Aug 97 | 423 | -0.248 | 0.097 | -2.557 |
| Mar 98 | 545 | -0.259 | 0.086 | -3.012 |
| Aug 98 | 670 | -0.290 | 0.077 | -3.766 |

The first look was taken at 125 events and the analysis of the data showed the value of $\hat{\delta} = -0.283$ and $SE(\hat{\delta}) = 0.179$. First, click the blank row in the IM Dashboard and then click the [Enter Interim Data](#) icon. Now you can enter the first analysis results into the TS calculator and click Recalc. The Test Statistic value will be computed and the TS

calculator will appear as shown below.



Now click on the button 'OK' to get the first look details into IM Dashboard. The following message will appear that some required computations are being carried out.



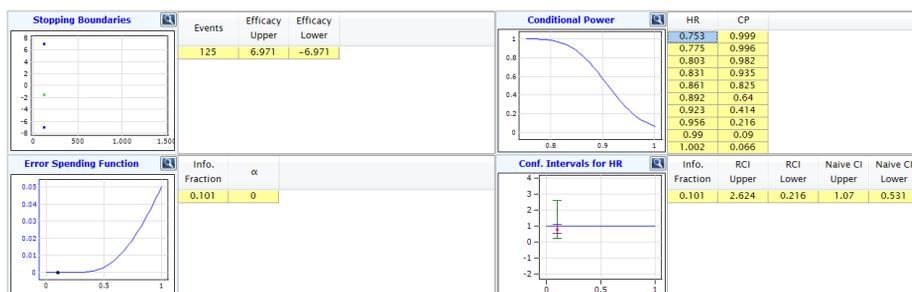
After the computations are over, the output for the first look will appear in the IM

43 Tutorial: Survival Endpoint

Dashboard as shown below.

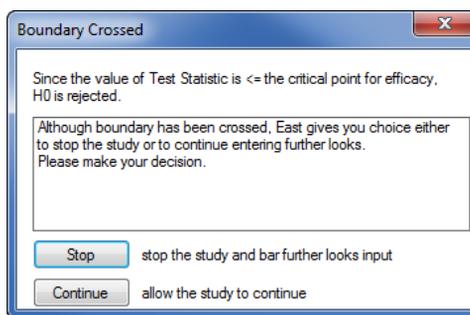
| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of HR | Est. of δ | Std. Error of Est. of δ | Efficacy | | 95% RCI for HR | | 95% RCI for δ | | Repeat... p-value | CP | Predicti... Power |
|--------|----------------------|-------------------|----------------|------------|------------------|--------------------------------|----------|--------|----------------|-------|----------------------|--------|-------------------|-------|-------------------|
| | | | | | | | Upper | Lower | Upper | Lower | Upper | Lower | | | |
| 1 | 0.101 | 125 | -1.581 | 0.754 | -0.283 | 0.179 | 6.971 | -6.971 | 2.624 | 0.216 | 0.965 | -1.531 | 1 | 0.999 | 0.853 |
| 2 | | | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | | | | |

For the first look at total number of events, 125, the Information Fraction works out to be 0.101. The efficacy boundaries for this information fraction are newly computed. The Repeated 95% Confidence Interval limits and Repeated p-value are computed and displayed. You may also see that the charts at the bottom of the IM Dashboard have been updated with relevant details appearing on the side.



In a similar way, enter the interim analysis results for the next 4 looks in the IM Dashboard.

At the fifth look, the boundary is crossed. A message window appears as shown below.



Click Stop and you will see the details of all the looks in the IM Dashboard as shown below.

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of HR | Est. of δ | Std. Error of Est. of δ | Efficacy | | 95% RCI for HR | | 95% RCI for δ | | Repeat... p-value | CP | Predicti... Power |
|--------|----------------------|-------------------|----------------|------------|------------------|--------------------------------|----------|--------|----------------|-------|----------------------|--------|-------------------|-------|-------------------|
| | | | | | | | Upper | Lower | Upper | Lower | Upper | Lower | | | |
| 1 | 0.101 | 125 | -1.581 | 0.754 | -0.283 | 0.179 | 6.971 | -6.971 | 2.624 | 0.216 | 0.965 | -1.531 | 1 | 0.999 | 0.853 |
| 2 | 0.241 | 299 | -1.681 | 0.823 | -0.195 | 0.116 | 4.423 | -4.423 | 1.374 | 0.493 | 0.318 | -0.708 | 0.66 | 0.95 | 0.795 |
| 3 | 0.34 | 423 | -2.557 | 0.78 | -0.248 | 0.097 | 3.672 | -3.672 | 1.114 | 0.547 | 0.108 | -0.604 | 0.212 | 0.999 | 0.962 |
| 4 | 0.438 | 545 | -3.012 | 0.772 | -0.259 | 0.086 | 3.206 | -3.206 | 1.017 | 0.586 | 0.017 | -0.535 | 0.069 | 1 | 0.993 |
| 5 | 0.539 | 670 | -3.766 | 0.748 | -0.29 | 0.077 | 2.872 | -2.872 | 0.933 | 0.6 | -0.069 | -0.511 | 0.008 | NA | NA |

The final Adjusted Inference output also appears as displayed below.

| Final Inference | |
|--|--------|
| Final Outputs at Look # | 5 |
| Adj. p-value | 0.001 |
| Adj. Pt. Est. for HR | 0.767 |
| Adj. 95% CI for HR | |
| Upper Confidence Bound | 0.902 |
| Lower Confidence Bound | 0.655 |
| Adj. Pt. Est. for δ | -0.266 |
| Adj. 95% CI for δ | |
| Upper Confidence Bound | -0.104 |
| Lower Confidence Bound | -0.424 |

One important point to note here is that this study got over almost about 2 years ahead of planned schedule, because of the very favorable interim analysis results.

This completes the Interim Monitoring exercise in this trial.

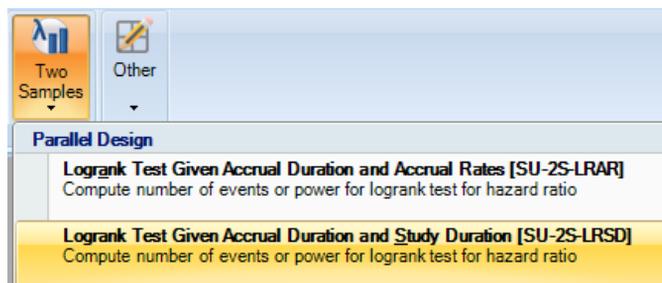
43.3 User Defined R Function

East allows you to customize simulations by inserting user-defined R functions for one or more of the following tasks: generate response, compute test statistic, randomize subjects, generate arrival times, and generate dropout information. The R functionality for arrivals and dropouts will be available only if you have entered such information at the design stage. Although the R functions are also available for all normal and binomial endpoints, we will illustrate this functionality for a time-to-event endpoint. Specifically, we will use an R function to generate Weibull survival responses.

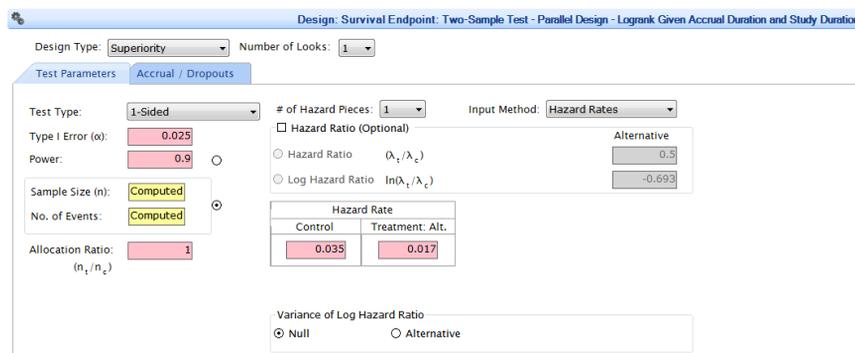
Start East afresh. On the **Design** tab, click **Survival: Two Samples** and then **Logrank**

43 Tutorial: Survival Endpoint

Test Given Accrual Duration and Study Duration.



Choose the design parameters as shown below. In particular, select a one sided test with type-1 error of $\alpha = 0.025$.



Click **Compute** and save this design (Des1) to the **Library**. Right-click Des1 in the **Library** and click **Simulate**. In the **Simulation Control Info** tab, check the box for **Suppress All Intermediate Output**. Type 10000 for **Number of**

Simulations and select **Clock** for **Random Number Seed**.

Test Parameters | Response Generation | Accrual / Dropouts | **Simulation Controls**

Number of Simulations:
 Refresh Frequency:
 Random Number Seed
 Clock
 Fixed
 Suppress All Intermediate Output

Output Options
 Output Type:
 Save summary statistics for every simulation run
 Save subject level data for simulation runs
 Note: Max. 100,000 records will be saved.

In the top right-hand corner for the input window, click **Include Options**, and then click **User Defined R Function**.

Include Options

- Site Info
- Randomization Info
- User Defined R Function
- Stratification Info

Go to the **User Defined R Function** tab. For now, leave the box **Initialize R simulation (optional)** unchecked. This optional task can be used to load required libraries, set seeds for simulations, and initialize global variables.

Select the row for **Generate Response**, click **Browse...**, and navigate to the folder containing your R file. Select the file and click **Open**. The path should now be displayed under **File Name**.

Test Parameters | Response Generation | Accrual / Dropouts | **User Defined R Function** | Simulation Controls

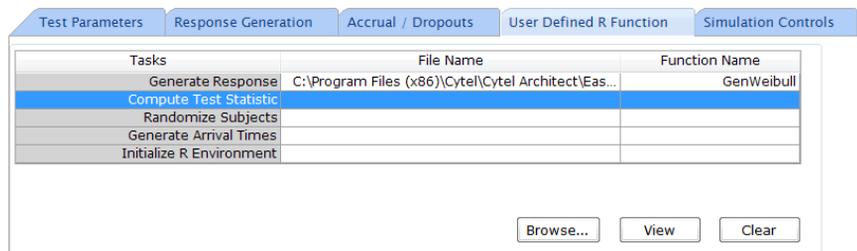
| Tasks | File Name | Function Name |
|--------------------------|---|---------------|
| Generate Response | C:\Program Files (x86)\Cytel\Cytel Architect\Eas... | |
| Compute Test Statistic | | |
| Randomize Subjects | | |
| Generate Arrival Times | | |
| Initialize R Environment | | |

43 Tutorial: Survival Endpoint

Click **View** to open a notepad application to view your R file. In this example, we are generating survival responses for both control and treatment arms from a Weibull with shape parameter = 2 (i.e. exponential), with the same hazard rate in both arms. This sample file is available in the folder named **R Samples** under installation directory of East 6.

```
SurvivalWeibull.r - Notepad
File Edit Format View Help
GenWeibull <- function(NumSub, NumArm, TreatmentID, SurvMethod, NumPrd, PrdTime, SurvParam)
{
  time <- c()
  null.rate <- SurvParam[1,1]
  for(m in 1:NumSub)
  {
    j <- TreatmentID[m]
    time[m] <- rweibull(n=1, shape=2, scale=1 / null.rate)
  }
  return(list(SurvivalTime = as.double(time), ErrorCode = as.integer(0)) )
}
```

Copy the function name (in this case *GenWeibull*) and paste it into the cell for **Function Name**. Save and close the R file, and click **Simulate**.



Return to the tab for **User Defined R Function**, select the **Generate Response** row, and click **View**. In the R function, change the shape parameter = 1, to generate responses from a Weibull distribution with increasing hazards. Save and close the R

file, and click **Simulate**. You may have to save this file on some other location.

```
SurvivalWeibull.r - Notepad
File Edit Format View Help
GenWeibull <- function(NumSub, NumArm, TreatmentID, SurvMethod, NumPrd, PrdTime, SurvParam)
{
  time <- c()
  null.rate <- SurvParam[1,1]
  for(m in 1:NumSub)
  {
    j <- TreatmentID[m]
    time[m] <- rweibull(n=1, shape=1, scale=1 / null.rate)
  }
  # time[m] <- rweibull(n=1, shape=1, scale=1 / SurvParam[1, j+1])
  return(list(SurvivalTime = as.double(time), ErrorCode = as.integer(0)) )
}
```

Select both simulations (Sim1 and Sim2) from the **Output Preview**, and on the toolbar, click  to display in the **Output Summary**.

| | Sim1 | Sim2 |
|-------------------------------------|--------------------|--------------------|
| Mnemonic | SU-2S-LRSD | SU-2S-LRSD |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| Test Type | 1-Sided | 1-Sided |
| Power | 0.027 | 0.026 |
| No. of Looks | 1 | 1 |
| Test Statistic | Logrank | Logrank |
| Model Parameters | | |
| No. of Hazard Pieces | 1 | 1 |
| Hazard Ratio | 0.5 | 0.5 |
| Sample Size | | |
| Maximum | 182 | 182 |
| Events | | |
| Maximum | 88 | 88 |
| Simulation Results (Overall) | | |
| Average Study Duration | 34.624 | 30.708 |
| Average Sample Size | 182 | 182 |
| Other Parameters | | |
| Subjects Followed-up | Until End of Study | Until End of Study |

Notice that the type-1 error appears to be controlled in both cases. When we simulated from the exponential (Sim2), the average study duration (30.7 months) was close to what was calculated at Des1 for the expected study duration under the null. However, when we simulated from the Weibull with decreasing hazards (Sim1), the average study duration increased to 34.6 months.

43 *Tutorial: Survival Endpoint*

The ability to use custom R functions for many simulation tasks allows considerable flexibility in performing sensitivity analyses and assessment of key operating characteristics.

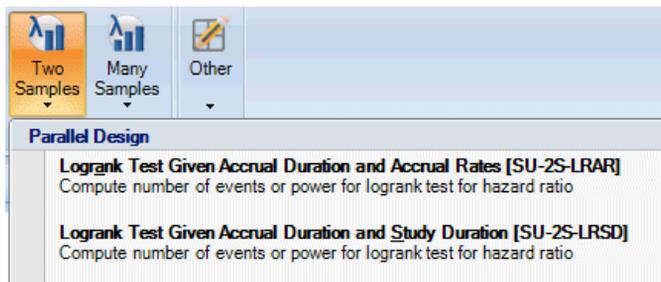
44 Superiority Trials with Variable Follow-Up

This chapter will illustrate through a worked example how to design, monitor and simulate a two-sample superiority trial with a time-to-event trial endpoint. Each subject who has not dropped out or experienced the event is followed until the trial ends. This implies that a subject who is enrolled earlier could potentially be followed for a longer time than a subject who is enrolled later on in the trial. In East we refer to such designs as **variable follow-up designs**.

44.1 The RALES Clinical Trial: Initial Design

The RALES trial (Pitt et al., 1999) was a double blind study of aldosterone-receptor blocker spironolactone at a daily dose of 25 mg in combination with standard doses of an ACE inhibitor (treatment arm) versus standard therapy of an ACE inhibitor (control arm) in patients who had severe heart failure as a result of systolic left ventricular dysfunction. The primary endpoint was death from any cause. Six equally-spaced looks at the data using the Lan-DeMets-O'Brien-Fleming spending function were planned. The trial was designed to detect a hazard ratio of 0.83 with 90% power at a two-sided 0.05 level of significance. The hazard rate of the control arm was estimated to be 0.38/year. The trial was expected to enroll 960 patients/year.

We begin by using East to design RALES under these basic assumptions. Open East, click **Design** tab and then **Two Samples** button in **Survival** group. You will see the following screen.



Note that there are two choices available in the above list; **Logrank Test Given Accrual Duration and Accrual Rates** and **Logrank Test Given Accrual Duration and Study Duration**. The option **Logrank Test Given Accrual Duration and Study Duration** is explained later in Chapter 48. Now click **Logrank Test Given Accrual**

44 Superiority Trials with Variable Follow-Up

Duration and Accrual Rates and you will get the following input dialog box.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

Design Type: Superiority Number of Looks: 1

Test Parameters Accrual / Dropouts

Test Type: 1-Sided # of Hazard Pieces: 1 Input Method: Hazard Rates

Type I Error (α): 0.025

Power: 0.9

No. of Events: Computed

Allocation Ratio: 1
(n_1/n_2)

Hazard Ratio (Optional) Alternative

Hazard Ratio (λ_t/λ_c) 0.5

Log Hazard Ratio $\ln(\lambda_t/\lambda_c)$ -0.693

| Hazard Rate | |
|-------------|-----------------|
| Control | Treatment: Alt. |
| 0.035 | 0.017 |

Variance of Log Hazard Ratio

Null Alternative

Assurance (Probability of Success)

In the above dialog box, enter 6 for **Number of Looks**, keep the default choices of **Design Type: Superiority**, change the **Test Type to 2-Sided**, **Type I Error (α) to 0.05**, **Power : 0.9**, and the **Allocation Ratio: 1**.

Further, keep the default choices of **# of Hazard Pieces** as 1 and the **Input Method:** as **Hazard Rates**. Click the check box against **Hazard Ratio** and enter the **Hazard Ratio** as 0.83. Enter **Hazard Rate (Control)** as 0.38. You will see the **Hazard Rate (Treatment:Alt)** computed as 0.3154. Also, keep the **Variance of Log Hazard Ratio** to be used as under **Null**. Now the **Test Parameters** tab of the input

dialog will appear as shown below.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

Design Type: Superiority Number of Looks: 6

Test Parameters Boundary Accrual / Dropouts

Test Type: 2-Sided # of Hazard Pieces: 1 Input Method: Hazard Rates

Type I Error (α): 0.05
 Power: 0.9
 No. of Events: Computed
 Allocation Ratio: 1
 (n_1/n_2)

Hazard Ratio (Optional) Alternative
 Hazard Ratio (λ_1/λ_2) 0.83
 Log Hazard Ratio $\ln(\lambda_1/\lambda_2)$ -0.186

| Hazard Rate | |
|-------------|-----------------|
| Control | Treatment: Alt. |
| 0.38 | 0.315 |

Variance of Log Hazard Ratio
 Null Alternative

Assurance (Probability of Success)

Now click on the tab **Boundary**. You will see the following input dialog box.

Design Parameters Boundary Info Accrual / Dropout Info

Efficacy
 Boundary Family: Spending Functions
 Spending Function: Lan-DeMets
 Parameter: OF
 Type I Error (α): 0.05

Futility
 Boundary Family: None

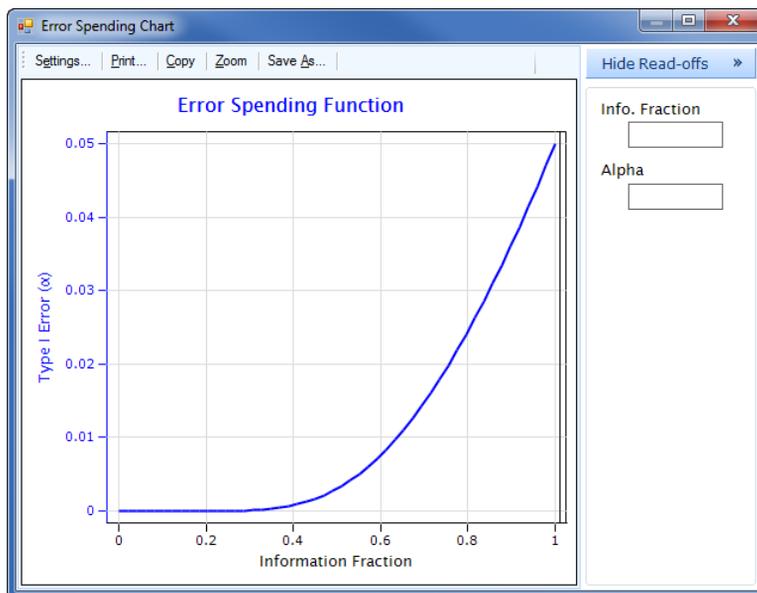
Spacing of Looks Equal Unequal Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.167 | 0.000 | 5.367 | -5.367 |
| 2 | 0.333 | 0.000 | 3.710 | -3.710 |
| 3 | 0.500 | 0.003 | 2.970 | -2.970 |
| 4 | 0.667 | 0.012 | 2.539 | -2.539 |
| 5 | 0.833 | 0.028 | 2.252 | -2.252 |

Keep all the default specifications for the boundaries to be used in the design. You can

44 Superiority Trials with Variable Follow-Up

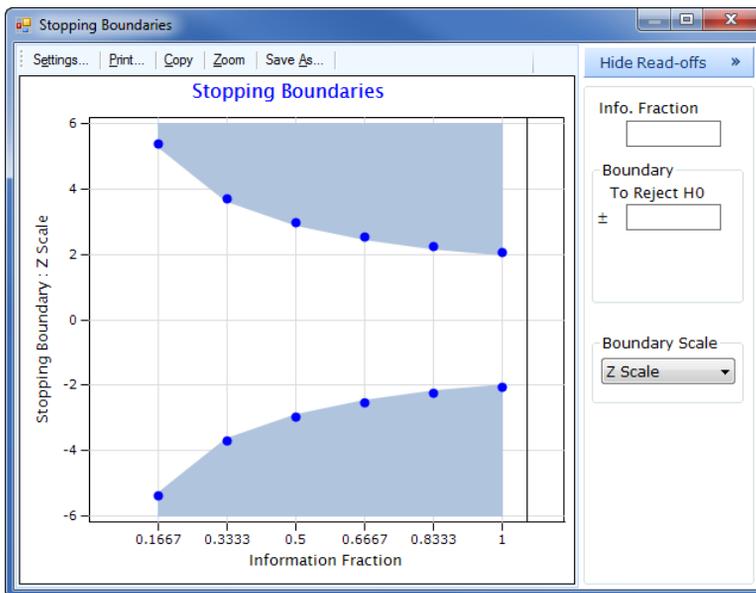
look at the Error Spending Chart by clicking on the icon 



Close this chart.

If you click on the boundary chart icon , you will see the boundary chart as

displayed below.



Close this chart.

Now click **Accrual/Dropouts** tab. Keep the default choice **Until End of Study** for the input **Subjects are followed:**. Keep the **# of Accrual Periods** as 1 and enter 960/year as the accrual rate. For this example, assume no dropouts. The dialog box

44 Superiority Trials with Variable Follow-Up

will look as shown below.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

Design Type: Superiority Number of Looks: 6

Test Parameters Boundary Accrual / Dropouts

Subjects are followed: Until End of Study

Accrual Info

of Accrual Periods: 1

Accrual Rate: 960

Piecewise Dropout Information

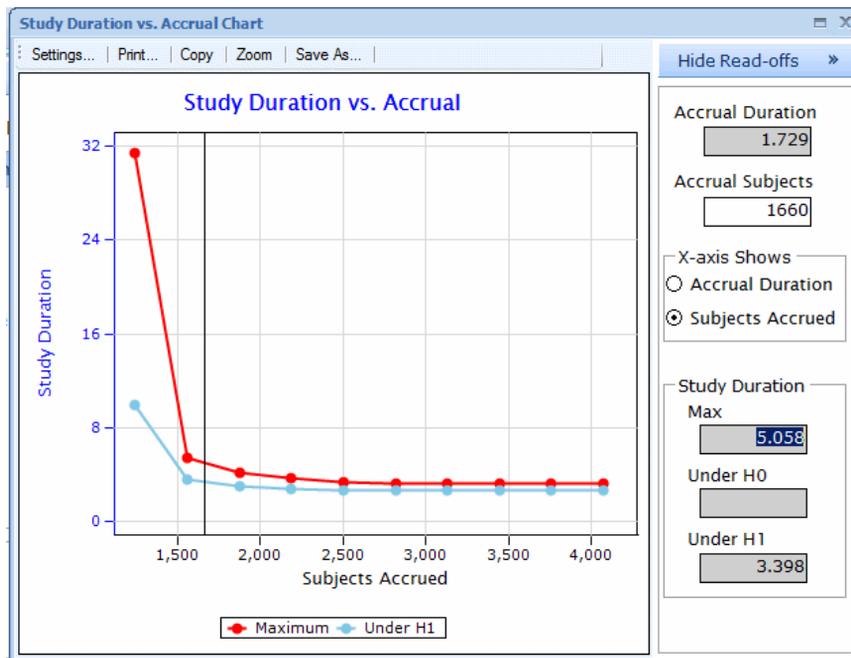
of Pieces: 0

Accrual

| | Min. | Comtd. | Sugg. Max. |
|--|-------|--------|------------|
| <input type="radio"/> Duration: | 1.295 | 2.268 | 3.241 |
| <input checked="" type="radio"/> Subjects: | 1243 | 2177 | 3111 |

Under the **Accrual** section and in column titled **Comtd.** (committed), you see two radio buttons **Durations** and **Subjects** with the latter selected by default. The selected item will appear as the x-axis item in the **Study Duration vs. Accrual** chart, which you can get by clicking on the icon displayed on the side. Against **Durations** and **Subjects** you see two rows of three cells each. The first and third cells will show the min and max values for the row item and the middle cell, mid value between min and max values. From these results, you see that any sample size in the range 1243 to 3111 will suffice to attain the desired 90% power and selects 2177, the mid-point of the allowable range, as the default sample size. Depending on the needs of the study, you may wish to use a different sample size within the allowable range. The choice of sample size generally depends on how long you wish the study to last. The larger you make the patient accrual the shorter will be the total study duration, consisting of accrual time plus follow up time. To understand the essence of this trade-off, bring up

the **Study Duration vs. Accrual** chart by clicking on the icon  .



Based on this chart, a sample size of 1660 subjects is selected. Close the chart and enter 1660 for **Committed Accrual (subjects)**. Click on **Compute** and see the results in the new design created under **Output Preview**. Click the  icon to see the design summary. This sample size ensures that the maximum study duration will be slightly more than 4.9 years. Additionally, under the alternative hypothesis, the

44 Superiority Trials with Variable Follow-Up

expected study duration will be only about 3.3 years.

| Sample Size | |
|-------------------|----------|
| Maximum | 1660 |
| Expected Under H0 | 1659.987 |
| Expected Under H1 | 1659.985 |
| Events | |
| Maximum | 1243 |
| Expected Under H0 | 1233.984 |
| Expected Under H1 | 903.595 |
| Study Duration | |
| Maximum | 4.905 |
| Expected Under H0 | 4.506 |
| Expected Under H1 | 3.337 |
| Accrual Duration | |
| Maximum | 1.729 |
| Expected Under H0 | 1.729 |
| Expected Under H1 | 1.729 |

44.2 Incorporating Drop-Outs

The investigators expect 5% of the patients in both the groups to drop out each year. To incorporate this drop-out rate into the design, in the **Piecewise Constant Dropout Rates** tab, select 1 for the number of pieces and change the Input Method from **Hazard Rates** to **Prob. of Dropout**. Then enter 0.05 as the probability of dropouts at 1 year for both the groups.

Piecewise Dropout Information

of Pieces: Input Method:

By Time:

Prob. of Dropout (Control)

Prob. of Dropout (Treatment)

Note: Period 1 hazard rates apply after time 1.

To make Des1 and Des2 comparable, change the sample size of Des2 to 1660 by

typing this value into the **Committed Accrual (Subjects)** cell. Click on **Compute** and see the results in the new design created under **Output Preview**. Select the two designs and click on  icon to see them side-by-side.

| | Des1 | Des2 |
|---|--------------------|--------------------|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 6 | 6 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Hazard Ratio (Alt.) | 0.83 | 0.83 |
| Var (Log HR) | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Boundary Parameters | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) |
| Accrual & Dropout Parameters | | |
| Accrual Rate | 960 | 960 |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 0 | 1 |
| Sample Size | | |
| Maximum | 1660 | 1660 |
| Expected Under H0 | 1659.987 | 1659.992 |
| Expected Under H1 | 1659.985 | 1659.986 |
| Events | | |
| Maximum | 1243 | 1243 |
| Expected Under H0 | 1233.984 | 1233.984 |
| Expected Under H1 | 903.595 | 903.595 |
| Study Duration | | |
| Maximum | 4.905 | 5.87 |
| Expected Under H0 | 4.506 | 5.258 |
| Expected Under H1 | 3.337 | 3.687 |
| Accrual Duration | | |
| Maximum | 1.729 | 1.729 |
| Expected Under H0 | 1.729 | 1.729 |
| Expected Under H1 | 1.729 | 1.729 |

A comparison of two designs reveals that, because of the drop-outs, the maximum study duration will be prolonged from 4.9 years under Des1 to 5.9 years under Des2. The expected study duration will likewise be prolonged from 3.3 years to 3.7 years under the alternative hypothesis, and from 4.5 years to 5.3 years under the null hypothesis.

44 Superiority Trials with Variable Follow-Up

44.3 Incorporating Non-Constant Accrual Rates

In many clinical trials, the enrollment rate is low in the beginning and reaches its maximum expected level a few months later when all the sites enrolling patients have been recruited. Suppose that patients are expected to enroll at an average rate of 400/year for the first six months and at an average rate of 960/year thereafter. Click on



the **Input** icon on the bottom of your screen to go back to the input window of Des2. Now in **Accrual** section, specify that there are two accrual periods and enter the accrual rate for each period in the dialog box as shown below.

Accrual Info

of Accrual Periods:

| Period # | Starting At | Accrual Rate |
|----------|-------------|--------------|
| 1 | 0 | 400.000 |
| 2 | 0.5 | 960.000 |

Accrual

| | | | | |
|--|------------------------------------|------------------------------------|------------------------------------|--|
| | Min. | Comtd. | Sugg. Max. | |
| <input type="radio"/> Duration: | <input type="text" value="1.779"/> | <input type="text" value="2.021"/> | <input type="text" value="3.618"/> | |
| <input checked="" type="radio"/> Subjects: | <input type="text" value="1428"/> | <input type="text" value="1660"/> | <input type="text" value="3193"/> | |

Once again let the sample size be 1660 to make Des3 comparable to the other two designs. Click on **Compute** to complete the design. Select all the three designs in the

Output Preview area and click on  icon to see them side-by-side.

| | Des1 | Des2 | Des3 |
|---|--------------------|--------------------|--------------------|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR | SU-2S-LRAR |
| Test Parameters | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 6 | 6 | 6 |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.9 | 0.9 | 0.9 |
| Model Parameters | | | |
| Hazard Ratio (Alt.) | 0.83 | 0.83 | 0.83 |
| Var (Log HR) | Null | Null | Null |
| Allocation Ratio (nt / nc) | 1 | 1 | 1 |
| Boundary Parameters | | | |
| Spacing of Looks | Equal | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) |
| Accrual & Dropout Parameters | | | |
| Accrual Rate | 960 | 960 | Multiple |
| Subjects are Followed | Until End of Study | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 | 2 |
| No. of Dropout Pieces | 0 | 1 | 1 |
| Sample Size | | | |
| Maximum | 1660 | 1660 | 1660 |
| Expected Under H0 | 1659.987 | 1659.992 | 1659.99 |
| Expected Under H1 | 1659.985 | 1659.986 | 1659.985 |
| Events | | | |
| Maximum | 1243 | 1243 | 1243 |
| Expected Under H0 | 1233.984 | 1233.984 | 1233.984 |
| Expected Under H1 | 903.595 | 903.595 | 903.595 |
| Study Duration | | | |
| Maximum | 4.905 | 5.87 | 6.15 |
| Expected Under H0 | 4.506 | 5.258 | 5.538 |
| Expected Under H1 | 3.337 | 3.687 | 3.966 |
| Accrual Duration | | | |
| Maximum | 1.729 | 1.729 | 2.021 |
| Expected Under H0 | 1.729 | 1.729 | 2.021 |
| Expected Under H1 | 1.729 | 1.729 | 2.021 |

Notice that the enrollment period has increased from 1.7 years to 2 years. Likewise, the maximum study duration and the expected study durations under H_0 and H_1 have also increased relative to Designs 1 and 2. Now the maximum study duration is 6.15 years.

44.4 Incorporating Piecewise Constant Hazards

Prior studies had suggested that the survival curves might not follow an exponential distribution. Suppose it is believed that the hazard rate for failure on the control arm decreases after the first 12 months from 0.38 to 0.35. We will assume that the hazard ratio is still 0.83. We can enter the appropriate piecewise hazard rates into East as

follows. Click on  icon on the bottom of your screen to go back to

44 Superiority Trials with Variable Follow-Up

the input window and go to **Test Parameters** tab.

of Hazard Pieces: Input Method:

Hazard Ratio

Hazard Ratio (λ_t / λ_c) Alternative

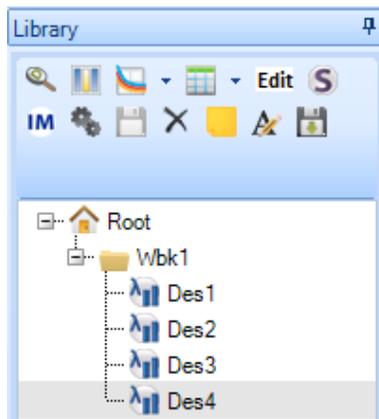
Log Hazard Ratio $\ln(\lambda_t / \lambda_c)$

| Period # | Starting at Time | Hazard Rate (Control) | Hazard Rate (Treatment: Alt.) |
|----------|------------------|-----------------------|-------------------------------|
| 1 | 0.000 | 0.38 | 0.315 |
| 2 | 1.000 | 0.35 | 0.291 |

Change the sample size to 1660 on **Accrual/Dropouts** tab for comparability with the previous designs. Click on **Compute** and see the results of the design in the **Output Summary** mode.

| | Des1 | Des2 | Des3 | Des4 |
|---|--------------------|--------------------|--------------------|--------------------|
| Test Parameters | | | | |
| Design Type | Superiority | Superiority | Superiority | Superiority |
| No. of Looks | 6 | 6 | 6 | 6 |
| Test Type | 2-Sided | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 | 0.05 |
| Power | 0.9 | 0.9 | 0.9 | 0.9 |
| Model Parameters | | | | |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 | 1 |
| Hazard Ratio (Alt.) | 0.83 | 0.83 | 0.83 | 0.83 |
| Var (Log HR) | Null | Null | Null | Null |
| Boundary Parameters | | | | |
| Spacing of Looks | Equal | Equal | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) | LD (OF) |
| Accrual & Dropout Parameters | | | | |
| Accrual Rate | 960 | 960 | Multiple | Multiple |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 | 1 | 2 | 2 |
| No. of Dropout Pieces | 0 | 1 | 1 | 1 |
| Sample Size | | | | |
| Maximum | 1660 | 1660 | 1660 | 1660 |
| Expected Under H0 | 1659.987 | 1659.992 | 1659.99 | 1659.991 |
| Expected Under H1 | 1659.985 | 1659.986 | 1659.985 | 1659.985 |
| Events | | | | |
| Maximum | 1243 | 1243 | 1243 | 1243 |
| Expected Under H0 | 1233.984 | 1233.984 | 1233.984 | 1233.984 |
| Expected Under H1 | 903.595 | 903.595 | 903.595 | 903.595 |
| Study Duration | | | | |
| Maximum | 4.905 | 5.87 | 6.15 | 6.555 |
| Expected Under H0 | 4.506 | 5.258 | 5.538 | 5.868 |
| Expected Under H1 | 3.337 | 3.687 | 3.966 | 4.136 |
| Accrual Duration | | | | |
| Maximum | 1.729 | 1.729 | 2.021 | 2.021 |
| Expected Under H0 | 1.729 | 1.729 | 2.021 | 2.021 |
| Expected Under H1 | 1.729 | 1.729 | 2.021 | 2.021 |

We observe that the impact of changing from a constant hazard rate to a piecewise constant hazard rate is substantial. The maximum study duration has increased from 6.15 years for Des3 to 6.56 years for Des4. Before proceeding further, save all the four designs in the workbook.



44.5 Simulating a Trial with Proportional Hazards

44.5.1 Simulation Worksheet

44.5.2 Simulating Under H_1

44.5.3 Simulating...

It would be useful to verify the operating characteristics of the various designs created in the previous section by simulation. The new survival simulation capabilities in East permit this. Let us use these capabilities to simulate Des4. Save this design in the workbook. Right-click on this design node and select the menu item **Simulate**. You'll see the following **Survival Simulation** worksheet.

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

Number of Looks: 6

Test Parameters | Response Generation | Accrual / Dropouts | Simulation Controls

Trial Type: Superiority
 Test Type: 2-Sided
 Fix at Each Look: Total No. of Events
 Total No. of Events: 1243

Test Statistic: Logrank

| Look # | Info. Fraction | Cum. α Spent | | Efficacy Z | |
|--------|----------------|--------------|-------|------------|--------|
| | | Upper | Lower | Upper | Lower |
| 1 | 0.167 | 0.000 | 0.000 | 5.369 | -5.369 |
| 2 | 0.333 | 0.000 | 0.000 | 3.712 | -3.712 |
| 3 | 0.500 | 0.002 | 0.002 | 2.968 | -2.968 |
| 4 | 0.667 | 0.006 | 0.006 | 2.538 | -2.538 |
| 5 | 0.833 | 0.014 | 0.014 | 2.252 | -2.252 |

44.5.1 Components of the Simulation Worksheet

This simulation worksheet consists four tabs - **Test Parameters**, **Response Generation**, **Accrual/Dropouts**, and **Simulation Controls**. The **Test Parameters** tab displays all the parameters of the simulation. If desired, you may modify one or more

44 Superiority Trials with Variable Follow-Up

of these parameter values before carrying out simulation. The second tab **Response Generation** will appear as shown below.

Survival Information

of Hazard Pieces: 2 Input Method: Hazard Rates

Hazard Ratio

| Piece | Starting at Time | Hazard Rates | | Hazard Ratio |
|-------|------------------|--------------|-----------|--------------|
| | | Control | Treatment | |
| 1 | 0.000 | 0.380 | 0.315 | 0.830 |
| 2 | 1.000 | 0.350 | 0.291 | 0.830 |

In this tab, you may modify values of response parameters before carrying out simulation. The third tab **Accrual/Dropouts** will display information relating to accrual and dropouts.

Sample Size: 1660
 Total No. of Events: 1243
 Subjects are followed: Until End of Study

Distribution of Accrual Time: Uniform

Accrual Info

of Accrual Periods: 2 Input Method: Accrual Rates

| Period # | Starting at Time | Accrual Rate |
|----------|------------------|--------------|
| 1 | 0 | 400.000 |
| 2 | 0.5 | 960.000 |

Piecewise Dropout Information

of Pieces: 1 Input Method: Prob. of Dropout

By Time: 1

Probability of Dropout - Control: 0.05
 Probability of Dropout - Treatment: 0.05

Note: Period 1 hazard rates apply after time 1.

As in the case of other tabs, you may modify one or more values appearing in this tab before simulation is carried out.

In the **Simulation Controls**, you may specify the simulation parameters like

number of simulations required and the desired simulation seed etc.

The screenshot shows the 'Simulation Controls' tab of a software interface. It contains several input fields and checkboxes. On the left, there are fields for 'Number of Simulations' (10000) and 'Refresh Frequency' (1000). Below these is a 'Random Number Seed' section with radio buttons for 'Clock' (selected) and 'Fixed' (100). Further down are checkboxes for 'Suppress All Intermediate Output', 'Pause after Refresh', and 'Pause at End' (checked). On the right, there is an 'Output Options' section with a dropdown for 'Output Type' (Case Data) and two checkboxes: 'Save summary statistics for every simulation run' and 'Save subject level data for' (1) simulation runs. A note below states 'Note: Max. 100,000 records will be saved.'

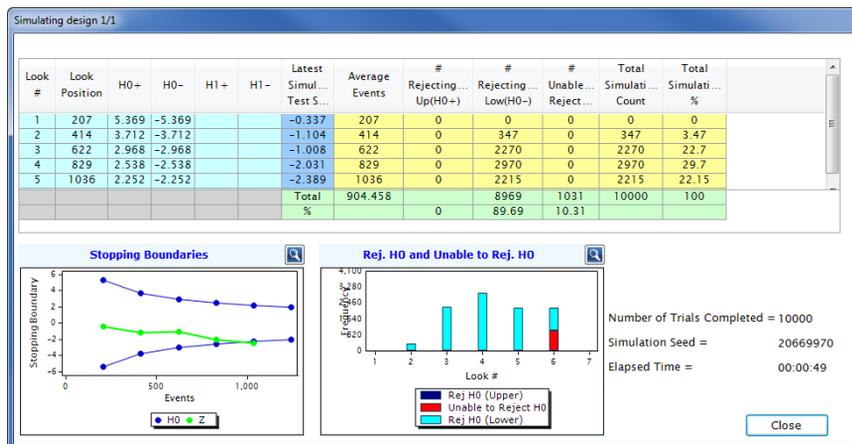
Also optionally, you may bring out one more tab **Randomization** by clicking on **Include Options** button on the right hand top corner. In the **Randomization**, you may alter the allocation ratio of the design before carrying out simulation. The other tabs under the **Include Options** will be discussed elsewhere in this manual.

The screenshot shows the 'Randomization' tab of the software interface. It features a dropdown menu for 'Randomization Method' set to 'Complete Randomized Design' and a text input field for 'Allocation Ratio (n₁/n_c):' with the value '1' entered.

Keeping all the default parameter values same as in the different tabs, click **Simulate**. You can see the progress of the simulation process summarized as shown in the following screen shot. The complete summary of simulations will be displayed

44 Superiority Trials with Variable Follow-Up

at the end of simulations.



Close this window. The simulation results appear in a row in the **Output Preview** as shown below.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Accrual Rate | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events | Average Study Duration | Average Events |
|------|-------------|--------------|-----------|--------------------|-------|-------|------------------|-------------------|--------------|-------------|------------------|------------------|----------------|------------------------|----------------|
| Sim1 | Superiority | 6 | 2-Sided | | 0.897 | 1 | User Specified | User Specified | Multiple | 1660 | | | 1243 | 4.136 | 904.458 |

The output summary can be seen by clicking on the icon  after selecting the

simulation row in the Output Preview.

| | Sim1 |
|---|--------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Power | 0.897 |
| No. of Looks | 6 |
| Test Statistic | Logrank |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| No. of Hazard Pieces | 2 |
| Boundary Parameters | |
| Spacing of Looks | User Specified |
| Efficacy Boundary | User Specified |
| Accrual & Dropout Parameters | |
| Subjects Followed-up | Until End of Study |
| Accrual Rate | Multiple |
| No. of Accrual Periods | 2 |
| Sample Size | |
| Maximum | 1660 |
| Events | |
| Maximum | 1243 |
| Simulation Results (Overall) | |
| Average Study Duration | 4.166 |
| Average Sample Size | 1659.759 |

Now save the simulation results to the workbook by selecting the simulation results row and then clicking on . On this newly added workbook node for simulation, right-click and select **Details**. You will see the complete details simulation

44 Superiority Trials with Variable Follow-Up

appearing on the output pane. The core part is shown below.

☰ Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 1137.768 | 111.854 | 95.146 | 15.136 | 15.493 | 1.476 | 0.525 |
| 2 | 1654.199 | 222.647 | 191.353 | 30.446 | 31.447 | 2.05 | 0.729 |
| 3 | 1660 | 332.113 | 289.887 | 45.977 | 47.99 | 2.641 | 1.104 |
| 4 | 1660 | 436.841 | 392.159 | 61.915 | 65.1 | 3.428 | 1.491 |
| 5 | 1660 | 538.522 | 497.478 | 77.842 | 82.436 | 4.555 | 1.882 |
| 6 | 1660 | 637.603 | 605.397 | 93.507 | 99.86 | 6.537 | 2.275 |
| Average | 1659.759 | 481.734 | 427.9 | 67.754 | 72.454 | 4.166 | 1.646 |

☰ Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Events | Boundaries | | Stopping For | | Total Simulations | |
|--------|--------|------------|--------|----------------|----------------|-------------------|---------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 207 | 5.369 | -5.369 | 0 | 1 | 1 | 0.010% |
| 2 | 414 | 3.712 | -3.712 | 0 | 335 | 335 | 3.350% |
| 3 | 622 | 2.968 | -2.968 | 0 | 2194 | 2194 | 21.940% |
| 4 | 829 | 2.538 | -2.538 | 0 | 2964 | 2964 | 29.640% |
| 5 | 1036 | 2.252 | -2.252 | 0 | 2248 | 2248 | 22.480% |
| 6 | 1243 | 2.045 | -2.045 | 0 | 1224 | 2258 | 22.580% |
| Total | | | | 0 | 8966 | 10000 | |
| % | | | | 0.000% | 89.660% | | |

44.5.2 Simulating Under H_1

Notice that in the above simulations, we did not change anything on the **Response Generation** tab which indicates that we executed 10000 simulations under the designs assumptions or in other words, under alternative hypothesis.

Let us examine these 10000 simulations more closely. The actual values may differ from the manual, depending on the starting seed used.

The column labeled **Events** in the second table, displays the number of events after which each interim look was taken. The column labeled **Avg. Look Time** in the first table, displays the average calendar times at which each interim look was taken. Thus, the first interim look (taken after observing 207 events) occurred after an average elapse of about 1.5 years; the second interim look (taken after observing 414 events) occurred after an average elapse of about 2.1 years; and so on. The remaining columns of the simulation output are self-explanatory. The columns labeled **Stopping For** show that 8966 of the 10000 simulations crossed the lower stopping boundary, thus confirming (up to Monte Carlo accuracy) that this design has 90% power. The detailed output tables also show how the events, drop-outs, accruals, and average follow-up times were observed at each interim analysis.

44.5.3 Simulating Under H_0

To simulate under the null hypothesis we must go back to the input window of Sim1 and then to the **Response Generation** tab. In this pane change the hazard rate for the treatment arm to 0.38 for the first piece and to 0.35 for the second piece of the hazard function.

| Survival Information | | | | |
|---------------------------------------|-------------|----------------------------|-----------|--------------|
| # of Hazard Pieces | 2 | Input Method: Hazard Rates | | |
| <input type="checkbox"/> Hazard Ratio | | | | |
| Piece | Starting At | Hazard Rates | | Hazard Ratio |
| | | Control | Treatment | |
| 1 | 0.000 | 0.380 | 0.38 | 1.000 |
| 2 | 1.000 | 0.350 | 0.35 | 1.000 |

This change implies that we will be simulating under the null hypothesis. Click on the **Simulate** button. A new row in Output Preview will be added now. Select this row and add to the library node. By double-clicking on this node, you will see the detailed simulation output as shown below. The results are displayed below.

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

| Test Parameters | |
|--------------------------------------|----------------------------|
| Simulation ID | Sim2 |
| Design Type | Superiority |
| Number of Looks | 6 |
| Test Type | 2-Sided |
| Sample Size (n) | 1660 |
| Fix at Each Look | Total No. of Events |
| Test Statistic | Logrank |
| Average Events | 1233.685 |
| Total Accrual Duration | 2.021 |
| Avg. Power at Termination | 0.052 |
| Accrual / Dropouts Parameters | |
| Prob. of Dropout by Time = | 1 |
| Control | 0.05 |
| Treatment | 0.05 |
| Randomization Parameters | |
| Method | Complete Randomized Design |
| Allocation Ratio (n/n _c) | 1 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 1089.173 | 103.447 | 103.553 | 14.027 | 14.065 | 1.426 | 0.501 |
| 2 | 1615.584 | 206.873 | 207.127 | 28.239 | 28.372 | 1.977 | 0.682 |
| 3 | 1660 | 310.907 | 311.093 | 42.844 | 42.898 | 2.511 | 1.006 |
| 4 | 1660 | 414.403 | 414.597 | 57.851 | 57.849 | 3.217 | 1.358 |
| 5 | 1660 | 517.905 | 518.095 | 72.995 | 73.061 | 4.222 | 1.714 |
| 6 | 1660 | 621.518 | 621.482 | 88.051 | 88.218 | 5.93 | 2.07 |
| Average | 1659.981 | 616.777 | 616.908 | 87.388 | 87.511 | 5.865 | 2.054 |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Events | Boundaries Efficacy | | Stopping For | | Total Simulations | |
|--------|--------|---------------------|--------|----------------|----------------|-------------------|---------|
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| | | | | | | | |
| 2 | 414 | 3.712 | -3.712 | 1 | 2 | 3 | 0.030% |
| 3 | 622 | 2.968 | -2.968 | 15 | 8 | 23 | 0.230% |
| 4 | 829 | 2.538 | -2.538 | 61 | 49 | 110 | 1.100% |
| 5 | 1036 | 2.252 | -2.252 | 80 | 69 | 149 | 1.490% |
| 6 | 1243 | 2.045 | -2.045 | 112 | 120 | 9715 | 97.150% |
| Total | | | | 269 | 249 | 10000 | |
| % | | | | 2.690% | 2.480% | | |

Out of 10000 simulated trials only 27 crossed the upper stopping boundary and 25 crossed the lower stopping boundary thus confirming (up to Monte Carlo accuracy) that the type-1 error is preserved for this design.

44 Superiority Trials with Variable Follow-Up

44.6 Simulating a Trial with Non-Proportional Hazards

44.6.1 Single-Look Design

44.6.2 Single-Look Design

44.6.3 Group Seq. Design

A new agent is to be tested against placebo in a large cardiovascular study with the endpoint being time to stroke, MI or death. The control arm has a 12-month event-free rate of 97%. We wish to design the study to detect a hazard ratio of 0.75 with 90% power, using a two-sided test conducted at the 0.05 level. An important design consideration is that treatment differences are expected to emerge only after one year of therapy. Subjects will enroll at the rate of 1000/month and be followed to the end of the study. The dropout rate is expected to be 10% per year for both treatment arms. Finally, the study should be designed for maximum study duration of 50 months.

The usual design options in East are not directly applicable to this trial because they require the hazard ratio to be constant under the alternative hypothesis. Here, however, we are required to power the trial to detect a hazard ratio of 0.75 that only emerges after patients have been on the study for 12 months. The simulation capabilities of East can help us with the design.

44.6.1 Single-Look Design with Proportional Hazards

We begin by creating a single-look design powered to detect hazard ratio of 0.75, ignoring the fact that the two survival curves separate out only after 12 months. Open a new survival design worksheet by clicking on **Design>Survival>Logrank Test Given Accrual Duration and Accrual Rates**. In the resulting **Test Parameters** tab, enter the parameters values as shown below.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

Design Type: Superiority Number of Looks: 1

Test Parameters Accrual / Dropouts

Test Type: 2-Sided # of Hazard Pieces: 1 Input Method: Cum. % Survival

Type I Error (α): 0.05

Power: 0.9

No. of Events: Computed

Allocation Ratio: 1 (n_1/n_2)

Hazard Ratio (Optional) Alternative

Hazard Ratio (λ_1/λ_2) 0.75

Ratio of % Survivals at Period # 1 (S_1/S_2) 1.008

| By Time | Cum. % Survival | |
|---------|-----------------|-----------------|
| | Control | Treatment: Alt. |
| 12 | 97 | 97.741 |

Note: Period 1 hazard rates apply after time 12.

Variance of Log Hazard Ratio

Null Alternative

Assurance (Probability of Success)

Click on the tab **Accrual/Dropouts** and enter the values as shown below,

excluding the Accrual tab.

Design Type: Superiority Number of Looks: 1

Test Parameters | **Accrual / Dropouts**

Subjects are followed: Until End of Study

Accrual Info

of Accrual Periods: 1

Accrual Rate: 1000

Piecewise Dropout Information

of Pieces: 1 Input Method: Prob. of Dropout

By Time: 12

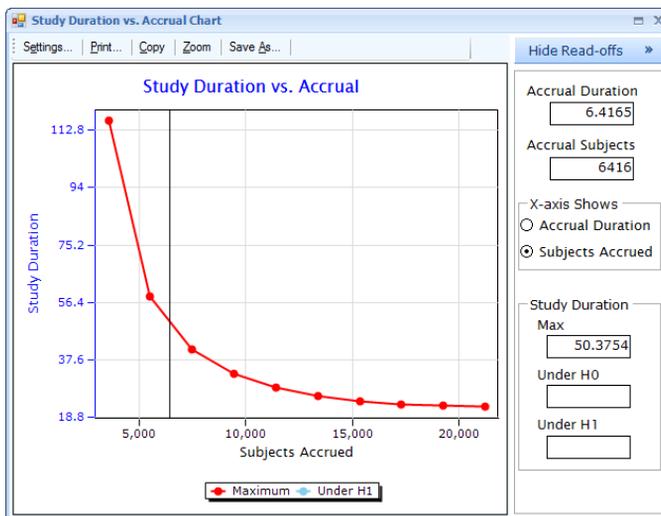
Prob. of Dropout (Control): 0.1

Prob. of Dropout (Treatment): 0.1

Note: Period 1 hazard rates apply after time 12.

| Accrual | Min. | Comtd. | Sugg. Max. |
|--|-------|--------|------------|
| <input type="radio"/> Duration: | 2.524 | 12.392 | 22.26 |
| <input checked="" type="radio"/> Subjects: | 2524 | 12392 | 22260 |

East informs you in the Accrual tab, that any sample size in the range 2524 to 22260 will suffice to attain the desired 90% power. However, the study will end sooner if we enroll more patients. Recall that we wish the trial to last no more than 50 months, inclusive of accrual and follow-up. The **Accrual-Duration** chart can provide guidance on sample size selection. This chart reveals that if 6400 subjects are enrolled, the expected maximum duration of a trial is close to 50 months.



44 Superiority Trials with Variable Follow-Up

Now change the **Comtd.** number of subjects to 6400 and click on **Compute** to complete the design. A new row is added for this design in the Output Preview. Select this row and add it to a library node under a workbook. Now you double-click on this node, you will see the detailed output. A section of it is shown below:

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

| Test Parameters | |
|--|-------------|
| Design ID | Des5 |
| Design Type | Superiority |
| Number of Looks | 1 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| HR = λ_1/λ_2 | |
| Under H0 | 1 |
| Under H1 | 0.75 |
| Cum. % Surv. by Time = | 12 |
| Control (S _c) | 97 |
| Treatment (S _t) | 97.741 |
| Var (Log HR) | Null |
| Allocation Ratio (n ₁ /n ₂) | 1 |
| Accrual / Dropouts Parameters | |
| Accrual Rate | 1000 |
| Dropout | Yes |
| Prob. of Dropout by Time = | 12 |
| Control | 0.1 |
| Treatment | 0.1 |

| Sample Size Information | |
|-----------------------------|------|
| Sample Size (n) | 6400 |
| Treatment (n _t) | 3201 |
| Control (n _c) | 3199 |
| Events (s) | 508 |
| Treatment (s _t) | 220 |
| Control (s _c) | 288 |
| Dropouts (d) | 2010 |
| Treatment (d _t) | 1012 |
| Control (d _c) | 998 |
| Information (I) | 127 |

| Accrual and Study Duration | |
|----------------------------|--------|
| Accrual Duration | 6.4 |
| Max. Study Duration | 48.677 |

| Critical Points | |
|----------------------|------|
| Lower Critical Point | 1.96 |
| Upper Critical Point | 1.96 |

Variable Follow-Up Design: All subjects are followed until failure, drop out or end of study.

We can verify the operating characteristics of Des1 by simulation. With the cursor on Des1 node, Click on Simulation icon from the library menu bar. You'll be taken to the survival simulation worksheet. In the **Simulation Control** tab, specify the number of simulations to be 1000. Now click on **Simulate** button. This will generate 1000 simulations from the survival curves specified in the design. Each simulation will consist of survival data on 6400 subjects entering the trial uniformly at the rate of 1000/month. Events (failures) will be tracked and the simulated trial will be terminated when the total number of events equals 508. Subjects surviving past this termination time point will have their survival times censored. The resulting survival data will be summarized in terms of the logrank test statistic. Each simulation records two important quantities:

- the calendar time at which the last of the specified 508 events arrived;
- whether or not the logrank test statistic rejected the null hypothesis.

We would expect that, on average, the 508 events will occur in about 48.7 months and about 90% of the simulations will reject the null hypothesis. The simulation summary

is shown in the following screen shot.

| | Sim3 |
|---|--------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Power | 0.913 |
| No. of Looks | 1 |
| Test Statistic | Logrank |
| Model Parameters | |
| No. of Hazard Pieces | 1 |
| Hazard Ratio | 0.75 |
| Accrual & Dropout Parameters | |
| Subjects Followed-up | Until End of Study |
| Accrual Rate | 1000 |
| Sample Size | |
| Maximum | 6400 |
| Events | |
| Maximum | 508 |
| Simulation Results (Overall) | |
| Average Study Duration | 48.691 |
| Average Sample Size | 6400 |

Indeed we observe that the average study duration for this set of 1000 simulations was 48.691 months, and that 913 of the 1000 simulated trials crossed the critical value and rejected H_0 and hence the power attained is 0.913. This serves as an independent verification of the operating characteristics of Des1, up to Monte Carlo accuracy.

44.6.2 Single-Look Design with Non-Proportional Hazards

Were it not for the fact that the hazard ratio of 0.75 only emerges after 12 months of therapy, Des1 would meet the goals of this study. However, the impact of the late separation of the survival curves must be taken into consideration. This is accomplished, once again, by simulation. Click the Edit Simulation icon while the cursor is on the last simulation node. In the resulting simulation sheet click on **Response Generation** tab. In this tab, specify that the hazard rates for the control and treatment arms are identical and equal to 0.0025 for the first 12 months and the hazard ratio is 0.75 thereafter. This is done by making appropriate entries in this

44 Superiority Trials with Variable Follow-Up

tab as shown below.

| Survival Information | | | | |
|---------------------------------------|------------------|----------------------------|-----------|--------------|
| # of Hazard Pieces | 2 | Input Method: Hazard Rates | | |
| <input type="checkbox"/> Hazard Ratio | | | | |
| Piece | Starting at Time | Hazard Rates | | Hazard Ratio |
| | | Control | Treatment | |
| 1 | 0.000 | 0.0025 | 0.0025 | 1.015 |
| 2 | 12.000 | 0.0025 | 0.0019 | 0.75 |

Click on the **Simulate** button. This will generate 10000 simulations from survival curves specified in the **Survival Parameters Pane**. As before, each simulation will consist of survival data on 6400 subjects entering the trial uniformly at the rate of 1000/month. Events (failures) will be tracked and the simulated trial will be terminated when the total number of events equals 508. The summary output of this simulation

run as shown below.

| | |
|---|--------------------|
| | Sim2 |
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Power | 0.522 |
| No. of Looks | 1 |
| Test Statistic | Logrank |
| Model Parameters | |
| No. of Hazard Pieces | 2 |
| Accrual & Dropout Parameters | |
| Followup Duration | Until End of Study |
| Accrual Rate | 1000 |
| No. of Accrual Periods | 1 |
| Sample Size | |
| Maximum | 6400 |
| Events | |
| Maximum | 508 |
| Simulation Results (Overall) | |
| Average Study Duration | 46.962 |
| Average Sample Size | 6400 |
| Average Events | 508 |

This time only 522 of the 1000 trials were able to reject H_0 . The drop in power is of course due to the fact that the two survival curves do not separate out until 12 months have elapsed. Thus events that arise within the first 12 months arrive at the same rate for both arms and are not very informative about treatment differences.

We need to increase the power of the study to 90%. This can be accomplished in one of two ways:

1. Prolonging the study duration until a sufficient number of events are obtained to achieve 90% power.
2. Increasing the sample size.

The first approach cannot be used because the study duration is not permitted to exceed 50 months. The simulations have shown that the study duration is already almost 50 months, and it has only achieved 56.5% power. Thus we must resort to increasing the sample size.

44 Superiority Trials with Variable Follow-Up

Now if we increase the sample size while keeping the total number of events fixed at 508, the average study duration will drop. The power, however, may not increase. In fact it might even decrease since a larger fraction of the 508 events will arise in the first 12 months, before the two survival curves have separated. To see this, increase the sample size from 6400 to 10000 in the **Accrual/Dropouts** tab. Then click on **Simulate** button. From this simulation run, you will get the output summary as shown below.

| | Sim3 |
|---|--------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Power | 0.261 |
| No. of Looks | 1 |
| Test Statistic | Logrank |
| Model Parameters | |
| No. of Hazard Pieces | 2 |
| Accrual & Dropout Parameters | |
| Followup Duration | Until End of Study |
| Accrual Rate | 1000 |
| No. of Accrual Periods | 1 |
| Sample Size | |
| Maximum | 10000 |
| Events | |
| Maximum | 508 |
| Simulation Results (Overall) | |
| Average Study Duration | 29.668 |
| Average Sample Size | 10000 |
| Average Events | 508 |

Notice that the average study duration has dropped to 29.7 months. But the power has dropped also. This time only 261 of the 10000 simulations could reject the null hypothesis.

To increase power we must increase sample size while keeping the study duration fixed at about 50 months. This is accomplished by selecting the **Look Time** option from the drop-down box in the **Fix at Each Look** section of the **Survival Parameters Pane** and choosing a 50 month Total Study Durn., while keeping the

sample size increase from 6400 to 10000.

The screenshot shows a software interface with two tabs: "Test Parameters" and "Response Generation". The "Response Generation" tab is active. Below the tabs, there are four settings:

- Number of Looks: 1 (dropdown menu)
- Trial Type: Superiority (dropdown menu)
- Test Type: 2-Sided (dropdown menu)
- Fix at Each Look: Look Time (dropdown menu)
- Study Duration: 50 (text input field)

We will now run 10000 simulations in each of which 10000 subjects are enrolled at the rate of 1000/year. Each simulated trial will be terminated at the end of 50 months of calendar time and a logrank test statistic will be derived from the data. Click on the **Simulate** button. Add the simulation run output to library node and see the

44 Superiority Trials with Variable Follow-Up

following output summary.

| | Sim4 |
|---|--------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Power | 0.735 |
| No. of Looks | 1 |
| Test Statistic | Logrank |
| Model Parameters | |
| No. of Hazard Pieces | 2 |
| Accrual & Dropout Parameters | |
| Followup Duration | Until End of Study |
| Accrual Rate | 1000 |
| No. of Accrual Periods | 1 |
| Sample Size | |
| Maximum | 10000 |
| Events | |
| Maximum | 810.638 |
| Simulation Results (Overall) | |
| Average Study Duration | 50 |
| Average Sample Size | 10000 |
| Average Events | 810.638 |

For more details, you can click  icon after selecting the saved simulation node.

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

| Simulation Parameters | |
|-------------------------------|-------------|
| Simulation ID | Sim4 |
| Design Type | Superiority |
| Number of Looks | 1 |
| Test Type | 2-Sided |
| Sample Size (n) | 10000 |
| Fix at Each Look | Look Time |
| Test Statistic | Logrank |
| Average Events | 810.638 |
| Total Accrual Duration | 10 |
| Avg. Power at Termination | 0.735 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

☞ Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 10000 | 440.811 | 369.827 | 1547.908 | 1557.214 | 50 | 35.373 |
| Average | 10000 | 440.811 | 369.827 | 1547.908 | 1557.214 | 50 | 35.373 |

☞ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Look Time | Boundaries Efficacy | | Stopping For | | Total Simulations | |
|--------|-----------|---------------------|-------|----------------|----------------|-------------------|---|
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| | | | | | | | |
| Total | | | | 0 | 7345 | 10000 | |
| % | | | | 0.000% | 73.450% | | |

☞ Response Generation Parameters

No. of Hazard Pieces: 2
 Input Method: Hazard Rates

| Piece # | Starting at Time | Control | Treatment | Hazard Ratio |
|---------|------------------|---------|-----------|--------------|
| 1 | 0 | 0.003 | 0.003 | 1.015 |
| 2 | 12 | 0.003 | 0.002 | 0.75 |

Now you can see, the power of the study has increased to 73.5%. On average 811 events occurred during the 50 months that the study remained open. Since we require 90% power, the sample size must be increased even further. This can be done by trial and error over several simulation experiments. Eventually we discover that a sample size of 18000 patients will provide about 90% power with an average of 1358 events.

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

| Simulation Parameters | |
|-------------------------------|-------------|
| Simulation ID | Sim8 |
| Design Type | Superiority |
| Number of Looks | 1 |
| Test Type | 2-Sided |
| Sample Size (n) | 18000 |
| Fix at Each Look | Look Time |
| Test Statistic | Logrank |
| Average Events | 1358.741 |
| Total Accrual Duration | 18 |
| Avg. Power at Termination | 0.892 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

☞ Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 18000 | 736.419 | 622.323 | 2585.87 | 2600.32 | 50 | 32.82 |
| Average | 18000 | 736.419 | 622.323 | 2585.87 | 2600.32 | 50 | 32.82 |

☞ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Look Time | Boundaries Efficacy | | Stopping For | | Total Simulations | |
|--------|-----------|---------------------|-------|----------------|----------------|-------------------|---|
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| | | | | | | | |
| Total | | | | 0 | 8922 | 10000 | |
| % | | | | 0.000% | 89.220% | | |

☞ Response Generation Parameters

No. of Hazard Pieces: 2
 Input Method: Hazard Rates

| Piece # | Starting at Time | Control | Treatment | Hazard Ratio |
|---------|------------------|---------|-----------|--------------|
| 1 | 0 | 0.003 | 0.003 | 1.015 |
| 2 | 12 | 0.003 | 0.002 | 0.75 |

It is evident from these simulations that the proportional hazards assumption is simply not appropriate if the survival curves separate out late. In the present example the

44 Superiority Trials with Variable Follow-Up

proportional hazards assumption would have led to a sample size of 6400 whereas the sample size actually needed was 18000.

44.6.3 Group Sequential Design with Non-Proportional Hazards

The single-look design discussed in the previous section required a sample size of 17200 subjects. A group sequential design, monitored by an independent data monitoring committee, is usually more efficient for large studies of this type. Such a trial can be designed with efficacy stopping boundaries or with efficacy and futility stopping boundaries. Consider first a design with five equally spaced efficacy boundaries. Go back to the library, click on Des1 node, and then click on  . In the resulting design input dialog window, change the entry in the **Number of Looks** cell from 1 to 5. Click on **Compute** button and save the plan as Des2 in the library. Select Des1 and Des2 nodes and then click on  to see the following

details for both the designs.

| | Wbk2:Des1 | Wbk2:Des2 |
|---|--------------------|--------------------|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 5 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.9 | 0.9002 |
| Model Parameters | | |
| Hazard Ratio (Alt.) | 0.75 | 0.75 |
| Var (Log HR) | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Boundary Parameters | | |
| Spacing of Looks | | Equal |
| Efficacy Boundary | | LD (OF) |
| Accrual & Dropout Parameters | | |
| Accrual Rate | 1000 | 1000 |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 |
| Sample Size | | |
| Maximum | 6400 | 12555 |
| Expected Under H0 | 6400 | 12554.9964 |
| Expected Under H1 | 6400 | 12554.1223 |
| Events | | |
| Maximum | 508 | 520 |
| Expected Under H0 | 508 | 516.5858 |
| Expected Under H1 | 508 | 385.5073 |
| Accrual Duration | | |
| Maximum | 6.4 | 12.555 |
| Expected Under H0 | 6.4 | 12.555 |
| Expected Under H1 | 6.4 | 12.5541 |
| Study Duration | | |
| Maximum | 48.677 | 27.2317 |
| Expected Under H0 | 41.817 | 24.2646 |
| Expected Under H1 | 48.677 | 21.4502 |

Des2 reveals that a group sequential design, with five equally spaced looks, taken after observing 104, 208, 312, 416 and 520 events, respectively, utilizing the default Lan-DeMets-O'Brien-Fleming (LD (OF)) spending function, achieves 90% power with a maximum sample size of 12555 and a maximum study duration of 27.232 months. The expected study duration under H_1 is 21.451 months. However, these operating characteristics are based on the assumption that the hazard ratio is constant and equals 0.75. Since in fact the hazard ratio is 0.75 only after 12 months of treatment, the actual power of this design is unlikely to be 90%. We can use simulation to determine the actual power. With the cursor in any cell of Des2 node, select 

44 Superiority Trials with Variable Follow-Up

from the menu bar. You will be taken to the simulation worksheet. In the **Response Generation** tab, make the changes in the hazard rates as shown below.

| Survival Information | | | | |
|---------------------------------------|------------------|---------------|--------------|--------------|
| # of Hazard Pieces | 2 | Input Method: | Hazard Rates | |
| <input type="checkbox"/> Hazard Ratio | | | | |
| Piece | Starting at Time | Hazard Rates | | Hazard Ratio |
| | | Control | Treatment | |
| 1 | 0.000 | 0.0025 | 0.0025 | 1.000 |
| 2 | 12.000 | 0.0025 | 0.0019 | 0.760 |

After changing the number of simulations as 1000 in the **Simulation Control**, click on the **Simulate** button to run 1000 simulations of Des2 with data being generated from the survival distributions that were specified in the **Response Generation** tab.

The results of this simulation run are as shown below.

| Wbk2:Des2:Sim6 | |
|---|--------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Test Statistic | Logrank |
| Power | 0.187 |
| No. of Looks | 5 |
| Model Parameters | |
| No. of Hazard Pieces | 2 |
| Boundary Parameters | |
| Efficacy Boundary | User Specified |
| Spacing of Looks | User Specified |
| Accrual & Dropout Parameters | |
| Followup Duration | Until End of Study |
| Accrual Rate | 1000 |
| No. of Accrual Periods | 1 |
| Sample Size | |
| Maximum | 12555 |
| Events | |
| Maximum | 520 |
| Study Duration | |
| Maximum | 25.2768 |

Only 187 of the 1000 simulated trials were able to reject the null hypothesis indicating that the study is grossly underpowered. We can improve on this performance by extending the total study duration so that additional events may be observed. To increase study duration, go to the **Simulation Parameters** tab and select the **Look Time** option under **Fix at Each Look**. We had specified at the outset that the total study duration should not exceed 50 months. Let us therefore fix the total study duration at 50 months and space each interim look 10 months apart by editing

44 Superiority Trials with Variable Follow-Up

the **Study Duration**.

The screenshot shows the 'Test Parameters' tab of a trial design software. The 'Trial Type' is set to 'Superiority', 'Test Type' is '2-Sided', and 'Fix at Each Look' is 'Look Time'. The 'Study Duration' is set to 50. The 'Test Statistic' is 'Logrank'. Below the settings is a table showing efficacy Z values for 5 looks.

| Look # | Analysis Time | Efficacy Z | |
|--------|---------------|------------|--------|
| | | Upper | Lower |
| 1 | 10.000 | 4.877 | -4.877 |
| 2 | 20.000 | 3.357 | -3.357 |
| 3 | 30.000 | 2.680 | -2.680 |
| 4 | 40.000 | 2.290 | -2.290 |
| 5 | 50.000 | 2.031 | -2.031 |

We are now ready to simulate a 5-look group sequential trial in which the **LD (OF)** stopping boundaries are applied and the looks are spaced 10 months apart. Each simulated trial will enroll 12555 subjects at the rate of 1000/month. The simulation data will be generated from survival distributions in which the hazard rates of both arms are 0.0025 for the first 12 months and the hazard ratio is 0.75 thereafter. To generate 1000 simulations of this design click on the **Simulate** button. These simulations do indeed show a substantial increase in power, from 18.7% previously to 79.9% .

| | Wbk2:Des2:Sim6 | Wbk2:Des2:Sim7 |
|---|--------------------|--------------------|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| Test Type | 2-Sided | 2-Sided |
| Test Statistic | Logrank | Logrank |
| Power | 0.187 | 0.799 |
| No. of Looks | 5 | 5 |
| Model Parameters | | |
| No. of Hazard Pieces | 2 | 2 |
| Boundary Parameters | | |
| Efficacy Boundary | User Specified | User Specified |
| Spacing of Looks | User Specified | User Specified |
| Accrual & Dropout Parameters | | |
| Followup Duration | Until End of Study | Until End of Study |
| Accrual Rate | 1000 | 1000 |
| No. of Accrual Periods | 1 | 1 |
| Sample Size | | |
| Maximum | 12555 | 12555 |
| Events | | |
| Maximum | 520 | 879.219 |
| Study Duration | | |
| Maximum | 25.2768 | 50 |

The design specifications stated, however, that the trial should have 90% power. In order to achieve this amount of power we will have to increase the sample size. By trial and error, upon increasing the sample size to 18200 on the **Simulation Parameters** tab we observe that the power has increased to 90 % (up to Monte Carlo accuracy).

| | Wbk2:Des2:Sim6 | Wbk2:Des2:Sim7 | Wbk2:Des2:Sim8 |
|---|-----------------------|-----------------------|-----------------------|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR | SU-2S-LRAR |
| Test Parameters | | | |
| Design Type | Superiority | Superiority | Superiority |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Test Statistic | Logrank | Logrank | Logrank |
| Power | 0.187 | 0.799 | 0.904 |
| No. of Looks | 5 | 5 | 5 |
| Model Parameters | | | |
| No. of Hazard Pieces | 2 | 2 | 2 |
| Boundary Parameters | | | |
| Efficacy Boundary | User Specified | User Specified | User Specified |
| Spacing of Looks | User Specified | User Specified | User Specified |
| Accrual & Dropout Parameters | | | |
| Followup Duration | Until End of Study | Until End of Study | Until End of Study |
| Accrual Rate | 1000 | 1000 | 1000 |
| No. of Accrual Periods | 1 | 1 | 1 |
| Sample Size | | | |
| Maximum | 12555 | 12555 | 18300 |
| Events | | | |
| Maximum | 520 | 879.219 | 1169.605 |
| Study Duration | | | |
| Maximum | 25.2768 | 50 | 50 |

44 Superiority Trials with Variable Follow-Up

44.7 Simulating a Trial with Stratification variables

The data presented in Appendix I of Kalbfleisch and Prentice (1980) on lung cancer patients were used as a basis for this example. We will design a trial to compare two treatments (Standard and Test) in a target patient group where patients had some prior therapy. The response variable is the survival time in days of lung cancer patients. First, we will create a design for 3 looks, to compare the two treatment groups. Next, using this design, we will carry out simulation with stratification variables. Three covariates in the data are used here as stratum variables: a) type of cancer cell (small, adeno, large, squamous,), b) age in years (≤ 50 , > 50), and c) performance status score (≤ 50 , > 50 and ≤ 70 , > 70).

The input data for base design are as follows: Trial type:superiority; test type:2-sided; type I error:0.05; power:0.90; allocation ratio:1; hazard rate (control):0.009211; hazard rate (treatment):0.004114; number of looks:3; Boundary family:spending functions; spending function:Lan-DeMets (OF); subjects are followed:until end of study; subjects accrual rate:12 per day.

The input data for stratified simulation are as given below: The number of stratum variables=3 (cell type; age group; performance status score).

Table 44.1: Input data for stratified simulation

| Cell type | Proportion | Hazard ratio |
|--------------------------------|------------|--------------|
| small | 0.28 | Baseline |
| adeno | 0.13 | 2.127 |
| large | 0.25 | 0.528 |
| squamous | 0.34 | 0.413 |
| Age group | Proportion | Hazard ratio |
| ≤ 50 years | 0.28 | Baseline |
| > 50 years | 0.72 | 0.438 |
| Performance status score group | Proportion | Hazard ratio |
| ≤ 50 | 0.43 | Baseline |
| > 50 and ≤ 70 | 0.37 | 0.164 |
| > 70 | 0.20 | 0.159 |

44.7.1 Creating the design

First we will create a design using the input data. Open East, click **Design** tab and then **Two Samples** button in **Survival** group. Now click **Logrank Test: Given Accrual Duration and Accrual Rates**. In the resulting screen, enter the input data in the dialog

boxes under the different tabs. Finally click on **Compute** button. Now the dialog boxes under the different tabs will appear as shown below.

The **Test Parameters** tab is shown below, where you can see the computed value of **No.of Events**.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary Accrual / Dropouts

Test Type: 2-Sided # of Hazard Pieces: 1 Input Method: Hazard Rates

Type I Error (α): 0.05

Power: 0.9

No. of Events: Computed

Allocation Ratio: 1
(n_1/n_2)

Hazard Ratio (Optional) Alternative: 0.4466

Hazard Ratio (λ_c/λ_t)

Log Hazard Ratio $\ln(\lambda_c/\lambda_t)$ -0.806

| Hazard Rate | |
|-------------|-----------------|
| Control | Treatment: Alt. |
| 0.009211 | 0.004114 |

Variance of Log Hazard Ratio

Null Alternative

Assurance (Probability of Success)

The **Boundary** will appear as shown below, where all the input data are seen.

Test Parameters Boundary Accrual / Dropouts

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.05

Futility

Boundary Family: None

Spacing of Looks Equal Unequal

Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|--------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.333 | 0.000 | 3.710 | -3.710 |
| 2 | 0.667 | 0.012 | 2.511 | -2.511 |
| 3 | 1.000 | 0.050 | 1.993 | -1.993 |

44 Superiority Trials with Variable Follow-Up

The **Accrual/Dropouts** tab containing the input data will be as shown below.

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary **Accrual / Dropouts**

Subjects are followed: Until End of Study

Accrual Info

of Accrual Periods: 1

Accrual Rate: 12

Piecewise Dropout Information

of Pieces: 0

| Accrual | Min. | Comtd. | Sugg. Max. |
|--|------|--------|------------|
| <input type="radio"/> Duration: | 5.5 | 28 | 42.5 |
| <input checked="" type="radio"/> Subjects: | 66 | 288 | 510 |

After the design is completed and saved in a workbook, select the design node and

click on the **output summary** icon to see the following output display.

| Wbk3:Des 1 | |
|---|--------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 3 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9023 |
| Model Parameters | |
| Hazard Ratio (Alt.) | 0.4466 |
| Var (Log HR) | Null |
| Allocation Ratio (nt/nc) | 1 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Accrual & Dropout Parameters | |
| Accrual Rate | 12 |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 288 |
| Expected Under H0 | 287.9915 |
| Expected Under H1 | 288 |
| Events | |
| Maximum | 66 |
| Expected Under H0 | 65.7293 |
| Expected Under H1 | 52.8227 |
| Accrual Duration | |
| Maximum | 24 |
| Expected Under H0 | 23.9993 |
| Expected Under H1 | 24 |
| Study Duration | |
| Maximum | 52.0169 |
| Expected Under H0 | 40.3527 |
| Expected Under H1 | 43.2957 |

44.7.2 Running Stratified Simulation

After selecting the design node, click on **Simulate** icon. You will see simulation screen with the dialog boxes under different tabs. Click on **Include Options** and select **Stratification**.

44 Superiority Trials with Variable Follow-Up

The dialog box under **Test Parameters** will be as shown below. Keep the default test statistic **LogRank** and the default choice of **Use Stratified Statistic**.

| Look # | Info. Fraction | Cum. α Spent | | Efficacy Z | |
|--------|----------------|--------------|-------|------------|--------|
| | | Upper | Lower | Upper | Lower |
| 1 | 0.333 | 0.000 | 0.000 | 3.710 | -3.710 |
| 2 | 0.667 | 0.006 | 0.006 | 2.511 | -2.511 |
| 3 | 1.000 | 0.025 | 0.025 | 1.993 | -1.993 |

After entering the stratification input information, the dialog box under **Stratification** will appear as shown below.

| Cell type | Level | Fraction | Age group | Level | Fraction | PerfStatus | Level | Fraction |
|-----------|---------|----------|-----------|---------|----------|------------|-------|----------|
| small | <=50... | 0.28 | > 50... | <=50... | 0.28 | Perf_1 | 0.43 | |
| | adeno | 0.13 | | > 50... | 0.72 | Perf_2 | 0.37 | |
| large | large | 0.25 | squa... | large | 0.34 | Perf_3 | 0.2 | |
| | squa... | 0.34 | | Perf_3 | 0.2 | | | |

| Stratum ID | Label | Fraction |
|------------|---------------------------|----------|
| SID01 | small <=50 yrs Perf_1 | 0.034 |
| SID02 | small <=50 yrs Perf_2 | 0.029 |
| SID03 | small <=50 yrs Perf_3 | 0.016 |
| SID04 | small > 50 yrs Perf_1 | 0.087 |
| SID05 | small > 50 yrs Perf_2 | 0.075 |
| SID06 | small > 50 yrs Perf_3 | 0.040 |
| SID07 | adeno <=50 yrs Perf_1 | 0.016 |
| SID08 | adeno <=50 yrs Perf_2 | 0.013 |
| SID09 | adeno <=50 yrs Perf_3 | 0.007 |

After entering adding response related input information, the dialog box under

Response Generation will display details as shown in the following screen shots.

The screenshot shows a software interface with several tabs: Simulation Parameters, Stratification Info, Response Generation Info, Accrual/Dropout Info, and Simulation Control Info. The 'Response Generation Info' tab is active. It displays 'Survival Information' with a dropdown for 'Stratum ID' set to 'SID01:small | <=50 yrs | Perf_1'. There are radio buttons for 'Using Hazard Rates' (selected) and 'Using Cum. % Survival'. Below this is a table showing 'Hazard Rates' for 'Control' and 'Treatment' across different strata.

| Piece | Starting At | Hazard Rates | | Hazard Ratio |
|-------|-------------|--------------|-----------|--------------|
| | | Control | Treatment | |
| 1 | 0.0000 | 0.0092 | 0.0041 | 0.4466 |

Below the main interface, three detailed views of the 'Specify Hazard Ratio' table are shown for different variables: Cell type, Age group, and PerfStatus.

Variables: Treatment, Cell type, Age group, PerfStatus

Specify Hazard Ratio:

| Cell type | | |
|-----------|----------|--------------|
| Level | Fraction | Hazard Ratio |
| small | 0.280 | Baseline |
| adeno | 0.130 | 2.127 |
| large | 0.250 | 0.528 |
| squamous | 0.340 | 0.413 |

Variables: Treatment, Cell type, Age group, PerfStatus

Specify Hazard Ratio:

| Age group | | |
|-----------|----------|--------------|
| Level | Fraction | Hazard Ratio |
| <= 50 yrs | 0.280 | Baseline |
| > 50 yrs | 0.720 | 0.438 |

Variables: Treatment, Cell type, Age group, PerfStatus

Specify Hazard Ratio:

| PerfStatus | | |
|------------|----------|--------------|
| Level | Fraction | Hazard Ratio |
| Perf_1 | 0.430 | Baseline |
| Perf_2 | 0.370 | 0.164 |
| Perf_3 | 0.200 | 0.159 |

44 Superiority Trials with Variable Follow-Up

The **Accrual/Dropout** dialog box will appear as shown below.

Simulation Parameters | Stratification Info | Response Generation Info | **Accrual / Dropout Info** | Simulation Control Info

Sample Size: Distribution of Accrual Time: Uniform

Subjects are followed:

Accrual Info

of Accrual Periods: Input Method:

| Period # | Starting At | Accrual Rate |
|----------|-------------|--------------|
| 1 | 0.0000 | 12.0000 |

Piecwise Constant Dropout Rates

of Pieces: Input Method:

| Period # | Starting At | Hazard Rate (Control) | Hazard Rate (Treatment) |
|----------|-------------|-----------------------|-------------------------|
| | | | |

In the **Simulation Control** tab, specify number of simulations as 1000 and select the choices under output options to save simulation data. The dialog box will appear as shown below.

Simulation Parameters | Stratification Info | Response Generation Info | **Accrual / Dropout Info** | **Simulation Control Info**

Number of Simulations:

Refresh Frequency:

Random Number Seed

Clock

Fixed

Output Options

Save summary statistics for every simulation run

Save subject-level data for simulation runs

Note: Max. 100,000 records will be saved.

Suppress All Intermediate Output

Pause after Refresh

After clicking on **Simulate** button, the results will appear in the Output Preview row. Click on it and save it in the workbook. Select this simulation node and click on

Output Summary icon to see the following stratification simulation output summary.

| | Wbk3:Des1:Sim1 |
|---|-----------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Test Statistic | Stratified Logrank |
| Power | 0.856 |
| No. of Looks | 3 |
| Boundary Parameters | |
| Efficacy Boundary | User Specified |
| Spacing of Looks | User Specified |
| Accrual & Dropout Parameters | |
| Followup Duration | Until End of Study |
| Accrual Rate | 12 |
| No. of Accrual Periods | 1 |
| Sample Size | |
| Maximum | 288 |
| Events | |
| Maximum | 66 |
| Study Duration | |
| Maximum | 172.4444 |
| Stratum Information | |
| No. of Stratum Variables | 3 |
| No. of Strata | 24 |
| Allocate Fractions to Strata | Marginally |
| Specification of Hazard Rates | Model based |

The stratified simulation results show that the attained power 0.856 is slightly less than the design specified power of 0.90.

45 Superiority Trials with Fixed Follow-Up

This chapter will illustrate through a worked example how to design, monitor and simulate a two-sample superiority trial with a time-to-event endpoint in which each subject who has not dropped out or experienced the event is followed for a fixed duration only. This implies that each subject who does not drop-out or experience the event within a given time interval, as measured from the time of randomization, will be administratively censored at the end of that interval. In East we refer to such designs as **fixed follow-up designs**.

45.1 Clinical Trial of Drug Eluting Stents

Drug-eluting coronary-artery stents were shown to decrease the risks of death from cardiac causes, myocardial infarction and target-vessel revascularization as compared to uncoated stents in patients undergoing primary percutaneous coronary intervention (PCI) in two randomized clinical trials published in the September 14, 2006 issue of the New England Journal of Medicine. In the Paclitaxel-Eluting Stent versus Conventional Stent in Myocardial Infarction with ST-Segment Elevation (PASSION) trial, Laarman et al. (2006) randomly assigned 619 patients to receive either a paclitaxel-eluting stent or an uncoated stent. The primary endpoint was the percentage of cardiac deaths, recurrent myocardial infarctions or target-lesion revascularizations at 12 months. A marginally lower 12-month failure rate was observed in the paclitaxel-stent group compared with the uncoated-stent group (8.8% versus 12.8%, $p = 0.09$). The Trial to Assess the Use of the Cypher Stent in Acute Myocardial Infarction Treated with Balloon Angioplasty (TYPHOON), (Spaulding et al., 2006) showed even more promising results. In this trial of 712 patients the sirolimus-eluting stents had a significantly lower target-vessel failure rate at 12 months than the uncoated stents (7.3% versus 14.3%, $p = 0.004$). Based on these results an editorial by Van de Werf (2006) appeared in the same issue of the New England Journal of Medicine as the Typhoon and PASSION trials, recommending that studies with a larger sample size and a hard clinical endpoint be conducted so that drug-eluting stents might be routinely implanted in patients undergoing PCI. In this chapter we will use East to design and monitor a possible successor to the PASSION trial using a time-to-event endpoint with one year of fixed follow-up for each subject.

45.2 Single-Look Design

45.2.1 Initial Design

The primary endpoint for the trial is the time to target-vessel failure, with a failure being defined as target-vessel related death, recurrent myocardial infarction, or target-vessel revascularization. Each subject will be followed for 12 months. Based on the PASSION data we expect that 87.2% of subjects randomized to the uncoated stents will be event-free at 12 months. We will design the trial for 90% power to detect an increase to 91.2% in the paclitaxel-stents group, using a two-sided level-0.05 test. Enrollment is expected to be at the rate of 30 subjects per month.

45.2.1 Initial Design

We begin by opening a new East Workbook and selecting **Logrank Test Given Accrual Duration and Accrual Rates**.

This will open the input window for the design as shown below. Select **2-Sided** for **Test Type**, and enter **0.05** for **Type I error**.

The right hand side panel of this input window is to be used for entering the relevant time-to-event information.

| Test Parameters | Accrual / Dropouts |
|------------------------------------|--------------------|
| Test Type: | 2-Sided |
| Type I Error (α): | 0.05 |
| Power: | 0.9 |
| No. of Events: | Computed |
| Allocation Ratio: (n_1/n_2) | 1 |

The default values in the above dialog box must be changed to reflect the time-to-event parameters specified for the design. Select **% Cumulative Survival** for the **Input**

45 Superiority Trials with Fixed Follow-Up

Method and enter the relevant 12-month event-free percentages.

of Hazard Pieces: Input Method:

Hazard Ratio (Optional) Alternative

Hazard Ratio (λ_t / λ_c)

Ratio of % Survivals at Period # 1 (S_t / S_c)

| By Time | Cum. % Survival | |
|---------------------------------|-----------------------------------|-----------------------------------|
| | Control | Treatment: Alt. |
| <input type="text" value="12"/> | <input type="text" value="87.2"/> | <input type="text" value="91.2"/> |

Note: Period 1 hazard rates apply after time 12.

Change the **Input Method** to Hazard Rates. You will see the information you entered converted as shown below. Note that you may need to change the decimal display options for hazard rates using the  icon to see these numbers with more decimal places.

of Hazard Pieces: Input Method:

Hazard Ratio (Optional) Alternative

Hazard Ratio (λ_t / λ_c)

Log Hazard Ratio $\ln(\lambda_t / \lambda_c)$

| Hazard Rate | |
|-------------------------------------|-------------------------------------|
| Control | Treatment: Alt. |
| <input type="text" value="0.0114"/> | <input type="text" value="0.0077"/> |

Another parameter to be decided is the **Variance** which specifies whether the calculation of the required number of events is to be based on the variance estimate of log hazard ratio under the null hypothesis or the alternative hypothesis. The default choice in East is **Null**. Most textbooks recommend this choice as well (see, for example Collett, 1994, equation (2.21) specialized to no ties). It will usually not be necessary to change this default. For a technical discussion of this issue refer to

Appendix B, Section B.5.3

The second tab, labeled **Accrual/Dropouts** is used to enter the patient accrual rate and, for fixed follow-up designs, the duration of patient follow-up and the dropout information. In this example the clinical endpoint is progression-free survival for 12 months. Patients who are still on study at month 12 and who have not experienced the endpoint will be treated as censored. Therefore, in the first panel out of two, we select the entry from the dropdown that indicates that subjects are followed **For Fixed Period** and enter the number 12 in the corresponding edit box. Suppose that the anticipated rate of enrollment is 30 patients per month. This number is also entered into the dialog box as shown below. Let the committed accrual of subjects be same as 2474.

Test Parameters | **Accrual / Dropouts**

Subjects are followed: For Fixed Period [12]

Accrual Info

of Accrual Periods: [1]

Accrual Rate: [30]

| | Min. | Comtd. | Sugg. Max. |
|--|--------|--------|------------|
| <input type="radio"/> Duration: | 82.467 | 82.467 | 88.267 |
| <input checked="" type="radio"/> Subjects: | 2474 | 2474 | 2648 |

The second panel, labeled **Piecewise Constant Dropout Rates**, is used to enter the rate at which we expect patients to drop out of the study. For the present we will assume that there are no drop-outs.

— Piecewise Dropout Information —

of Pieces: [0]

45 Superiority Trials with Fixed Follow-Up

An initial design, titled Des1, is created in the **Output Preview** pane upon clicking

the **Compute** button. Click on  icon to save the design in a workbook or on

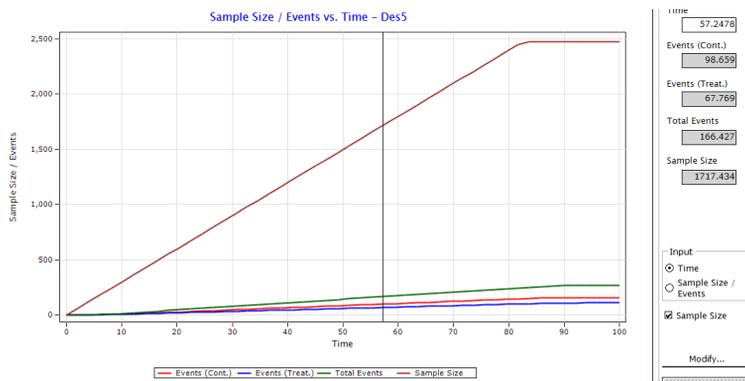


icon to see the output summary of this design.

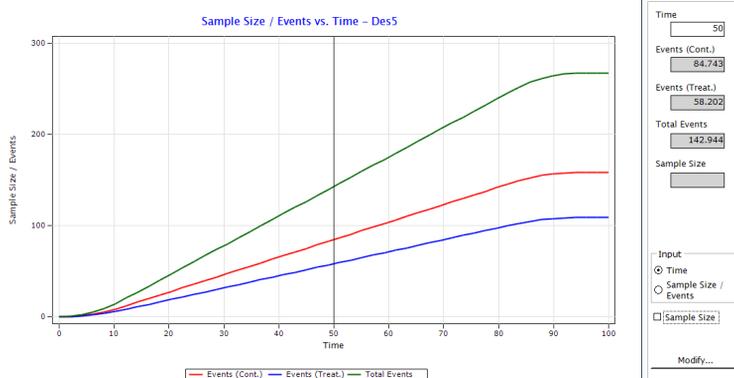
| | Des5 |
|---|-----------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.901 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.673 |
| Var (Log HR) | Null |
| Accrual & Dropout Parameters | |
| Accrual Rate | 30 |
| Subjects are Followed | For Fixed Period (12) |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 2474 |
| Expected Under H0 | 2269.642 |
| Expected Under H1 | 2474 |
| Events | |
| Maximum | 268 |
| Expected Under H0 | 268 |
| Expected Under H1 | 268 |
| Study Duration | |
| Maximum | 94.467 |
| Expected Under H0 | 75.655 |
| Expected Under H1 | 94.467 |
| Accrual Duration | |
| Maximum | 82.467 |
| Expected Under H0 | 75.655 |
| Expected Under H1 | 82.467 |

East reveals that 268 events are required in order to obtain 90% power. If each patient can only be followed for a maximum of 12 months, we must commit to enrolling a total of 2474 patients over a period of 82.5 months. With this commitment we expect to see the required 268 events within 12 months of the last patient being enrolled. So the total study duration is expected to be $82.5 + 12 = 94.5$ months. To see how the

events are expected to arrive over time, invoke a plot of **Sample Size/ Events vs. Time** by clicking the **Plots** icon  from the toolbar.



Uncheck the **Sample Size** box, to see the events graphs on a larger scale as shown below.



45.3 Shortening the Study Duration

45.3.1 Increasing the Sample Size

45.3.2 Patient Follow-Up

45.3.3 Increasing the Rate of Enrollment

Under Des1 the trial will last for 94.5 months, with 82.5 months of patient enrollment (i.e., a sample size of 2474 subjects). This is not considered to be satisfactory to the trial sponsor. There are three possible ways in which the study duration might be shortened; by increasing the sample size, by increasing the duration of patient follow-up, or by increasing the rate of patient enrollment.

45.3.1 Increasing the Sample Size

45 Superiority Trials with Fixed Follow-Up

Unlike trials with variable patient follow-up, in a fixed follow-up design the gain from increasing the sample size is limited. This is evident from the relatively narrow range between the minimum accrual duration (82.5 months) and the suggested maximum accrual duration (88.3 months).

| Accrual | | | |
|--|--------|--------|------------|
| | Min. | Comtd. | Sugg. Max. |
| <input type="radio"/> Duration: | 82.467 | 82.467 | 88.267 |
| <input checked="" type="radio"/> Subjects: | 2474 | 2474 | 2648 |

Notice that if we were to increase the enrollment duration to the say, 88.3 months, the total study duration would only decrease by 5.9 months; from 94.5 months to 88.6 months. To see this, edit Des1 and create Des2 and enter the number 88.267 into the cell for **Committed Accrual (Duration)** as shown below:

| Accrual | | | |
|--|--------|--------|------------|
| | Min. | Comtd. | Sugg. Max. |
| <input type="radio"/> Duration: | 82.467 | 88.267 | 88.267 |
| <input checked="" type="radio"/> Subjects: | 2474 | 2648 | 2648 |

Des2 is created in the **Output Preview** pane upon clicking the **Compute** button.

Click on  icon to save the design in a workbook. Select Des1 and Des2 in the workbook and click on  icon to see the side-by-side comparison of the two

designs.

| | Wbk1:Des1 | Wbk1:Des2 |
|---|-----------------------|-----------------------|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 1 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.901 | 0.901 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Hazard Ratio (Alt.) | 0.673 | 0.673 |
| Var (Log HR) | Null | Null |
| Accrual & Dropout Parameters | | |
| Accrual Rate | 30 | 30 |
| Subjects are Followed | For Fixed Period (12) | For Fixed Period (12) |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 0 | 0 |
| Sample Size | | |
| Maximum | 2474 | 2648 |
| Expected Under H0 | 2269.642 | 2269.642 |
| Expected Under H1 | 2474 | 2648 |
| Events | | |
| Maximum | 268 | 268 |
| Expected Under H0 | 268 | 268 |
| Expected Under H1 | 268 | 268 |
| Study Duration | | |
| Maximum | 94.467 | 88.602 |
| Expected Under H0 | 75.655 | 75.655 |
| Expected Under H1 | 94.467 | 88.602 |
| Accrual Duration | | |
| Maximum | 82.467 | 88.267 |
| Expected Under H0 | 75.655 | 75.655 |
| Expected Under H1 | 82.467 | 88.267 |

The calculation of the minimum and maximum of the range of accrual durations is discussed on page 2308 of Appendix B, section B.5.2.

East has determined that if the enrollment (at the pre-specified rate of 30 patients per month) is stopped **before** 82.467 months have elapsed, and every patient still on study is followed for precisely 12 months, we will obtain fewer than 268 events on average, and the trial will be underpowered. Therefore East specifies that the minimum duration of enrollment must be 82.467 months. The user has the option to increase the enrollment duration beyond 82.467 months. In that case, however, if all patients still on study are followed for 12 months, more than 268 events will accumulate, on average, by the time the trial is terminated. Therefore it will not be necessary to follow the later enrollees for the entire 12 month period. In the extreme case, if we extend the enrollment duration to 88.267 months, the required 268 events will arrive, on average, by the end of the enrollment period itself thus making it unnecessary to have any follow-up after the last patient has been enrolled. The study duration cannot be shortened any further by increasing the enrollment beyond 88.267 months (i.e., extending the sample size beyond 2648). For these reasons, for fixed follow-up

45 Superiority Trials with Fixed Follow-Up

designs, East selects the **minimum** of the range of enrollment durations (in this case 82.467) as the default enrollment duration. We note that in contrast, for variable follow-up designs, East selects the **mid-point** of the range of suggested enrollment durations as the default. Of course, the user is free to change the default enrollment duration for both types of designs.

45.3.2 Increasing the Length of Patient Follow-Up

Since this is a fixed follow-up design we might consider increasing the length of

patient follow-up, at present equal to 12 months. Edit Des1 by clicking the icon  to create Des3. Increase the length of patient follow-up from 12 months to 18 months, and commit to the minimum sample size 1698.

Test Parameters

Accrual / Dropouts

Subjects are followed: For Fixed Period 18

Accrual Info

of Accrual Periods: 1

Accrual Rate: 30

| Accrual | Min. | Comtd. | Sugg. Max. |
|--|------|--------|------------|
| <input type="radio"/> Duration: | 56.6 | 56.6 | 65.267 |
| <input checked="" type="radio"/> Subjects: | 1698 | 1698 | 1958 |

Click on **Compute** to get Des3 as shown below.

| | Wbk1:Des3 |
|---|-----------------------|
| Mnemonic | SU-25-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.901 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.673 |
| Var (Log HR) | Null |
| Accrual & Dropout Parameters | |
| Accrual Rate | 30 |
| Subjects are Followed | For Fixed Period (18) |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 1698 |
| Expected Under H0 | 1698 |
| Expected Under H1 | 1698 |
| Events | |
| Maximum | 268 |
| Expected Under H0 | 268 |
| Expected Under H1 | 268 |
| Study Duration | |
| Maximum | 74.6 |
| Expected Under H0 | 56.795 |
| Expected Under H1 | 74.6 |
| Accrual Duration | |
| Maximum | 56.6 |
| Expected Under H0 | 56.6 |
| Expected Under H1 | 56.6 |

By increasing the duration of patient follow-up to 18 months, the required 268 events can be obtained with fewer patients. It is now only necessary to have enrollment duration of 56.6 months. The study is expected to terminate 18 months after the last patient has enrolled for total study duration of 74.6. Increasing the length of patient follow-up has indeed shortened the total study duration. We note, however, that it might not always be feasible to increase patient follow-up in this manner, particularly if the clinical endpoint of interest determines how long one should wait for the endpoint to occur.

45 Superiority Trials with Fixed Follow-Up

45.3.3 Increasing the Rate of Enrollment

In cases where the primary endpoint determines the duration of the fixed follow-up, the option to shorten the study duration by increasing the follow-up duration is not available. In that case the only possibility is to increase the rate of enrollment by opening up more sites. Edit Des1 to create Des4 and increase the rate of enrollment from 30 patients/month to 45 patients/month, while committing to accruing the minimum number of subjects 2474.

Subjects are followed: For Fixed Period 12

Accrual Info

of Accrual Periods: 1

Accrual Rate: 45

| Accrual | Min. | Comtd. | Sugg. Max. |
|--|--------|--------|------------|
| <input type="radio"/> Duration: | 54.978 | 54.978 | 60.8 |
| <input checked="" type="radio"/> Subjects: | 2474 | 2474 | 2736 |

With this enrollment rate East calculates that an enrollment duration of 55 (sample size of 2474) and 12 additional months of follow-up will produce the desired 268 events on average. Thus the total study duration is expected to be 67 months.

| | Des4 |
|---|-----------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.901 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.673 |
| Var (Log HR) | Null |
| Accrual & Dropout Parameters | |
| Accrual Rate | 45 |
| Subjects are Followed | For Fixed Period (12) |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 2474 |
| Expected Under H0 | 2357.587 |
| Expected Under H1 | 2474 |
| Events | |
| Maximum | 268 |
| Expected Under H0 | 268 |
| Expected Under H1 | 268 |
| Study Duration | |
| Maximum | 66.978 |
| Expected Under H0 | 52.391 |
| Expected Under H1 | 66.978 |
| Accrual Duration | |
| Maximum | 54.978 |
| Expected Under H0 | 52.391 |
| Expected Under H1 | 54.978 |

Now try an enrollment rate of 51.5 patients/month, and remember to maintain the 2474 accrual. At this enrollment rate the study is fully powered with a sample size of 2474 subjects, enrolled over a period of 48 months. The required 268 events will arrive on average 12 months after the last patient has enrolled so that the trial is expected to terminate at month 60.

45 Superiority Trials with Fixed Follow-Up

Subjects are followed: For Fixed Period

Accrual Info

of Accrual Periods:

Accrual Rate:

Accrual

| | Min. | Comtd. | Sugg. Max. |
|--|--------|--------|------------|
| <input type="radio"/> Duration: | 48.039 | 48.039 | 53.864 |
| <input checked="" type="radio"/> Subjects: | 2474 | 2474 | 2774 |

Click on **Compute** to see the design as shown below.

| | Des5 |
|---|-----------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.901 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.673 |
| Var (Log HR) | Null |
| Accrual & Dropout Parameters | |
| Accrual Rate | 51.5 |
| Subjects are Followed | For Fixed Period (12) |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 2474 |
| Expected Under H0 | 2395.697 |
| Expected Under H1 | 2474 |
| Events | |
| Maximum | 268 |
| Expected Under H0 | 268 |
| Expected Under H1 | 268 |
| Study Duration | |
| Maximum | 60.039 |
| Expected Under H0 | 46.518 |
| Expected Under H1 | 60.039 |
| Accrual Duration | |
| Maximum | 48.039 |
| Expected Under H0 | 46.518 |
| Expected Under H1 | 48.039 |

45 Superiority Trials with Fixed Follow-Up

45.4 Group Sequential Design

45.4.1 Incorporating Drop-Outs

45.4.2 Non-Const. Accr. Rates

45.4.3 Piece-wise Exp. Survival

Edit Des5 and change the number of looks from 1 to 5, equally spaced, with the default **LD (OF)** spending function. This will create Des6. Click on **Boundary** tab to choose the boundary family and alpha spending function shown below:

Design Type: Superiority Number of Looks: 5

Design Parameters Boundary Info Accrual / Dropout Info

Efficacy Boundary Family: Spending Functions Futility Boundary Family: None

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.05

Spacing of Looks Equal Unequal Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.200 | 0.000 | 4.877 | -4.877 |
| 2 | 0.400 | 0.001 | 3.357 | -3.357 |
| 3 | 0.600 | 0.008 | 2.680 | -2.680 |
| 4 | 0.800 | 0.024 | 2.290 | -2.290 |
| 5 | 1.000 | 0.050 | 2.031 | -2.031 |

Change the committed accrual to the minimum: 2531. Click on **Compute** button to

get the design shown below.

| | Des5 | Des6 |
|---|-----------------------|-----------------------|
| Mirrored | SU-ZS-LRKR | SU-ZS-LRKR |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 5 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.901 | 0.901 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Hazard Ratio (Alt.) | 0.673 | 0.673 |
| Var (Log HR) | Null | Null |
| Boundary Parameters | | |
| Spacing of Looks | | Equal |
| Efficacy Boundary | | LD (OF) |
| Accrual & Dropout Parameters | | |
| Accrual Rate | 51.5 | 51.5 |
| Subjects are Followed | For Fixed Period (12) | For Fixed Period (12) |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 0 | 0 |
| Sample Size | | |
| Maximum | 2474 | 2531 |
| Expected Under H0 | 2395.697 | 2428.535 |
| Expected Under H1 | 2474 | 2103.71 |
| Events | | |
| Maximum | 268 | 274 |
| Expected Under H0 | 268 | 272.203 |
| Expected Under H1 | 268 | 202.949 |
| Study Duration | | |
| Maximum | 60.039 | 61.146 |
| Expected Under H0 | 46.518 | 47.156 |
| Expected Under H1 | 60.039 | 43.892 |
| Accrual Duration | | |
| Maximum | 48.039 | 49.146 |
| Expected Under H0 | 46.518 | 47.156 |
| Expected Under H1 | 48.039 | 40.849 |

We note that the 5-look design requires an up-front commitment of 274 events compared to the 268 events for the single-look design. At an enrollment rate of 51.5 subjects/month we need to enroll 2531 subjects over 49.1 months. The maximum study duration is expected to be 61.1 months, only 1.1 months longer than the single-look design. However, because of the possibility of early stopping the expected study duration, under the alternative hypothesis that the negative of the log hazard ratio is $\ln(0.673) = -0.397$, is only 43.9 months a savings of more than 16 months.

45.4.1 Incorporating Drop-Outs

The sample size will have to be increased appropriately if we expect drop-outs. Suppose we expect a drop out rate of 0.05 by 12 months for each treatment arm. Edit Des6 and enter the drop-out rates in the appropriate design dialog box as shown below.

45 Superiority Trials with Fixed Follow-Up

Change the committed accrual to the minimum: 2595.

Test Parameters
Boundary
Accrual / Dropouts

Subjects are followed: For Fixed Period 12

Accrual Info

of Accrual Periods: 1

Accrual Rate: 51.5

Piecewise Dropout Information

of Pieces: 1 Input Method: Prob. of Dropout

By Time: 12

Prob. of Dropout (Control) 0.05

Prob. of Dropout (Treatment) 0.05

Accrual

| | Min. | Comtd. | Sugg. Max. |
|--|--------|--------|------------|
| <input type="radio"/> Duration: | 50.388 | 50.388 | 56.175 |
| <input checked="" type="radio"/> Subjects: | 2595 | 2595 | 2893 |

Note: Period 1 hazard rates apply after time 12.

Click on **Compute**. Now you will get Des7 as shown below.

| | Des6 | Des7 |
|---|-----------------------|-----------------------|
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 5 | 5 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.901 | 0.901 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Hazard Ratio (Alt.) | 0.673 | 0.673 |
| Var (Log HR) | Null | Null |
| Boundary Parameters | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) |
| Accrual & Dropout Parameters | | |
| Accrual Rate | 51.5 | 51.5 |
| Subjects are Followed | For Fixed Period (12) | For Fixed Period (12) |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 0 | 1 |
| Sample Size | | |
| Maximum | 2531 | 2595 |
| Expected Under H0 | 2428.535 | 2479.626 |
| Expected Under H1 | 2103.71 | 2149.292 |
| Events | | |
| Maximum | 274 | 274 |
| Expected Under H0 | 272.203 | 272.203 |
| Expected Under H1 | 202.949 | 202.949 |
| Study Duration | | |
| Maximum | 61.146 | 62.388 |
| Expected Under H0 | 47.156 | 48.148 |
| Expected Under H1 | 43.892 | 44.777 |
| Accrual Duration | | |
| Maximum | 49.146 | 50.388 |
| Expected Under H0 | 47.156 | 48.148 |
| Expected Under H1 | 40.849 | 41.734 |

The 5% drop-out rate has resulted in a sample size increase from 2531 subjects to 2595

subjects. However, the impact on maximum study duration and expected study duration is small. Under the alternative hypothesis the study is expected to last for 44.7 months in Des7 as compared to 43.9 months in Des6.

45.4.2 Incorporating Non-Constant Accrual Rates

Des7 was designed with the assumption that patients would be enrolled at the rate of 51.5/month. Suppose that this enrollment rate cannot be achieved from the get-go. Instead, assume that for the first 12 months patients are enrolled at an average rate of 25/month and thereafter the average enrollment rate is 51.5/month. To see the impact of this change on the study design, edit Des7, enter the two enrollment rates into the appropriate dialog box as shown below and create Des8.

Subjects are followed: For Fixed Period 12

Accrual Info

of Accrual Periods: 2

| Period # | Starting at Time | Accrual Rate |
|----------|------------------|--------------|
| 1 | 0.000 | 25 |
| 2 | 12.000 | 51.5 |

Accrual

| | Min. | Comtd. | Sugg. Max. |
|--|--------|--------|------------|
| <input type="radio"/> Duration: | 56.563 | 56.563 | 62.35 |
| <input checked="" type="radio"/> Subjects: | 2595 | 2595 | 2893 |

45 Superiority Trials with Fixed Follow-Up

Click on **Compute** button to create Des8 as shown below.

| | Des7 | Des8 |
|---|-----------------------|-----------------------|
| Mnemonic | SU-Z3-ERAK | SU-Z3-ERAK |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 5 | 5 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.901 | 0.901 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Hazard Ratio (Alt.) | 0.673 | 0.673 |
| Var (Log HR) | Null | Null |
| Boundary Parameters | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) |
| Accrual & Dropout Parameters | | |
| Accrual Rate | 51.5 | Multiple |
| Subjects are Followed | For Fixed Period (12) | For Fixed Period (12) |
| No. of Accrual Periods | 1 | 2 |
| No. of Dropout Pieces | 1 | 1 |
| Sample Size | | |
| Maximum | 2595 | 2595 |
| Expected Under H0 | 2479.626 | 2479.627 |
| Expected Under H1 | 2149.292 | 2149.291 |
| Events | | |
| Maximum | 274 | 274 |
| Expected Under H0 | 272.203 | 272.203 |
| Expected Under H1 | 202.949 | 202.949 |
| Study Duration | | |
| Maximum | 62.388 | 68.563 |
| Expected Under H0 | 48.148 | 54.323 |
| Expected Under H1 | 44.777 | 50.951 |
| Accrual Duration | | |
| Maximum | 50.388 | 56.563 |
| Expected Under H0 | 48.148 | 54.323 |
| Expected Under H1 | 41.734 | 47.909 |

The total sample size has not changed between Des7 and Des8. However, the total duration of the enrollment phase has increased by about six months. Moreover, because of the slower enrollment rate for the first 12 months, the maximum total study duration has increased from 62.4 months to 68.6 months and the expected study duration under the alternative hypothesis has increased from 44.7 months to 50.9 months.

45.4.3 Incorporating Piece-Wise Exponential Survival

Suppose that the mechanism of action of the stents is such that the hazard rate for failure decreases after the first six months. We will assume that the average hazard rate for the uncoated stents arm is 0.0114 for the first six months and decreases thereafter to an average rate of 0.0075. We will continue to assume that the hazard ratio is unchanged, at 0.673. Therefore the hazard rate for the coated stents arm decreases from $0.673 * 0.0114 = 0.0077$ to $0.673 * 0.0075 = 0.005$. Edit Des8 as shown below

to create Des9.

of Hazard Pieces: Input Method:

Hazard Ratio

Hazard Ratio (λ_t/λ_c) Alternative

Log Hazard Ratio $\ln(\lambda_t/\lambda_c)$

| Period # | Starting at Time | Hazard Rate (Control) | Hazard Rate (Treatment: Alt.) |
|----------|------------------|-----------------------|-------------------------------|
| 1 | 0.000 | 0.0114 | 0.008 |
| 2 | 6.000 | 0.0075 | 0.005 |

Change the committed accrual to the minimum: 3095, and click **Compute**. Now you will get the edited Des9 as shown below.

| | Des 8 | Des 9 |
|---|-----------------------|-----------------------|
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 5 | 5 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.901 | 0.901 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Hazard Ratio (Alt.) | 0.673 | 0.673 |
| Var (Log HR) | Null | Null |
| Boundary Parameters | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) |
| Accrual & Dropout Parameters | | |
| Accrual Rate | Multiple | Multiple |
| Subjects are Followed | For Fixed Period (12) | For Fixed Period (12) |
| No. of Accrual Periods | 2 | 2 |
| No. of Dropout Pieces | 1 | 1 |
| Sample Size | | |
| Maximum | 2595 | 3095 |
| Expected Under H0 | 2479.627 | 2865.504 |
| Expected Under H1 | 2149.291 | 2497.442 |
| Events | | |
| Maximum | 274 | 274 |
| Expected Under H0 | 272.203 | 272.203 |
| Expected Under H1 | 202.949 | 202.949 |
| Study Duration | | |
| Maximum | 68.563 | 78.272 |
| Expected Under H0 | 54.323 | 61.816 |
| Expected Under H1 | 50.951 | 57.712 |
| Accrual Duration | | |
| Maximum | 56.563 | 66.272 |
| Expected Under H0 | 54.323 | 61.816 |
| Expected Under H1 | 47.909 | 54.669 |

Since the hazard ratio is unchanged, Des8 and Des9 require the same number of events, 274, in order to achieve the desired 90% power. Observe, however, that in order for these 274 events to arrive on average 12 months after the last patient has enrolled, Des9

45 Superiority Trials with Fixed Follow-Up

requires a sample size of 3095 subjects; 500 patients more than were required under Des8. The study duration is likewise prolonged. This is because the hazard rate slows down after the first six months on study. If, for example, the change in hazard rate were to occur after 12 months instead of after 6 months, the change would have no impact on sample size or study duration. To verify this, make the following change in Des9:

of Hazard Pieces: Input Method:

Hazard Ratio

Hazard Ratio (λ_t/λ_c) Alternative

Log Hazard Ratio $\ln(\lambda_t/\lambda_c)$

| Period # | Starting at Time | Hazard Rate (Control) | Hazard Rate (Treatment: Alt.) |
|----------|------------------|-----------------------|-------------------------------|
| 1 | 0.000 | 0.0114 | 0.008 |
| 2 | 12.000 | 0.0075 | 0.005 |

Change the committed accrual to the minimum: 2595, and click on **Compute**. You will see Des10 details as shown below.

| | Wbk1:Des8 | Wbk1:Des9 | Wbk1:Des10 |
|---|-----------------------|-----------------------|-----------------------|
| Mnemonic | SU-25-LRAR | SU-25-LRAR | SU-25-LRAR |
| Test Parameters | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 5 | 5 | 5 |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.901 | 0.901 | 0.901 |
| Model Parameters | | | |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Hazard Ratio (Alt.) | 0.673 | 0.673 | 0.673 |
| Var (Log HR) | Null | Null | Null |
| Boundary Parameters | | | |
| Spacing of Looks | Equal | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) |
| Accrual & Dropout Parameters | | | |
| Accrual Rate | Multiple | Multiple | Multiple |
| Subjects are Followed | For Fixed Period (12) | For Fixed Period (12) | For Fixed Period (12) |
| No. of Accrual Periods | 2 | 2 | 2 |
| No. of Dropout Pieces | 1 | 1 | 1 |
| Sample Size | | | |
| Maximum | 2595 | 3093 | 2595 |
| Expected Under H0 | 2479.627 | 2863.6 | 2479.627 |
| Expected Under H1 | 2149.291 | 2495.803 | 2149.291 |
| Events | | | |
| Maximum | 274 | 274 | 274 |
| Expected Under H0 | 272.203 | 272.203 | 272.203 |
| Expected Under H1 | 202.949 | 202.949 | 202.949 |
| Study Duration | | | |
| Maximum | 68.563 | 78.233 | 68.563 |
| Expected Under H0 | 54.323 | 61.779 | 54.323 |
| Expected Under H1 | 50.951 | 57.68 | 50.951 |
| Accrual Duration | | | |
| Maximum | 56.563 | 66.233 | 56.563 |
| Expected Under H0 | 54.323 | 61.779 | 54.323 |
| Expected Under H1 | 47.909 | 54.637 | 47.909 |

Notice that Des8 and Des10 are now identical.

45.5 Verification by Simulation

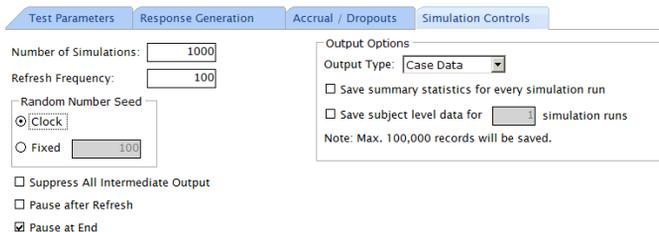
Click on the Des10 node in the **Library** click on  from the toolbar. A simulation input window comprising of four tabs - **Test Parameters**, **Response Generation**, **Accrual/Dropouts** and **Simulation Controls** is now invoked.

We will run simulations under different assumptions about the manner in which the data are generated.

45.5.1 Simulation Under the Alternative Hypothesis

We first run the simulations without making any changes to the default settings of the simulation input tabs. To see the default inputs for the simulations, click the tabs mentioned above.

Change the number of simulations to be run as 1000 on the last tab, **Simulation Control** and click **Simulate** button to run the simulations.



The screenshot shows the 'Simulation Controls' tab of a simulation input window. It contains the following settings:

- Number of Simulations: 1000
- Refresh Frequency: 100
- Random Number Seed:
 - Clock
 - Fixed: 100
- Suppress All Intermediate Output
- Pause after Refresh
- Pause at End
- Output Options:
 - Output Type: Case Data
 - Save summary statistics for every simulation run
 - Save subject level data for [] simulation runs
 - Note: Max. 100,000 records will be saved.

An entry for the simulation output gets added in the **Output Preview** pane. Save it in the workbook. Since we simulated Des10, a node named Sim1 will get associated with Des10. Double click on this node and see the detailed simulation output.

45 Superiority Trials with Fixed Follow-Up

Let us examine the results. Notice first that we have selected the option to fix the number of events for each look at their pre-planned values in the **Look Information** section. This can be seen in the **Simulation Parameters** tab.

Fix at Each Look: Total No. of Events ▾

Upon examining the simulation results in detail, however, we observe that the actual number of events at the final look is slightly lower than the pre-planned number of 274.

☉ Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 816.757 | 32.647 | 22.353 | 12.395 | 12.38 | 22.04 | 7.042 |
| 2 | 1344.949 | 65.138 | 44.862 | 24.692 | 24.893 | 32.303 | 8.578 |
| 3 | 1860.36 | 96.361 | 67.639 | 36.735 | 36.902 | 42.293 | 9.261 |
| 4 | 2368.408 | 125.772 | 93.228 | 48.63 | 49.139 | 52.228 | 9.661 |
| 5 | 2594.008 | 150.064 | 118.223 | 59.426 | 59.985 | 64.739 | 10.784 |
| Average | 2176.262 | 121.205 | 83.385 | 45.596 | 46.244 | 50.637 | 9.739 |

This is observed consistently. If you edit this simulation node and simulate this scenario again and again with different starting seeds, you will notice that the actual number of events at which the first four looks are taken match the corresponding pre-planned values, whereas there appears to be a systematic bias towards taking the fifth and final look with slightly fewer events than was pre-planned. As a result the trial is slightly underpowered.

In practice, the slight loss of power due to the systematic decrease in the number of events at the final look relative to the pre-planned number is of very little consequence. It is instructive, however, to understand why it arises at all. The reason for the small amount of systematic bias is that the maximum follow-up time for each patient is 12 months. Thus, no further follow-up is possible once 12 months have elapsed after the last subject has enrolled. Since the duration of the enrollment period has been fixed at 56.5 months, the trial must be terminated at the latest in $56.5 + 12 = 68.5$ months. Observe that the selection in the **Simulation Parameters** tab, that in the row titled Fixed at Each Look, the choice **Total No. of Events** has been selected from the drop down list.

This means that East has been instructed to perform simulations in which each look is taken after a fixed number of events has been observed, as pre-specified in the design. Specifically, the looks should be taken after 55, 110, 164, 219 and 274 events have been observed. The trial should be terminated early if a boundary is crossed at one of the

first four looks; otherwise it should continue until all 274 events have been obtained.

☰ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Events | Boundaries | | Stopping For | | Total Simulations | |
|--------|--------|------------|--------|----------------|----------------|-------------------|---------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 55 | 4.868 | -4.868 | 0 | 0 | 0 | 0.000% |
| 2 | 110 | 3.35 | -3.35 | 0 | 84 | 84 | 8.400% |
| 3 | 164 | 2.684 | -2.684 | 0 | 333 | 333 | 33.300% |
| 4 | 219 | 2.291 | -2.291 | 0 | 318 | 318 | 31.800% |
| 5 | 274 | 2.031 | -2.031 | 0 | 99 | 265 | 26.500% |
| Total | | | | 0 | 834 | 1000 | |
| % | | | | 0.000% | 83.400% | | |

We thus have two conflicting restrictions for the maximum study duration. The fixed follow-up design implies that the trial cannot proceed beyond month 68.5 whereas the **Planned # of Events** restriction implies that the trial cannot proceed beyond 274 events. East resolves the conflict by fixing the maximum study duration at the **earlier** of 68.5 months or the time at which 274 events have been observed. Thereby the average number of events at the fifth and final look becomes a random variable with an upper bound of 274 and an expected value that is slightly less than 274.

To get around this bias one should specify in the **Look Information** section that we will **Fix at Each Look** the **Look Time** rather than the **Total No. of Events**.

Fix at Each Look:

| Look # | Analysis Ti... | Efficacy Z | |
|--------|----------------|------------|--------|
| | | Upper | Lower |
| 1 | 22.070 | 4.868 | -4.868 |
| 2 | 32.283 | 3.350 | -3.350 |
| 3 | 42.238 | 2.684 | -2.684 |
| 4 | 52.377 | 2.291 | -2.291 |
| 5 | 68.563 | 2.031 | -2.031 |

With this specification the looks will occur at fixed calendar times of 22.1, 32.3, 42.2, 52.4, and 68.5 months regardless of the number of events that have been obtained at these looks. Although the actual number of events obtained at each of these five looks are now random variables, the **average** number of events obtained in repeated simulations will be 55, 110, 164, 219 and 274, respectively, under the alternative hypothesis. The bias due to fixing the maximum number of events at 274 will no longer occur and the study will be fully powered. To see this, run the simulations

45 Superiority Trials with Fixed Follow-Up

10,000 times after fixing the **Look Time** rather than the **# of Events**

The appropriate number of events is obtained at each look on average and the study is fully powered, up to Monte Carlo accuracy.

45.5.2 Simulation Under the Null Hypothesis

It is important to verify by simulation that the type-1 error is preserved. Accordingly, edit the node Sim2 and switch to the **Response Generation** tab. We may now make changes to the design by editing the entries in the cells that are white in color.

| Survival Information | | | | |
|---------------------------------------|------------------|----------------------------|-----------|--------------|
| # of Hazard Pieces | 2 | Input Method: Hazard Rates | | |
| <input type="checkbox"/> Hazard Ratio | | | | |
| Piece | Starting at Time | Hazard Rates | | Hazard Ratio |
| | | Control | Treatment | |
| 1 | 0.000 | 0.0114 | 0.0077 | 0.673 |
| 2 | 12.000 | 0.0075 | 0.0050 | 0.673 |

To simulate under the null hypothesis we must set the hazard rates of the Control and Treatment groups to be the same, as shown below:

| Survival Information | | | | |
|---------------------------------------|------------------|----------------------------|-----------|--------------|
| # of Hazard Pieces | 2 | Input Method: Hazard Rates | | |
| <input type="checkbox"/> Hazard Ratio | | | | |
| Piece | Starting at Time | Hazard Rates | | Hazard Ratio |
| | | Control | Treatment | |
| 1 | 0.000 | 0.0114 | 0.0114 | 1.000 |
| 2 | 12.000 | 0.0075 | 0.0075 | 1.000 |

Then click on the **Simulate** button to generate 10000 simulated trials under the null

hypothesis.

⊖ **Simulation Boundaries and Boundary Crossing Probabilities**

| Look # | Look Time | Boundaries | | Stopping For | | Total Simulations | |
|--------|-----------|------------|--------|----------------|----------------|-------------------|---------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 22.07 | 4.868 | -4.868 | 0 | 0 | 0 | 0.000% |
| 2 | 32.283 | 3.35 | -3.35 | 3 | 5 | 8 | 0.080% |
| 3 | 42.238 | 2.684 | -2.684 | 33 | 36 | 69 | 0.690% |
| 4 | 52.377 | 2.291 | -2.291 | 85 | 69 | 154 | 1.540% |
| 5 | 68.563 | 2.031 | -2.031 | 140 | 111 | 9769 | 97.690% |
| Total | | | | 261 | 221 | 10000 | |
| % | | | | 2.610% | 2.210% | | |

The type-1 error has been preserved, with an overall two-sided false positive rate less than 5%. The above simulations were run with fixed look times rather than with fixed numbers of events at each look. It is interesting to note that, for the same fixed look times, the average number of events at each look under the null hypothesis greatly exceeds the corresponding average number of events at each look under the alternative hypothesis. This is so because the events arrive faster when the treatment arm is no more effective than the control arm.

46 *Non-Inferiority Trials Given Accrual Duration and Accrual Rates*

This chapter will illustrate through a worked example how to design, monitor and simulate a two-sample non-inferiority trial with a time-to-event trial endpoint, when the accrual duration and accrual rates are fixed.

46.1 *Establishing the Non-Inferiority Margin*

The first step in designing a non-inferiority trial is to establish a suitable non-inferiority margin. This is typically done by performing a meta-analysis on past clinical trials of the active control versus placebo. Regulatory agencies then require the sponsor of the clinical trial to demonstrate that a fixed percentage of the active control effect (usually 50%) is retained by the new treatment. A further complication arises because the active control effect can only be estimated with error. We illustrate below with an example provided by reviewers at the FDA.

Rothman et al. (2003) have discussed a clinical trial to establish the non-inferiority of the test drug Xeloda (treatment t) relative to the active control (treatment c) consisting of 5-fluorouracil with leucovorin (5FU+LV) for metastatic colorectal cancer. In order to establish a suitable non-inferiority margin for this trial it is necessary to first establish the effect of 5FU+LV relative to the reference therapy of 5FU alone (treatment p , here regarded as placebo). To establish this effect the FDA conducted a ten-study random effects meta-analysis (FDA Medical-Statistical review for Xeloda, NDA 20-896, April 2001) of randomized comparisons of 5-FU alone versus 5-FU+LV. Letting λ_t , λ_c and λ_p denote the constant hazard rates for the new treatment, the active control and the placebo, respectively, the FDA meta-analysis established that

$$\ln(\widehat{\lambda_p/\lambda_c}) = 0.234$$

with standard error

$$\text{se}[\ln(\widehat{\lambda_p/\lambda_c})] = 0.075 .$$

Thus with 100 γ % confidence the active control effect lies inside the interval

$$\left[0.234 - 0.075\Phi^{-1}\left(\frac{1+\gamma}{2}\right), 0.234 + 0.075\Phi^{-1}\left(\frac{1+\gamma}{2}\right)\right] \quad (46.1)$$

The new study is required to demonstrate that some fraction (usually 50%) of the active control effect is retained. Rothman et al. (2003) state that the claim of non-inferiority for the new treatment relative to the active control can be demonstrated if the upper limit of a two-sided 100(1 - α)% confidence interval for $\ln(\lambda_t/\lambda_c)$ is less than a pre-specified fraction of the lower limit of a two-sided 100 γ % confidence interval for the active control effect established by the meta-analysis. This is known as

the “two confidence intervals procedure”. Specifically in order to claim non-inferiority in the current trial it is necessary to show that

$$\begin{aligned} & \ln(\widehat{\lambda_t/\lambda_c}) + \Phi^{-1}(1 - \alpha/2)\text{se}[\ln(\widehat{\lambda_t/\lambda_c})] \\ & < (1 - f_0)\{\ln(\widehat{\lambda_p/\lambda_c}) - \Phi^{-1}(\frac{1 + \gamma}{2})\text{se}[\ln(\widehat{\lambda_p/\lambda_c})]\} . \end{aligned} \quad (46.2)$$

We may re-write the non-inferiority condition (46.2) in terms of a one-sided Wald test of the form

$$\frac{\ln(\widehat{\lambda_t/\lambda_c}) - \delta_0}{\text{se}[\ln(\widehat{\lambda_t/\lambda_c})]} < \Phi^{-1}(1 - \alpha/2) , \quad (46.3)$$

where

$$\delta_0 = (1 - f_0)\{\ln(\widehat{\lambda_p/\lambda_c}) - \Phi^{-1}(\frac{1 + \gamma}{2})\text{se}[\ln(\widehat{\lambda_p/\lambda_c})]\} \quad (46.4)$$

is the non-inferiority margin.

The choice $f_0 = 1$ implies that the entire active control effect must be retained in the new trial and amounts to running a superiority trial. At the other end of the spectrum, the choice $f_0 = 0$ implies that none of the active control effect need be retained; i.e., the new treatment is only required to demonstrate effectiveness relative to placebo. The usual choice is $f_0 = 0.5$, implying that the new treatment is required to retain at least 50% of the active control effect. The usual choice for α is $\alpha = 0.05$. A conservative choice for the coefficient γ is $\gamma = (1 - \alpha) = 0.95$. Rothman et al. (2003) refer to this method of establishing the non-inferiority margin as the “two 95 percent two-sided confidence interval procedure” or the “95-95 rule”. In general this approach leads to rather tight margins unless the active control effect is substantial. Rothman et al. (2003) have also proposed more lenient margins that vary with the amount of power desired. Fleming (2007), however, argues for the stricter 95-95 rule on the grounds that it offers greater protection against an ineffective medical compound being approved in the event that the results of the previous trials used to establish the active control effect are of questionable relevance to the current setting. Accordingly we evaluate (46.4) with $\gamma = 0.95$, $f_0 = 0.5$, $\ln(\widehat{\lambda_p/\lambda_c}) = 0.234$ and $\text{se}[\ln(\widehat{\lambda_p/\lambda_c})] = 0.075$ thereby obtaining the non-inferiority margin to be $\delta_0 = 0.044$ for the log hazard ratio and $\exp(0.044) = 1.045$ for the hazard ratio.

46 Non-Inferiority Trials Given Accrual Duration and Accrual Rates

46.2 Design of Metastatic Colorectal Cancer Trial

46.2.1 Single-Look Design

46.2.2 Early Stopping for Futility

In this section we will use East to design a single-look non-inferiority trial comparing the test drug Xeloda (treatment t) to the active control 5FU+LV (treatment c) for the treatment of metastatic colorectal cancer. On the basis of a meta-analysis of ten previous studies of the active control versus placebo (Rothman et al., 2003), a non-inferiority margin of 1.045 for λ_t/λ_c has been established. Thus we are interested in testing the null hypothesis of inferiority $H_0: \lambda_t/\lambda_c \geq 1.045$ versus the one-sided alternative hypothesis that $\lambda_t/\lambda_c < 1.045$. Subjects are expected to enroll at the rate of 60/month and the median survival time for patients randomized to the active control arm is expected to be 18 months.

46.2.1 Single-Look Design

We will use East to create an initial single-look design having 80% power to detect the alternative hypothesis $H_1: \lambda_t/\lambda_c = 1$ with a one sided level 0.025 non-inferiority test.

To begin click **Survival: Two Samples** on the **Design** tab and then click **Parallel Design: Log Rank Test Given Accrual Duration and Accrual Rates**.

A new screen will appear. Enter the appropriate design parameters into the dialog box as shown below.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

Design Type: Noninferiority Number of Looks: 1

Test Parameters Accrual / Dropouts

Test Type: 1-Sided # of Hazard Pieces: 1 Input Method: Median Survival Times

Hazard Ratio (Optional)

| | Null | Alternative |
|---|-------|-------------|
| <input checked="" type="radio"/> Hazard Ratio (λ_t/λ_c) | 1.045 | 1 |
| <input type="radio"/> Ratio of Medians (m_t/m_c) | 0.957 | 1 |

| | Med.Surv.Time | |
|---------|-----------------|-----------------|
| Control | Treatment: Null | Treatment: Alt. |
| 18 | 17.225 | 18 |

Assurance (Probability of Success)

Variance of Log Hazard Ratio

Null Alternative

The box labeled **Variance of Log Hazard Ratio** specifies whether the calculation of the required number of events is to be based on the variance estimate of the log hazard ratio under the null hypothesis or the alternative hypothesis. The default choice in East is **Null**. Most textbooks recommend this choice as well (see, for example Collett,

1994, equation (2.21) specialized to no ties). It will usually not be necessary to change this default. For a technical discussion of this issue refer to Appendix B, Section B.5.3.

Next click on the **Accrual/Dropouts** tab. Here we will specify the accrual information and dropout rates. Enter an accrual rate of 60. Suppose that there are 5% drop-outs per year in each arm. Enter these values as shown below.

Test Parameters | **Accrual / Dropouts**

Subjects are followed:

Accrual Info

of Accrual Periods:

Accrual Rate:

Piecewise Dropout Information

of Pieces: Input Method:

By Time:

Prob. of Dropout (Control)

Prob. of Dropout (Treatment)

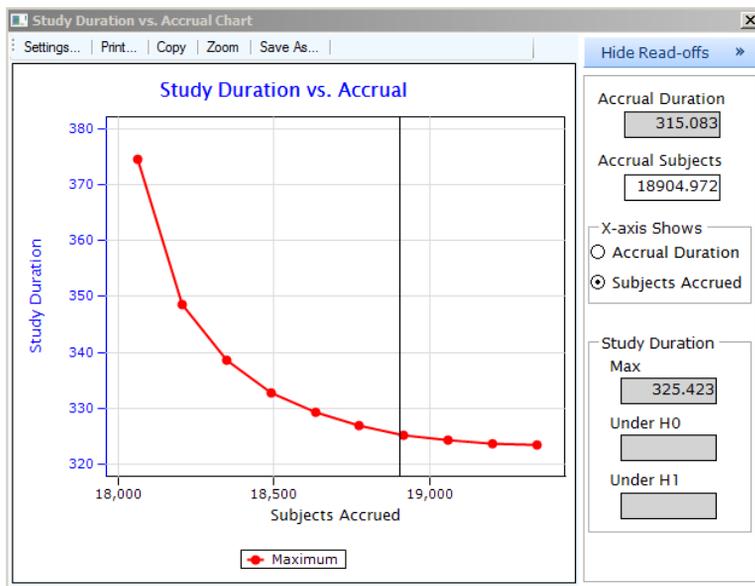
Note: Period 1 hazard rates apply after time 12.

| Accrual | | Min. | Comtd. | Sugg. Max. |
|--|--|--------|---------|------------|
| <input type="radio"/> Duration: | | 300.05 | 311.733 | 323.4 |
| <input checked="" type="radio"/> Subjects: | | 18003 | 18704 | 19404 |

On the bottom of this screen is where you can specify the accrual duration or number of subjects. East automatically computes a range that is necessary to achieve the desired power of the study and selects the midpoint of the range, as the committed accrual duration or subjects. If your study has a restriction on accrual duration or subject accrual, you may enter this value in the **Comtd.** column. In our example, East computes a minimum accrual duration of 300.05 months and a suggested maximum of 323.4 months. Also, if you click the  icon a chart which shows the relationship between accrual duration (or subject accrual, depending on whether you choose to

46 Non-Inferiority Trials Given Accrual Duration and Accrual Rates

specify accrual duration or subject accrual) and study duration.



Looking at this chart, choosing an accrual duration longer than 315 months will not result in a substantial decrease in study duration. Thus, we commit to an accrual duration of 315 months. Close this chart, select the radio button next to **Duration** and enter 315 in the **Comtd.** column.

Click on **Compute** to complete the design. The design is shown as a row in the **Output Preview** located in the lower pane of this window. You can select this design by clicking anywhere along the row in the **Output Preview**. With Des1 selected, click the  icon to display the details of this design in the upper pane, which are shown below. You may also wish to save this design. Select Des1 in the **Output Preview**

window and click the  to save this design to Workbook1 in the **Library**.

| Des 1 | |
|---|--------------------|
| Mnemonic | SU-25-LRAR |
| Test Parameters | |
| Design Type | Noninferiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.8 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Null) | 1.045 |
| Hazard Ratio (Alt.) | 1 |
| Var (Log HR) | Null |
| Accrual & Dropout Parameters | |
| Accrual Rate | 60 |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 1 |
| Sample Size | |
| Maximum | 18900 |
| Expected Under H0 | 18900 |
| Expected Under H1 | 18900 |
| Events | |
| Maximum | 16205 |
| Expected Under H0 | 16205 |
| Expected Under H1 | 16205 |
| Study Duration | |
| Maximum | 325.466 |
| Expected Under H0 | 323.839 |
| Expected Under H1 | 325.466 |
| Accrual Duration | |
| Maximum | 315 |
| Expected Under H0 | 315 |
| Expected Under H1 | 315 |

It is immediately evident that Des1 is untenable. It requires 16,205 events to be fully powered. The problem lies with trying to power the trial to detect a hazard ratio of 1 under the alternative hypothesis. Suppose instead that the investigators actually believe that the treatment is slightly superior to the active control, but the difference is too small to be detected in a superiority trial. In that case a non-inferiority design powered at a hazard ratio less than 1 (0.95, say) would be a better option because such a trial would require fewer events.

To see this create a new design by selecting Des1 in the **Library**, and clicking the  icon on the **Library** toolbar. Then edit this design by specifying a hazard ratio of 0.95

46 Non-Inferiority Trials Given Accrual Duration and Accrual Rates

under the alternative hypothesis as shown below.

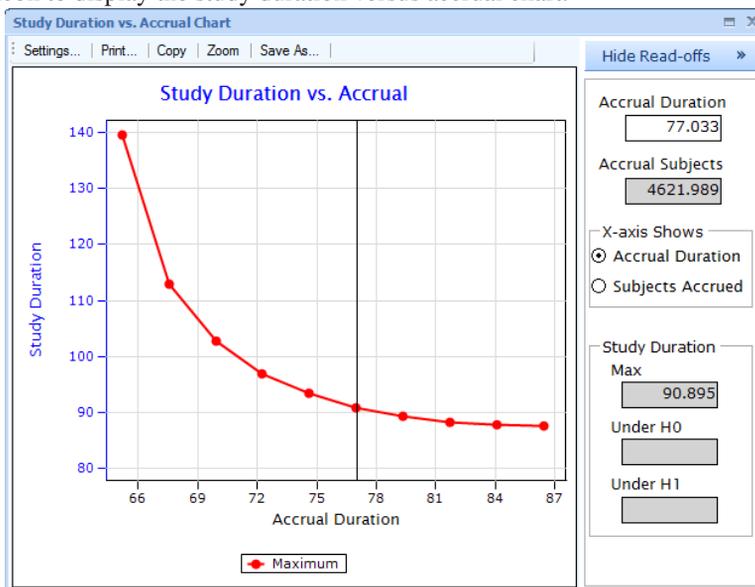
of Hazard Pieces: Input Method:

Hazard Ratio (Optional)

| | Null | Alternative |
|---|------------------------------------|------------------------------------|
| <input checked="" type="radio"/> Hazard Ratio (λ_t/λ_c) | <input type="text" value="1.045"/> | <input type="text" value="0.95"/> |
| <input type="radio"/> Ratio of Medians (m_t/m_c) | <input type="text" value="0.957"/> | <input type="text" value="1.053"/> |

| | Med.Surv.Time | |
|---------------------------------|-------------------------------------|-------------------------------------|
| Control | Treatment:Null | Treatment: Alt. |
| <input type="text" value="18"/> | <input type="text" value="17.225"/> | <input type="text" value="18.947"/> |

Next, click on the **Accrual/Dropouts** tab. Notice that the minimum and suggested maximum accrual have changed to 64.167 and 87.45 months, respectively. Click the  icon to display the study duration versus accrual chart.



Suppose that after examining this chart, you decide that an accrual duration longer than 77 months is not worth the small decrease in study duration one would gain from a longer accrual duration. Close this chart. Select the radio button next to **Duration** and

enter 77 in the **Comtd.** column.

Subjects are followed: Until End of Study

Accrual Info

of Accrual Periods: 1

Accrual Rate: 60

Piecewise Dropout Information

of Pieces: 1 Input Method: Prob. of Dropout

By Time: 12

Prob. of Dropout (Control) 0.05

Prob. of Dropout (Treatment) 0.05

Note: Period 1 hazard rates apply after time 12.

Accrual

| | Min. | Comtd. | Sugg. Max. |
|-------------|--------|--------|------------|
| ○ Duration: | 64.167 | 77 | 87.45 |
| ○ Subjects: | 3850 | 4620 | 5247 |

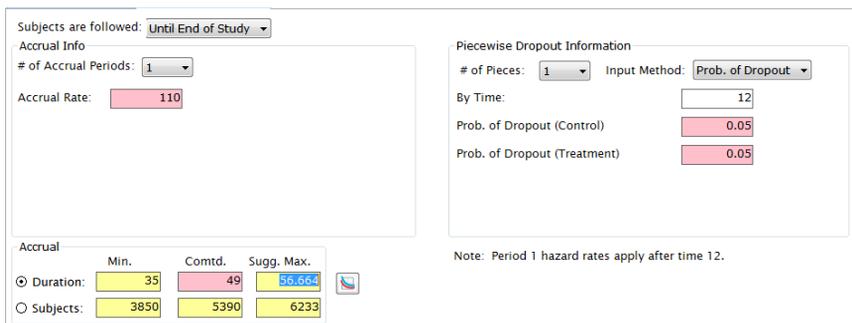
Click the **Compute** button to generate output for Des2. With Des2 selected in the **Output Preview**, click the icon to save Des2 to the **Library**. In the **Library**, select the rows for Des1 and Des2, by holding the Ctrl key, and then click the icon. The upper pane will display the details of the two designs side-by-side:

| | Des 1 | Des 2 |
|---|--------------------|--------------------|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| Test Parameters | | |
| Design Type | Noninferiority | Noninferiority |
| No. of Looks | 1 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.8 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Hazard Ratio (Null) | 1.045 | 1.045 |
| Hazard Ratio (Alt.) | 1 | 0.95 |
| Var (Log HR) | Null | Null |
| Accrual & Dropout Parameters | | |
| Accrual Rate | 60 | 60 |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 |
| Sample Size | | |
| Maximum | 18900 | 4620 |
| Expected Under H0 | 18900 | 4620 |
| Expected Under H1 | 18900 | 4620 |
| Events | | |
| Maximum | 16205 | 3457 |
| Expected Under H0 | 16205 | 3457 |
| Expected Under H1 | 16205 | 3457 |
| Study Duration | | |
| Maximum | 325.466 | 90.946 |
| Expected Under H0 | 323.839 | 88.962 |
| Expected Under H1 | 325.466 | 90.946 |
| Accrual Duration | | |
| Maximum | 315 | 77 |
| Expected Under H0 | 315 | 77 |
| Expected Under H1 | 315 | 77 |

46 Non-Inferiority Trials Given Accrual Duration and Accrual Rates

Des2 is clearly easier to implement than Des1. It requires only 3,457 events and 4620 subjects to be fully powered. Also note the marked decrease in study duration under either the null or alternative hypothesis. Nevertheless, Des2 is also unsatisfactory. The maximum study duration for Des2 (accrual plus follow-up) is 90.9 months with 77 months of that amount of time being utilized to enroll 4620 patients. It is necessary to shorten the maximum study duration further. One possible way to shorten the maximum study duration is to increase the rate of enrollment. Suppose that additional sites can be enlisted to enroll patients after the study is activated so that six months later the average rate of enrollment is increased to 110/month. To see the impact of the increased rate of enrollment select Des2 in the **Library**, and click on the  icon on the **Library** toolbar.

Next, click on the **Accrual/Dropouts** tab. Change the accrual rates as shown below.



The screenshot shows a software interface with two main panels. The left panel, titled 'Accrual Info', has a dropdown menu for 'Subjects are followed:' set to 'Until End of Study'. Below it, '# of Accrual Periods:' is set to '1'. The 'Accrual Rate:' is set to '110'. The right panel, titled 'Piecewise Dropout Information', has '# of Pieces:' set to '1' and 'Input Method:' set to 'Prob. of Dropout'. Under 'By Time:', there is a text box with '12'. Below that, 'Prob. of Dropout (Control)' is set to '0.05' and 'Prob. of Dropout (Treatment)' is set to '0.05'. A note at the bottom right says 'Note: Period 1 hazard rates apply after time 12.' At the bottom of the interface, there is an 'Accrual' table with columns 'Min.', 'Comtd.', and 'Sugg. Max.' and rows for 'Duration' and 'Subjects'.

| | Min. | Comtd. | Sugg. Max. |
|-----------|------|--------|------------|
| Duration: | 35 | 49 | 56.664 |
| Subjects: | 3850 | 5390 | 6233 |

Notice how East automatically updates the accrual duration and subject accrual. An accrual duration in the range of 35 to 56.664 months is sufficient to achieve the desired power. Suppose that after examining the study duration versus accrual chart, we decide on an accrual duration of 49 months. Enter 49 in the **Comtd.** column.

Click the **Compute** button to generate output for Des3. With Des3 selected in the **Output Preview**, click the  icon to save Des3 to the **Library**. In the **Library**, select the rows for Des1, Des2, and Des3 by holding the Ctrl key, and then click the



The upper pane will display the details of the three designs side-by-side:

| | Des 1 | Des 2 | Des 3 |
|---|--------------------|--------------------|--------------------|
| Mnemonic | SU-25-LRAR | SU-25-LRAR | SU-25-LRAR |
| Test Parameters | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 1 | 1 |
| Test Type | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 | 0.025 |
| Power | 0.8 | 0.8 | 0.8 |
| Model Parameters | | | |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Hazard Ratio (Null) | 1.045 | 1.045 | 1.045 |
| Hazard Ratio (Alt.) | 1 | 0.95 | 0.95 |
| Var (Log HR) | Null | Null | Null |
| Accrual & Dropout Parameters | | | |
| Accrual Rate | 60 | 60 | 110 |
| Subjects are Followed | Until End of Study | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 | 1 |
| Sample Size | | | |
| Maximum | 18900 | 4620 | 5390 |
| Expected Under H0 | 18900 | 4620 | 5390 |
| Expected Under H1 | 18900 | 4620 | 5390 |
| Events | | | |
| Maximum | 16205 | 3457 | 3457 |
| Expected Under H0 | 16205 | 3457 | 3457 |
| Expected Under H1 | 16205 | 3457 | 3457 |
| Study Duration | | | |
| Maximum | 325.466 | 90.946 | 58.523 |
| Expected Under H0 | 323.839 | 88.962 | 57.158 |
| Expected Under H1 | 325.466 | 90.946 | 58.523 |
| Accrual Duration | | | |
| Maximum | 315 | 77 | 49 |
| Expected Under H0 | 315 | 77 | 49 |
| Expected Under H1 | 315 | 77 | 49 |

Des3 also requires 3457 events. However, because of the faster rate of enrollment the time that it takes to obtain these events is cut down to 58.5 months.

46.2.2 Early Stopping for Futility

Under the null hypothesis Des3, with 3457 events, has an expected study duration of 57.2 months. This is a very long time commitment for a trial that is unlikely to be successful. Therefore it would be a good idea to introduce a futility boundary for possible early stopping. Since we wish to be fairly aggressive about early stopping for futility we will generate the futility boundary from the Gamma(-1) β -spending function. On the other hand, since there is no interest in early stopping for efficacy, we will not use an efficacy boundary.

Create a new design by selecting Des3 in the **Library**, and clicking the  icon on the **Library** toolbar. Change the number of looks from 1 to 3. Next, click on the **Boundary** tab. Enter the parameters as shown below. Be sure to select the **Non-Binding** option. This choice gives us the flexibility to continue the trial even if a futility boundary has been crossed. Data monitoring committees usually want this

46 Non-Inferiority Trials Given Accrual Duration and Accrual Rates

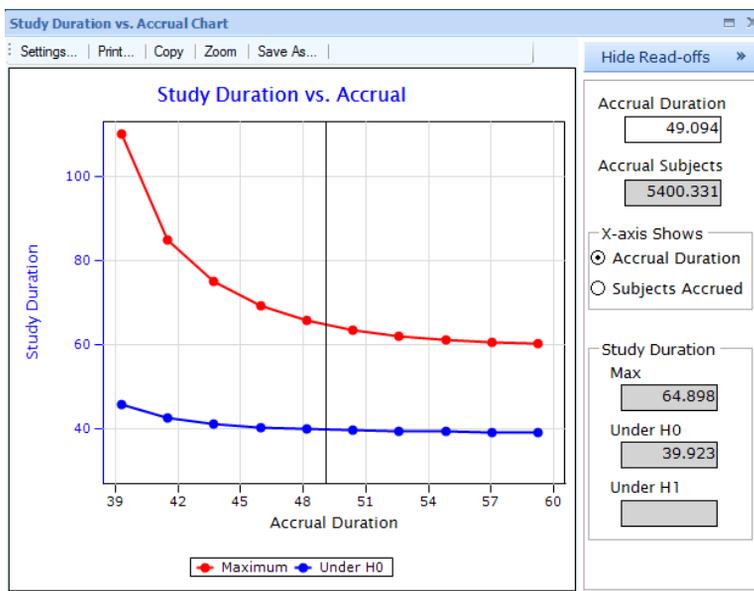
flexibility; for example, to follow a secondary endpoint.

The screenshot shows the 'Accrual / Dropouts' configuration window. It is divided into 'Efficacy' and 'Futility' sections. The 'Efficacy' section has 'Boundary Family' set to 'None'. The 'Futility' section has 'Boundary Family' set to 'Spending Functions', 'Spending Function' set to 'Gamma Family', 'Parameter (γ)' set to '-1', 'Type II Error (β)' set to '0.2', and 'Futility Boundary' set to 'Z Scale'. There are also radio buttons for 'Non-Binding' (selected) and 'Binding'. Below these settings is a table for 'Spacing of Looks'.

| Look # | Info. Fraction | Cum. β Spent | Futility Boundary |
|--------|----------------|--------------------|-------------------|
| 1 | 0.333 | 0.046 | -0.007 |
| 2 | 0.667 | 0.110 | -1.056 |
| 3 | 1.000 | 0.200 | -1.960 |

Next click on the **Accrual/Dropouts** tab. Once again, East automatically computes the minimum and suggested maximum values for the accrual duration and subject accrual.

Click the  icon to display the study duration versus accrual chart. Notice that another line is added to the chart. Now, we can see the maximum study duration vs accrual under the null hypothesis.



Suppose that after examining this chart, you decide to set the accrual duration at 49

months. Any increase in accrual duration past 49 months will not result in a substantial decrease in study duration. Close this chart. Select the radio button for **Duration** and enter 49 in the **Comtd.** column.

Click the **Compute** button to generate output for Des4. With Des4 selected in the **Output Preview**, click the  icon to save Des4 to the **Library**. In the **Library**, select the rows for Des3 and Des4 by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the two designs side-by-side:

| | Des 3 | Des 4 |
|---|--------------------|--------------------|
| Design Type | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Attained α | | 0.022 |
| Power | 0.8 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Hazard Ratio (Null) | 1.045 | 1.045 |
| Hazard Ratio (Alt.) | 0.95 | 0.95 |
| Var (Log HR) | Null | Null |
| Boundary Parameters | | |
| Spacing of Looks | | Unequal |
| Futility Boundary | | Gm (-1) (NB) |
| Accrual & Dropout Parameters | | |
| Accrual Rate | 110 | 110 |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 |
| Sample Size | | |
| Maximum | 5390 | 5390 |
| Expected Under H0 | 5390 | 4140.262 |
| Expected Under H1 | 5390 | 5271.053 |
| Events | | |
| Maximum | 3457 | 3780 |
| Expected Under H0 | 3457 | 2056.327 |
| Expected Under H1 | 3457 | 3583.036 |
| Study Duration | | |
| Maximum | 58.523 | 64.893 |
| Expected Under H0 | 57.158 | 39.546 |
| Expected Under H1 | 58.523 | 62.059 |
| Accrual Duration | | |
| Maximum | 49 | 49 |
| Expected Under H0 | 49 | 37.639 |
| Expected Under H1 | 49 | 47.919 |

Observe that while the maximum study duration has been inflated by about 6 months compared to Des3, the expected study duration under H_0 has been cut down by almost 18 months.

It would be useful to simulate Des4 under a variety of scenarios for the hazard ratio. Select Des4 in the **Library** and click the  icon. You will be taken to the

46 Non-Inferiority Trials Given Accrual Duration and Accrual Rates

following simulation worksheet.

Number of Looks: 3

Test Parameters | **Response Generation** | Accrual / Dropouts | Simulation Controls

Trial Type: Noninferiority
 Test Type: 1-Sided
 Fix at Each Look: Total No. of Events
 Total No. of Events: 3780

Noninf. Margin (ln(HR0)): 0.044
 Test Statistic: Logrank

| Look # | Info. Fraction | Futility Z |
|--------|----------------|------------|
| 1 | 0.333 | -0.007 |
| 2 | 0.667 | -1.056 |
| 3 | 1.000 | -1.960 |

We wish to simulate this trial under the null hypothesis that the hazard ratio is $\exp(0.044) = 1.045$. To this end click on the **Response Generation** tab. In this tab change the hazard ratio to 1.045.

Survival Information

of Hazard Pieces: 1 | Input Method: Median Survival Times

| Med.Surv.Time | | Hazard Ratio |
|---------------|-----------|--------------|
| Control | Treatment | |
| 18 | 17.225 | 1.045 |

Next, click the **Simulate** button to simulate 10000 trials. A new row labeled Sim1 will appear in the **Output Preview** window. Select Sim1 in the **Output Preview** and click the  icon to save it to the **Library**. In the **Library**, double-click Sim1. A portion of the output is displayed below. (The actual values may differ, depending on the

starting seed used).

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

| Test Parameters | |
|---|---------------------|
| Simulation ID | Sim1 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 5390 |
| Fix at Each Look | Total No. of Events |
| Noninferiority Margin (ln(HR0)) | 0.044 |
| Test Statistic | Logrank |
| Average Events | 2851.91 |
| Total Accrual Duration | 49 |
| Avg. Power at Termination | 0.023 |
| Response Generation Parameters | |
| HR = λ_1/λ_2 | 1.045 |
| Med. Surv. Time Control (m _c) | 18 |
| Med. Surv. Time Treatment (m _t) | 17.225 |
| Accrual / Dropouts Parameters | |
| Sample Size | 5390 |
| Subject are followed | Until End of Study |
| Accrual Rate | 110 |
| Prob. of Dropout by Time = | 12 |
| Control | 0.05 |
| Treatment | 0.05 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

☰ Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 3212.729 | 620.437 | 639.563 | 68.744 | 67.875 | 29.202 | 9.962 |
| 2 | 4961.993 | 1249.715 | 1270.285 | 137.006 | 136.414 | 45.102 | 12.899 |
| 3 | 5390 | 1883.993 | 1896.007 | 205.138 | 205.779 | 63.139 | 17.812 |
| Average | 4136.097 | 1009.548 | 1042.362 | 111.995 | 110.712 | 39.485 | 12.074 |

☰ Simulation Boundaries and Cumulative Boundary Crossing Probabilities

| Look # | Events | Boundaries | Stopping For | Reject H0 / Unable to Reject H1 | Total Simulations | |
|---------|--------|------------|--------------|---------------------------------|-------------------|----------|
| | | Futility | Futility | | Count | % |
| 1 | 1260 | -0.007 | 5050 | 0 | 5050 | 50.500% |
| 2 | 2520 | -1.056 | 8665 | 0 | 8665 | 86.650% |
| 3 | 3780 | -1.96 | 9766 | 234 | 10000 | 100.000% |
| Total % | | | 97.660% | 2.340% | | |

Simulation Seed and Elapsed Time

Starting Seed: 2272863
 Total Number of Simulations: 10000
 Elapsed Time: 00.01:15

Note that 234 out of the 10000 simulations rejected the null hypothesis when it was true. Thus confirming (up to Monte Carlo accuracy) that this design achieves a type-1 error of 2.5%. Also, observe that 50% of these trials have crossed the futility boundary at the very first interim look after only 29 months of study duration.

46.3 Interim Monitoring

Suppose we have adopted Des4. Let us monitor the trial with the help of the Interim Monitoring Worksheet. Select Des4 in the **Library**, and click the **IM** icon from the Library toolbar. Alternatively, right-click on Des4 and select **Interim Monitoring**. The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs. The bottom section contains four charts, each with a corresponding table to its right. These charts provide graphical and numerical descriptions of the progress of the clinical trial and are useful tools for decision making by a data monitoring committee.

Suppose that the first interim look is taken after observing 1300 events. The observed hazard ratio is 1.15 and the standard error of the log hazard ratio is 0.06. Enter this information into the interim monitoring worksheet using Test Statistic calculator. Click

46 Non-Inferiority Trials Given Accrual Duration and Accrual Rates

on **Enter Interim Data** and enter the data in the test statistic calculator as shown below.

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Read from Analysis Node

Select Workbook: [Dropdown]

Select Analysis Node: [Dropdown]

Cumulative Events: [Text Box] 1300

Input for Survival end point

Estimate of δ : [Text Box] 0.1398

$\delta = \ln(\lambda_t / \lambda_c)$

Standard Error of Estimate of δ : [Text Box] 0.06

Output

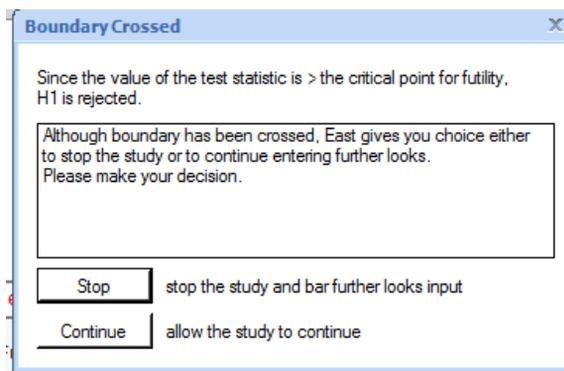
Estimate of $\delta - \delta_0$: [Text Box] 0.096

Test Statistic: [Text Box] 1.596

[Recalc] [OK] [Cancel]

Next, click **Recalc** and then **OK**. East will indicate that the H_1 (futility) boundary has been crossed and hence, the alternative hypothesis of non-inferiority is rejected in

favor of the null hypothesis of inferiority.



Click the **Stop** button to terminate the trial. You will see the IM sheet output including Final Inference details as shown below.

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of... | Est. of δ | Std. Error of Est. of δ | Futility | 97.5% RCI for HR | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|-------------------|----------------|------------|------------------|--------------------------------|----------|------------------|-------|------------------|----|------------------|
| | | | | | | | | Upper | Lower | | | |
| ▶ 1 | 0.344 | 1300 | 1.596 | 1.15 | 0.14 | 0.06 | -0.051 | NA | NA | NA | NA | NA |

Observe that the upper 97.5% Naive confidence bound for δ , 0.257, is above the non-inferiority margin of 0.044 (on the log hazard ratio scale).

Note - Click on  icon to hide or unhide the columns of your interest.

47 *Non-Inferiority Trials with Fixed Follow-Up*

This chapter will illustrate through a worked example how to design, monitor and simulate a two-sample non-inferiority trial with a time-to-event endpoint in which each subject who has not dropped out or experienced the event is followed for a fixed duration only. This implies that each subject who does not drop-out or experience the event within a given time interval, as measured from the time of randomization, will be administratively censored at the end of that interval. In East we refer to such designs as **fixed follow-up designs**.

47.1 *Type II Diabetes Trial*

A randomized non-inferiority clinical trial of a new monotherapy agent (treatment ‘t’) versus an active control (treatment ‘c’) is being planned for the treatment of type II diabetes. The primary endpoint is time to treatment failure, as measured by an elevated level of the HbA1c biomarker (greater than 8%). Each patient will be followed for up to 18 months or failure, whichever comes first. It is estimated that 50% of subjects on the active control will fail within four years. A major issue for non-inferiority trials is the selection of the non-inferiority margin for the new therapy. Since this question was discussed at length in Chapter 46, we will not repeat the discussion here. (See also, Rothman et al., 2003). Instead we will assume that, on the basis of an appropriate meta-analysis, the claim of non-inferiority can be sustained by demonstrating statistically that the treatment arm is at most 10% more hazardous than the control arm. This establishes a non-inferiority margin of $\lambda_t/\lambda_c = 1.1$ for the hazard ratio. Patient accrual will be at the rate of 1000/month for the first six months and 1500/month thereafter. The annual drop-out rate is expected to be 8% on each treatment arm.

We will design this trial to test the null hypothesis, $H_0: \lambda_t/\lambda_c \geq 1.1$, against the one sided alternative hypothesis, $H_1: \lambda_t/\lambda_c < 1.1$, with 90% power when $\lambda_t/\lambda_c = 1$. The investigators wish to select sample size that will enable the study to be completed within two years.

47.2 *Single-Look Design*

We begin by creating a single-look design for this study. To begin click **Survival: Two Samples** on the **Design** tab and then click **Parallel Design: Logrank Test Given Accrual Duration and Accrual Rates**.

This will open the input window for the design as shown below. Select **Noninferiority** from the **Design Type** dropdown. The right hand side panel of this input window is to be used for entering the relevant time-to event information. It appears with a default hazard ratio and default hazard rates for the control and treatment arms. Enter the survival information as mentioned in the design description.

- The hazard ratio under the null hypothesis (of non-inferiority) is 1.1. The hazard ratio under the alternative hypothesis at which 90% power is desired is 1.

Design Type: **Noninferiority** Number of Looks: **1**

Test Parameters | **Accrual / Dropouts**

Test Type: **1-Sided** # of Hazard Pieces: **1** Input Method: **Hazard Rates**

Type I Error (α): **0.025** Hazard Ratio (Optional)

Power: **0.9**

| | | |
|---|--------------|-------------|
| | Null | Alternative |
| <input checked="" type="radio"/> Hazard Ratio (λ_1/λ_2) | 1.1 | 1 |
| <input type="radio"/> Log Hazard Ratio $\ln(\lambda_1/\lambda_2)$ | 0.095 | 0 |

No. of Events: **Computed**

Allocation Ratio: **1**
 (n_1/n_2)

| | Hazard Rate | |
|---------|----------------|----------------|
| | Treatment:Null | Treatment:Alt. |
| Control | 0.035 | 0.038 |

Variance of Log Hazard Ratio
 Null Alternative

- Before leaving this window we must enter the hazard rate for the Active Control (Baseline) arm. We know that the four-year failure rate for the active control arm is 50%. This information can be directly entered by choosing the input method as **Cum % Survival** as shown below:

of Hazard Pieces: **1** Input Method: **Cum. % Survival**

Hazard Ratio (Optional)

Hazard Ratio (λ_1/λ_2)

| | | |
|------------|------------|-------------|
| | Null | Alternative |
| 1.1 | 1.1 | 1 |

Ratio of % Survivals at Period # 1 (S_1/S_2)

| | | |
|--------------|--------------|-------------|
| | Null | Alternative |
| 0.933 | 0.933 | 1 |

| By Time | Control | Cum. % Survival | Treatment:Alt. |
|-----------|-----------|-----------------|----------------|
| 48 | 50 | 46.652 | 50 |

Note: Period 1 hazard rates apply after time 48.

To see the conversion of this information into hazard rates, select as input method the **Hazard Rates** option. The cumulative % survival will be converted into hazard rates.

Another parameter to be decided is the **Variance** which specifies whether the null or alternative hypothesis variance will be used to convert information into sample size. Leave it at its default value. (If interested in the technical details of the choice of

47 Non-Inferiority Trials with Fixed Follow-Up

variance, refer to Appendix B, Section B.5.3.

Variance of Log Hazard Ratio

Nll Alternative

The second tab, labeled **Accrual/Dropout** is used to enter the patient accrual rate and, for fixed follow-up designs, the duration of patient follow-up and the dropout information

- In this study, each subject will be followed for up to 18 months. Therefore select the **For Fixed Period** entry from the dropdown of **Subjects are Followed** and enter 18 in the edit box.
- Enrollment begins at the rate of 1000/month and increases to 1500/month six months later. Enter this information as shown below.

Subjects are followed: **For Fixed Period**

Accrual Info

of Accrual Periods:

| Period # | Starting at Time | Accrual Rate |
|----------|------------------|--------------|
| 1 | 0.000 | 1000.000 |
| 2 | 6.000 | 1500.000 |

The second panel, labeled **Piecewise Constant Dropout Rates**, is used to enter the rate at which we expect patients to drop out of the study. Make the **# of Pieces** as 1, change the **Input Method** to **Dropout Rates** and enter the information that the annual drop-out rate is 8%, as shown below.

Piecewise Dropout Information

By Time:

Prob. of Dropout (Control)

Prob. of Dropout (Treatment)

Also, make the **Committed # of Subjects** equal to the **Min. Suggested**, 21446.

Accrual

| | | | | |
|--|--------|--------|-------|---|
| <input type="radio"/> Duration: | 16.297 | 16.297 | 24.72 |  |
| <input checked="" type="radio"/> Subjects: | 21446 | 21446 | 34080 | |

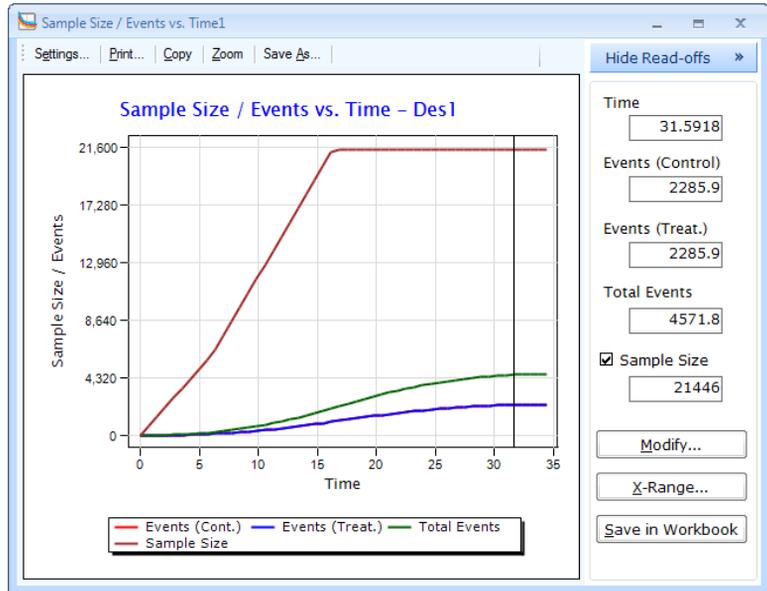
Click on **Compute** to complete the design.

47 Non-Inferiority Trials with Fixed Follow-Up

| | |
|---|-----------------------|
| | Des 1 |
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Noninferiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Null) | 1.1 |
| Hazard Ratio (Alt.) | 1 |
| Var (Log HR) | Null |
| Accrual & Dropout Parameters | |
| Accrual Rate | Multiple |
| Subjects are Followed | For Fixed Period (18) |
| No. of Accrual Periods | 2 |
| No. of Dropout Pieces | 1 |
| Sample Size | |
| Maximum | 21446 |
| Expected Under H0 | 21446 |
| Expected Under H1 | 21446 |
| Events | |
| Maximum | 4627 |
| Expected Under H0 | 4627 |
| Expected Under H1 | 4627 |
| Study Duration | |
| Maximum | 34.297 |
| Expected Under H0 | 29.268 |
| Expected Under H1 | 34.297 |
| Accrual Duration | |
| Maximum | 16.297 |
| Expected Under H0 | 16.297 |
| Expected Under H1 | 16.297 |

East reveals that any sample size between 21,446 and 34,080 will satisfy the 90% power requirement. With 21,446 patients enrolled the expected study duration is 34.3 months, consisting of 16.3 months during the enrollment phase and an additional 18 months of fixed follow-up for each patient - including the last one - to be enrolled. At the end of that 34.3 month period we expect 4627 events. This is the number of events needed to fully power the study. To see how the events arrive over time, click on the

Sample Size/Events vs. Time chart.



If you increase the sample size beyond 21,446, the total study duration will be shortened. For example consider increasing the sample to 30,000 patients by editing

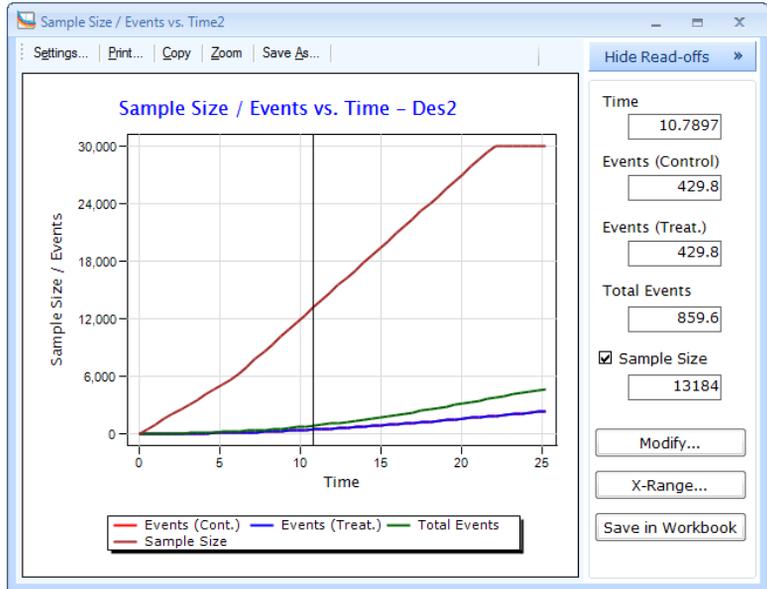
47 Non-Inferiority Trials with Fixed Follow-Up

Des1 ( icon) and creating Des2 with a new sample size.

| | Des1 | Des2 |
|---|-----------------------|-----------------------|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| Test Parameters | | |
| Design Type | Noninferiority | Noninferiority |
| No. of Looks | 1 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Hazard Ratio (Null) | 1.1 | 1.1 |
| Hazard Ratio (Alt.) | 1 | 1 |
| Var (Log HR) | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Accrual & Dropout Parameters | | |
| Accrual Rate | Multiple | Multiple |
| Subjects are Followed | For Fixed Period (18) | For Fixed Period (18) |
| No. of Accrual Periods | 2 | 2 |
| No. of Dropout Pieces | 1 | 1 |
| Sample Size | | |
| Maximum | 21446 | 30000 |
| Expected Under H0 | 21446 | 30000 |
| Expected Under H1 | 21446 | 30000 |
| Events | | |
| Maximum | 4627 | 4627 |
| Expected Under H0 | 4627 | 4627 |
| Expected Under H1 | 4627 | 4627 |
| Accrual Duration | | |
| Maximum | 16.297 | 22 |
| Expected Under H0 | 16.297 | 22 |
| Expected Under H1 | 16.297 | 22 |
| Study Duration | | |
| Maximum | 34.297 | 25.021 |
| Expected Under H0 | 29.268 | 24.28 |
| Expected Under H1 | 34.297 | 25.021 |

Now the total study duration is 25 months where the accrual phase alone lasts for 22

months.



In this case, every patient will not have been followed for the full 18 months by the time the 4627 events needed to fully power the study have arrived and the study has been closed. Only those who were enrolled early will have been followed for 18 months. The later enrollees will have been followed for a shorter time.

47 Non-Inferiority Trials with Fixed Follow-Up

47.3 Three-Look Design

Next we consider extending Des1 by permitting two equally spaced interim looks at the accruing data with a view to possible early stopping. Edit Des1, change the number of looks from 1 to 3 as shown below.

Design Type: **Noninferiority** Number of Looks: **3**

Test Parameters Boundary **Accrual / Dropouts**

Test Type: **1-Sided** # of Hazard Pieces: **1** Input Method: **Cum. % Survival**

Type I Error (α): **0.025** Hazard Ratio (Optional)

Power: **0.9**

| | | |
|---|--------------|-------------|
| | Null | Alternative |
| <input checked="" type="radio"/> Hazard Ratio (λ_1/λ_2) | 1.1 | 1 |
| <input type="radio"/> Ratio of % Survivals at Period # 1 (S_1/S_2) | 0.933 | 1 |

No. of Events: **Computed**

Allocation Ratio: **1**
(n_1/n_2)

| By Time | Cum. % Survival | |
|-----------|-----------------|-----------------|
| | Control | Treatment: Alt. |
| 48 | 50 | 50 |

Note: Period 1 hazard rates apply after time 48.

Variance of Log Hazard Ratio

Null Alternative

Change the **Committed # of Subjects** to 21701 on **accrual/Dropouts** tab. Click the **Compute** button.

| Wbk1:Des3 | |
|---|-----------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Noninferiority |
| No. of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Hazard Ratio (Null) | 1.1 |
| Hazard Ratio (Alt.) | 1 |
| Var (Log HR) | Null |
| Allocation Ratio (nt/nc) | 1 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Accrual & Dropout Parameters | |
| Accrual Rate | Multiple |
| Subjects are Followed | For Fixed Period (18) |
| No. of Accrual Periods | 2 |
| No. of Dropout Pieces | 1 |
| Sample Size | |
| Maximum | 21701 |
| Expected Under H0 | 21700.609 |
| Expected Under H1 | 21588.693 |
| Events | |
| Maximum | 4682 |
| Expected Under H0 | 4672.401 |
| Expected Under H1 | 3754.677 |
| Accrual Duration | |
| Maximum | 16.467 |
| Expected Under H0 | 16.467 |
| Expected Under H1 | 16.392 |
| Study Duration | |
| Maximum | 34.467 |
| Expected Under H0 | 29.351 |
| Expected Under H1 | 26.432 |

47 Non-Inferiority Trials with Fixed Follow-Up

Because the default Lan-DeMets-O’Brien-Fleming spending function **LD (OF)** was used in this design, the maximum study duration has been inflated very slightly, from 34.3 to 34.5 months. However, if the alternative hypothesis is true we expect to terminate the trial in 26.4 months, a savings of about 8 months. This can be seen from the table **Sample Size Information** of the design details window.

⊖ Sample Size Information

| | Control Arm | Treatment Arm | Total |
|--------------------------------|-------------|---------------|-----------|
| Sample Size (n) | | | |
| Maximum | 10850 | 10851 | 21701 |
| Expected H1 | 10794.347 | 10794.347 | 21588.693 |
| Expected H0 | 10850.304 | 10850.304 | 21700.609 |
| Events (s) | | | |
| Maximum | 2341 | 2341 | 4682 |
| Expected H1 | 2088.777 | 2088.777 | 3754.677 |
| Expected H0 | 2240.824 | 2436.398 | 4672.401 |
| Dropouts (d) | | | |
| Maximum | 1126 | 1127 | 2253 |
| Expected H1 | 903.221 | 903.66 | 1806.881 |
| Expected H0 | 1077.774 | 1064.823 | 2142.597 |
| Maximum Information (I):1170.5 | | | |

⊖ Accrual and Study Duration

| | Accrual Duration | Study Duration |
|-------------|------------------|----------------|
| Maximum | 16.467 | 34.467 |
| Expected H1 | 16.392 | 26.432 |
| Expected H0 | 16.467 | 29.351 |

47.4 Three-Look Design with Superiority Alternative

The preceding design required 21,701 subjects. This enormous up-front commitment might not be necessary if one actually believes that the new treatment is superior to the control treatment. Suppose that although the trial is still intended to reject the null hypothesis of inferiority at a non-inferiority margin of $\lambda_t/\lambda_c = 1.1$, it is believed that in fact λ_t/λ_c is less than 1; i.e., the treatment is actually superior to the active control. Ordinarily one would design a superiority trial in this situation. But now, suppose that the value λ_t/λ_c is believed to be about 0.95 under the alternative hypothesis. It would be very difficult to design a trial to prove superiority with this large a hazard ratio. (An extremely large sample size would be needed.) We can, however, use East to design a non-inferiority having 90% power at this alternative hypothesis. Edit Des3 and create Des4 by modifying the hazard ratio under the alternative hypothesis from 1 to 0.95 as shown below and **Committed Sample Size** equal to 9379.

Design Type: **Noninferiority** Number of Looks: **3**

Test Parameters Boundary Accrual / Dropouts

Test Type: **1-Sided** # of Hazard Pieces: **1** Input Method: **Hazard Rates**

Type I Error (α): **0.025**

Power: **0.9**

No. of Events: **Computed**

Allocation Ratio: **1**
 (n_1/n_2)

Hazard Ratio (Optional)

| | Null | Alternative |
|---|--------------|---------------|
| <input checked="" type="radio"/> Hazard Ratio (λ_t/λ_c) | 1.1 | 0.95 |
| <input type="radio"/> Log Hazard Ratio $\ln(\lambda_t/\lambda_c)$ | 0.095 | -0.051 |

| | Hazard Rate | |
|--------------|-----------------|-----------------|
| | Treatment: Null | Treatment: Alt. |
| 0.014 | 0.016 | 0.014 |

Variance of Log Hazard Ratio
 Null Alternative

Click the **Compute** button to complete the design.

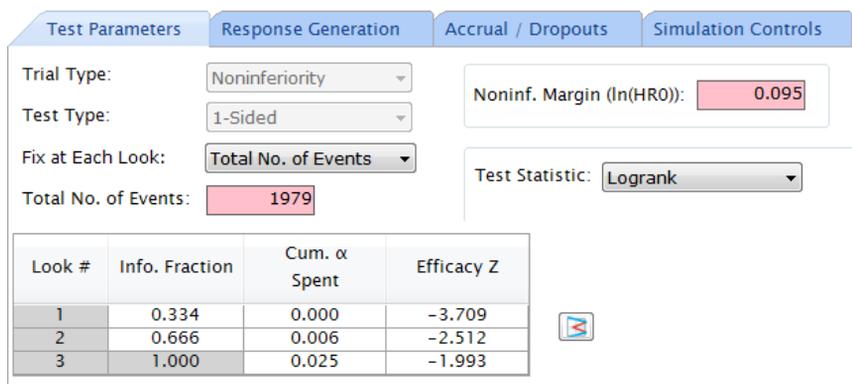
47 Non-Inferiority Trials with Fixed Follow-Up

| | Des4 |
|---|-----------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Noninferiority |
| No. of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Hazard Ratio (Null) | 1.1 |
| Hazard Ratio (Alt.) | 0.95 |
| Var (Log HR) | Null |
| Allocation Ratio (nt/nc) | 1 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Accrual & Dropout Parameters | |
| Accrual Rate | Multiple |
| Subjects are Followed | For Fixed Period (18) |
| No. of Accrual Periods | 2 |
| No. of Dropout Pieces | 1 |
| Sample Size | |
| Maximum | 9379 |
| Expected Under H0 | 9379 |
| Expected Under H1 | 9379 |
| Events | |
| Maximum | 1979 |
| Expected Under H0 | 1974.944 |
| Expected Under H1 | 1586.985 |
| Accrual Duration | |
| Maximum | 8.253 |
| Expected Under H0 | 8.253 |
| Expected Under H1 | 8.253 |
| Study Duration | |
| Maximum | 26.253 |
| Expected Under H0 | 21.956 |
| Expected Under H1 | 20.176 |

Des4 can achieve 90% power with only 9379 patients, and 1979 events. The maximum study duration is 26.3 months and the expected study duration is 20 months under the alternative hypothesis. When compared to Des3, the savings in sample size are enormous.

47.5 *Simulating a Non-Inferiority Trial*

Let us simulate Des4. Activate Des4 in the **Library** and click on  icon. You will be taken to the Simulation Input window.



| Look # | Info. Fraction | Cum. α Spent | Efficacy Z |
|--------|----------------|---------------------|------------|
| 1 | 0.334 | 0.000 | -3.709 |
| 2 | 0.666 | 0.006 | -2.512 |
| 3 | 1.000 | 0.025 | -1.993 |

To view the default simulation inputs for this design, navigate across the four tabs. The inputs are as follows:

The hazard rates displayed in the **Response Generation** tab are the ones that were specified under the alternative hypothesis; i.e., $\lambda_c = 0.0144$ and $\lambda_t = 0.0137$. Hence we expect the trial to have 90% power. To verify this click the **Simulate** button and observe that in 10000 simulated trials the null hypothesis of inferiority was rejected

47 Non-Inferiority Trials with Fixed Follow-Up

8958 times. Also note that the Average Study Duration is 19.765 months.

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

| Simulation Parameters | |
|---------------------------------|---------------------|
| Simulation ID | Sim1 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 9379 |
| Fix at Each Look | Total No. of Events |
| Noninferiority Margin (ln(HR0)) | 0.095 |
| Test Statistic | Logrank |
| Average Events | 1579.714 |
| Total Accrual Duration | 8.2527 |
| Avg. Power at Termination | 0.875 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 9379 | 338.102 | 321.899 | 162.616 | 163.277 | 9.833 | 5.001 |
| 2 | 9379 | 674.028 | 644.972 | 325.028 | 326.129 | 15.764 | 9.991 |
| 3 | 9379 | 989.234 | 973.423 | 484.427 | 484.709 | 25.307 | 14.87 |
| Average | 9379 | 808.531 | 771.184 | 389.332 | 390.603 | 19.765 | 11.966 |

Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Events | Boundaries | Stopping For | Total Simulations | |
|--------|--------|------------|--------------|-------------------|---------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 660 | -3.709 | 356 | 356 | 3.560% |
| 2 | 1319 | -2.512 | 5229 | 5229 | 52.290% |
| 3 | 1979 | -1.993 | 3168 | 4415 | 44.150% |
| Total | | | 8753 | 10000 | |
| % | | | 87.530% | | |

Next let us verify that this design also preserves the type-1 error. Edit the node Sim1 by clicking  icon. We now specify the hazard rates under the null hypothesis; i.e., $\lambda_c = 0.0144$ and $\lambda_t = 0.0144 * 1.1 = 0.0159$. We enter these hazard rates into the table labeled **Piecewise Hazards** as shown below. (Note - Consider taking the exact values of hazard rates with full precision to reproduce the results in this User Manual)

Survival Information

of Hazard Pieces: Input Method:

Hazard Ratio

| Piece | Starting at Time | Hazard Rates | | Hazard Ratio |
|-------|------------------|--------------|-----------|--------------|
| | | Control | Treatment | |
| 1 | 0.000 | 0.014 | 0.016 | 1.1 |

Generate 10000 simulated trials by clicking on the **Simulate** button.

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

| Simulation Parameters | |
|---------------------------------|---------------------|
| Simulation ID | Sim3 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 9379 |
| Fix at Each Look | Total No. of Events |
| Noninferiority Margin (ln(HR0)) | 0.095 |
| Test Statistic | Logrank |
| Average Events | 1975.497 |
| Total Accrual Duration | 8.2527 |
| Avg. Power at Termination | 0.023 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

☉ Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 9379 | 315.008 | 344.992 | 151.63 | 150.898 | 9.45 | 4.642 |
| 2 | 9379 | 630.637 | 688.363 | 303.606 | 301.186 | 14.936 | 9.278 |
| 3 | 9379 | 947.641 | 1031.354 | 456.285 | 450.992 | 22.047 | 13.919 |
| Average | 9379 | 946.199 | 1029.298 | 455.527 | 450.21 | 22.009 | 13.895 |

☉ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Events | Boundaries | Stopping For | Total Simulations | |
|--------|--------|------------|--------------|-------------------|---------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 660 | -3.709 | 1 | 1 | 0.010% |
| 2 | 1319 | -2.512 | 51 | 51 | 0.510% |
| 3 | 1979 | -1.993 | 173 | 9948 | 99.480% |
| Total | | | 225 | 10000 | |
| % | | | 2.250% | | |

We observe that only 225 of the 10000 trials rejected the null hypothesis thus confirming (up to Monte Carlo accuracy) that the type-1 error of 0.025 is preserved.

48 Superiority Trials Given Accrual Duration and Study Duration

This chapter will illustrate through a worked example how to design and simulate a two-sample superiority trial with a time-to-event trial endpoint, where the accrual duration and study duration are constrained. Most trials in the pharmaceutical industry setting are designed in this manner, time being a more rigid constraint than the accrual rate of patients. The duration of a clinical trial impacts the duration of a drug development program, and thus time to market and potential revenues. Therefore it is of interest to fix the study duration as well as the accrual duration to finish the clinical trial according to schedule. The option to design a trial in this way is available in East.

48.1 Calculating a Sample Size

For this design, East obtains the maximum number of events D_{max} from the maximum information I_{max} , as described in Appendix sections B.5 and B.5.3. To calculate the sample size, we first equate the expected number of events $d(S_a + S_f)$ (as calculated in Appendix D which depends on the accrual duration (S_a) and the duration of follow-up (S_f) to the maximum number of events D_{max} .

$$d(S_a + S_f) = D_{max} \quad (48.1)$$

In this type of design the accrual duration S_a and the study duration $S_a + S_f$ are given as input. East iterates between sample sizes, increasing onwards from a minimum value of D_{max} , enrolled over a duration of S_a until D_{max} events are found to occur within a study duration of $S_a + S_f$. The result is the unique sample size required to obtain the proper power for the study.

48.2 The RALES Clinical Trial: Initial Design

The RALES trial (Pitt et. al., 1999) was a double blind study of aldosterone-receptor blocker spironolactone at a daily dose of 25 mg in combination with standard doses of an ACE inhibitor (treatment arm) versus standard therapy of an ACE inhibitor (control arm) in patients who had severe heart failure as a result of systolic left ventricular dysfunction. The primary endpoint was death from any cause. Six equally-spaced looks at the data using the Lan-DeMets-O'Brien-Fleming spending function were planned. The trial was designed to detect a hazard ratio of 0.83 with 90% power at a two-sided 0.05 level of significance. The hazard rate of the control arm was estimated to be 0.38.

Randomization was scheduled to begin in March 1995 and complete in December 1996 for a total of 1.8 years of enrollment. Follow-up was planned through December

1999, so that the total study duration from first patient enrolled to last patient visit should be 4.8 years.

We begin by using East to design RALES under these basic assumptions. To begin click **Survival: Two Samples** on the **Design** tab and then click **Parallel Design: Logrank Test Given Accrual Duration and Study Duration**.

A new screen will appear. Enter the appropriate design parameters into the dialog box as shown below.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

Design Type: Superiority Number of Looks: 6

Test Parameters Boundary Accrual / Dropouts

Test Type: 2-Sided # of Hazard Pieces: 1 Input Method: Hazard Rates

Type I Error (α): 0.05 Hazard Ratio (Optional) Alternative

Power: 0.9 Hazard Ratio (λ₁/λ₂) 0.83

Sample Size (n): Computed Log Hazard Ratio ln(λ₁/λ₂) -0.186

No. of Events: Computed

| Hazard Rate | |
|-------------|-----------------|
| Control | Treatment: Alt. |
| 0.38 | 0.3154 |

Allocation Ratio: 1 (n₁/n₂)

Variance of Log Hazard Ratio

Null Alternative

The box labeled **Variance of Log Hazard Ratio** specifies whether the calculation of the required number of events is to be based on the variance estimate of the log hazard ratio under the null hypothesis or the alternative hypothesis. The default choice in East is **Null**. Most textbooks recommend this choice as well (see, for example Collett, 1994, equation (2.21) specialized to no ties). It will usually not be necessary to change this default. For a technical discussion of this issue refer to Appendix B, Section B.5.3.

Next, click on the **Boundary Info** tab. We will take six equally spaced looks at the data using the Lan-DeMets O’Brien-Fleming spending function. These are the default

48 Superiority Trials Given Accrual Duration and Study Duration

setting in East.

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.167 | 0.000 | 5.367 | -5.367 |
| 2 | 0.333 | 0.000 | 3.710 | -3.710 |
| 3 | 0.500 | 0.003 | 2.970 | -2.970 |
| 4 | 0.667 | 0.012 | 2.539 | -2.539 |
| 5 | 0.833 | 0.028 | 2.252 | -2.252 |
| 6 | 1.000 | 0.050 | 2.045 | -2.045 |

Note that we do not select a futility boundary in this case. Next click on the **Accrual/Dropout Info** tab. Here we will specify the accrual information and dropout rates. The software allows a specification of piecewise constant hazards and variable accrual rates but we start by looking at an example that does not require any of these options. In the drop-down menu next to **Subjects are followed:** select **Until End of Study**. Set the **Accrual Duration** to 1.8 years and the **Study Duration** to 4.8 years. Notice that East has changed the settings so that at 1.8 years the study should be 100% accrued. Keep the number of accrual periods equal to the default of 1. To the right of the **Accrual Info** box is the **Piecewise Constant Dropout Rates** box. This box is used to enter that rate at which we expect patients to drop out of the study. For the present we will assume that there are no drop-outs.

| Period # | By Time | Cum. % Accrued |
|----------|---------|----------------|
| 1 | 1.800 | 100.000 |

Click on **Compute** to complete the design. The design is shown as a row in the

Output Preview located in the lower pane of this window. You can select this design by clicking anywhere along the row in the **Output Preview**. With Des1 selected, click the  icon to display the details of this design in the upper pane, which are shown below. You may also wish to save this design. Select Des1 in the **Output Preview** window and click the  to save this design to Workbook1 in the **Library**.

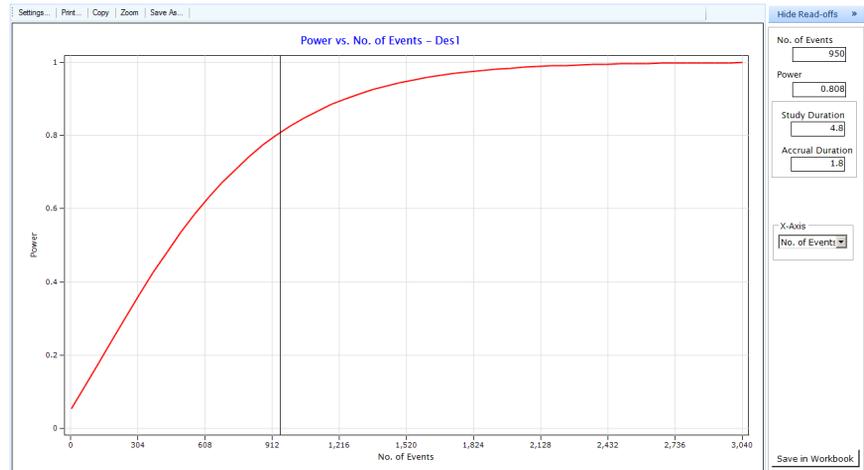
| Des 1 | |
|---|--------------------|
| Mnemonic | SU-2S-LRSD |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 6 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.83 |
| Var (Log HR) | Null |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Accrual & Dropout Parameters | |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 1689 |
| Expected Under H0 | 1688.978 |
| Expected Under H1 | 1687.556 |
| Events | |
| Maximum | 1243 |
| Expected Under H0 | 1233.984 |
| Expected Under H1 | 903.595 |
| Study Duration | |
| Maximum | 4.8 |
| Expected Under H0 | 4.416 |
| Expected Under H1 | 3.304 |
| Accrual Duration | |
| Maximum | 1.8 |
| Expected Under H0 | 1.8 |
| Expected Under H1 | 1.798 |

East notifies you that 1243 events and a sample size of 1689 are required to attain the desired 90% power in the allotted time.

East provides charts to examine the trade-offs between power and accrual duration, study duration, sample size or number of events. Select Des1 in the **Library** click the  icon and select **Power vs. Sample Size**.

48 Superiority Trials Given Accrual Duration and Study Duration

Switch the **X-Axis** to **No. of Events**. The power of the study is really tied to the number of events that are observed. This chart shows the direct relationship between power and number of events.



Note that 950 events give us about 81% power. You may wish to save this chart to the **Library** by clicking on the **Save in Workbook** button.

48.3 Incorporating Drop-Outs

The investigators expect 5% of the patients in the spironolactone group and the control group to drop out each year. Create a new design by selecting Des1 in the **Library**, and clicking the  icon on the **Library** toolbar. Next, click on the **Accrual/Dropout Info** tab. In the **Piecewise Constant Dropout Rates** box, select 1 for the number of pieces and change the **Input Method** from **Hazard Rates** to **Prob. of Dropout**. Then enter 0.05 dropouts by 1 year for the treatment and

control arm as shown below.

Piecewise Dropout Information

of Pieces: Input Method:

By Time:

Prob. of Dropout (Control)

Prob. of Dropout (Treatment)

Note: Period 1 hazard rates apply after time 1.

Click the **Compute** button to generate output for Des2. With Des2 selected in the **Output Preview**, click the  icon to save Des2 to the **Library**. In the **Library**, select the rows for Des1 and Des2, by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the two designs side-by-side.

A comparison of the two plans reveals that, because of the drop-outs, we require 1,824 subjects to be enrolled under Des2 rather than 1689 under Des1. Also, the expected study duration will not change much under the alternative and null hypotheses between Des1 and Des2.

48.4 Incorporating Non-Constant Accrual Rates

In many clinical trials the enrollment rate is low in the beginning and reaches its maximum expected level a few months later when all the sites enrolling patients are onboard. Suppose that 20% of the total accrual is expected to occur during the first six months with the rest happening during the remaining 1.3 years. Create a new design by selecting Des2 in the **Library**, and clicking the  icon on the **Library** toolbar. Next, click on the **Accrual/Dropout Info** tab. Specify that there are two accrual periods and enter the cumulative accrual for each period in the dialog box as shown

48 Superiority Trials Given Accrual Duration and Study Duration

below.

Accrual Info

Accrual Duration: Study Duration:

of Accrual Periods:

| Period # | By Time | Cum. % Accrued |
|----------|---------|----------------|
| 1 | 0.5 | 20.000 |
| 2 | 1.8 | 100.000 |

Click the **Compute** button to generate output for Des3. With Des3 selected in the **Output Preview**, click the  icon to save Des3 to the **Library**. In the **Library**, select the rows for Des1, Des2, and Des3 by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the three designs side-by-side.

Notice that we now need 1837 subjects to be enrolled to compensate for the overall later enrollment of subjects.

48.5 Simulation

48.5.1 Simulating Under H_1

48.5.2 Simulating Under H_0

It would be useful to verify the operating characteristics of the various plans created in the previous section by simulation. Select Des3 in the **Library** and click the  icon. You will be taken to the following simulation worksheet.

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

Number of Looks:

Test Parameters | Response Generation | **Accrual / Dropouts** | Simulation Controls

Trial Type: Test Statistic:

Test Type:

Fix at Each Look:

Total No. of Events:

| Look # | Info. Fraction | Cum. α Spent | | Efficacy Z | |
|--------|----------------|--------------|-------|------------|--------|
| | | Upper | Lower | Upper | Lower |
| 1 | 0.167 | 0.000 | 0.000 | 5.369 | -5.369 |
| 2 | 0.333 | 0.000 | 0.000 | 3.712 | -3.712 |
| 3 | 0.500 | 0.002 | 0.002 | 2.968 | -2.968 |
| 4 | 0.667 | 0.006 | 0.006 | 2.538 | -2.538 |
| 5 | 0.833 | 0.014 | 0.014 | 2.252 | -2.252 |

48.5.1 Simulating Under H_1

We will first simulate the trial under the alternative hypothesis H_1 . In the **Simulation Parameters** tab select **Total No. of Events** to fix at each look - the default option. Select **LogRank** from the drop-down menu next to **Test Statistic**. Other options for a test statistic include the Wilcoxon-Gehan and Harrington-Fleming. Next, click the **Simulate** button to simulate 10000 trials. A new row labeled Sim1 will

appear in the **Output Preview** window. Select that row and click the  icon to save it to the **Library**. In the **Library**, double-click Sim1. A portion of the output is displayed below. (The actual values may differ, depending on the starting seed used).

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

| Test Parameters | |
|---|---------------------|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 6 |
| Test Type | 2-Sided |
| Sample Size (n) | 1837 |
| Fix at Each Look | Total No. of Events |
| Test Statistic | Logrank |
| Average Events | 907.502 |
| Total Accrual Duration | 1.8 |
| Avg. Power at Termination | 0.893 |
| Response Generation Parameters | |
| HR = λ_1/λ_2 | 0.83 |
| Hazard Rate - Control (λ_0) | 0.38 |
| Hazard Rate - Treatment (λ_1) | 0.315 |
| Accrual / Dropouts Parameters | |
| Prob. of Dropout by Time = | 1 |
| Control | 0.05 |
| Treatment | 0.05 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

☉ Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 1221.648 | 111.89 | 95.12 | 15.122 | 15.482 | 1.256 | 0.489 |
| 2 | 1796.862 | 222.83 | 191.171 | 30.094 | 31.108 | 1.769 | 0.663 |
| 3 | 1837 | 332.985 | 289.015 | 45.087 | 46.9 | 2.254 | 0.976 |
| 4 | 1837 | 438.425 | 390.575 | 59.929 | 62.618 | 2.851 | 1.301 |
| 5 | 1837 | 540.806 | 495.194 | 74.683 | 78.427 | 3.637 | 1.626 |
| 6 | 1837 | 641.165 | 601.835 | 89.587 | 94.2 | 4.787 | 1.951 |
| Average | 1835.769 | 482.874 | 424.628 | 65.174 | 69.045 | 3.29 | 1.425 |

☉ Simulation Boundaries and Cumulative Boundary Crossing Probabilities

| Look # | Events | Boundaries Efficacy | | Stopping For | | Total Simulations | |
|---------|--------|---------------------|--------|----------------|----------------|-------------------|----------|
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| | | | | | | | |
| 2 | 414 | 3.712 | -3.712 | 0 | 337 | 337 | 3.370% |
| 3 | 622 | 2.968 | -2.968 | 0 | 2546 | 2546 | 25.460% |
| 4 | 829 | 2.538 | -2.538 | 0 | 5566 | 5566 | 55.660% |
| 5 | 1036 | 2.252 | -2.252 | 0 | 7757 | 7757 | 77.570% |
| 6 | 1243 | 2.045 | -2.045 | 0 | 8929 | 10000 | 100.000% |
| Total % | | | | 0.000% | 89.290% | | |

We will now run another 10000 simulations, this time fixing the calendar time of each look instead of fixing the number of events. Click the  icon on the left bottom corner to go back to the input window of Sim1. In the **Test Parameters** tab select **Look Time** from the drop-down menu next to **Fix at Each Look**:

Test Parameters
Response Generation
Accrual / Drop

Trial Type: Superiority

Test Type: 2-Sided

Fix at Each Look: Look Time

Study Duration: 4.797

Test Statistic

| Look # | Analysis Time | Efficacy Z | |
|--------|---------------|------------|--------|
| | | Upper | Lower |
| 1 | 1.256 | 5.369 | -5.369 |
| 2 | 1.769 | 3.712 | -3.712 |
| 3 | 2.254 | 2.968 | -2.968 |
| 4 | 2.852 | 2.538 | -2.538 |
| 5 | 3.640 | 2.252 | -2.252 |

48 Superiority Trials Given Accrual Duration and Study Duration

When the **Look Time** option is selected the locations of the interim looks at which stopping boundaries are computed are expressed in terms of the calendar time of each interim look instead of the number of events at each interim look.

Next, click the **Simulate** button to simulate 10000 trials. A new row labeled Sim2 will appear in the **Output Preview** window. Select that row and click the  icon to save it to the **Library**. In the **Library**, double-click Sim2. A portion of the output is displayed below. (The actual values may differ, depending on the starting seed used).

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

| Test Parameters | |
|---|-------------|
| Simulation ID | Sim2 |
| Design Type | Superiority |
| Number of Looks | 6 |
| Test Type | 2-Sided |
| Sample Size (n) | 1837 |
| Fix at Each Look | Look Time |
| Test Statistic | Logrank |
| Average Events | 900.309 |
| Total Accrual Duration | 1.8 |
| Avg. Power at Termination | 0.903 |
| Response Generation Parameters | |
| HR = λ_1/λ_0 | 0.83 |
| Hazard Rate - Control (λ_0) | 0.38 |
| Hazard Rate - Treatment (λ_1) | 0.315 |
| Accrual / Dropouts Parameters | |
| Prob. of Dropout by Time = | 1 |
| Control | 0.05 |
| Treatment | 0.05 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

⊖ Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 1221.045 | 111.825 | 94.843 | 15.136 | 15.357 | 1.256 | 0.488 |
| 2 | 1802.009 | 222.977 | 190.757 | 30.182 | 30.949 | 1.769 | 0.662 |
| 3 | 1837 | 333.126 | 288.822 | 45.116 | 46.722 | 2.254 | 0.976 |
| 4 | 1837 | 438.386 | 390.465 | 59.986 | 62.473 | 2.852 | 1.301 |
| 5 | 1837 | 540.595 | 494.764 | 74.945 | 78.405 | 3.64 | 1.628 |
| 6 | 1837 | 641.08 | 601.653 | 89.855 | 94.278 | 4.797 | 1.955 |
| Average | 1835.754 | 479.352 | 420.958 | 64.76 | 68.317 | 3.262 | 1.414 |

⊖ Simulation Boundaries and Cumulative Boundary Crossing Probabilities

| Look # | Look Time | Boundaries Efficacy | | Stopping For | | Total Simulations | |
|---------|-----------|---------------------|--------|----------------|----------------|-------------------|----------|
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 1.256 | 5.369 | -5.369 | 0 | 1 | 1 | 0.010% |
| 2 | 1.769 | 3.712 | -3.712 | 0 | 340 | 340 | 3.400% |
| 3 | 2.254 | 2.968 | -2.968 | 0 | 2648 | 2648 | 26.480% |
| 4 | 2.852 | 2.538 | -2.538 | 0 | 5696 | 5696 | 56.960% |
| 5 | 3.64 | 2.252 | -2.252 | 0 | 7874 | 7874 | 78.740% |
| 6 | 4.797 | 2.045 | -2.045 | 0 | 9027 | 10000 | 100.000% |
| Total % | | | | 0.000% | 90.270% | | |

48.5.2 Simulating Under H0

To simulate under the null hypothesis we must go to the **Response Generation Info** tab in the simulation worksheet. In this tab change the hazard rate for the treatment arm to 0.38.

Test Parameters
Response Generation
Accrual / Dropouts
Sir

Survival Information

of Hazard Pieces
Input Method:

| Hazard Rate | | Hazard Ratio |
|-------------|-----------|--------------------------------|
| Control | Treatment | |
| 0.38 | 0.38 | <input type="text" value="1"/> |

This change implies that we will be simulating under the null hypothesis. Next, click

on the **Test Parameters** tab and make sure that the **Total No. of Events** is fixed at each look. Next, click the **Simulate** button to simulate 10000 trials. A portion of the results are displayed below.

☰ **Simulation Boundaries and Cumulative Boundary Crossing Probabilities**

| Look # | Events | Boundaries | | Stopping For | | Total Simulations | |
|---------|--------|------------|--------|----------------|----------------|-------------------|----------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 207 | 5.369 | -5.369 | 0 | 0 | 0 | 0.000% |
| 2 | 414 | 3.712 | -3.712 | 0 | 2 | 2 | 0.020% |
| 3 | 622 | 2.968 | -2.968 | 9 | 22 | 31 | 0.310% |
| 4 | 829 | 2.538 | -2.538 | 55 | 76 | 131 | 1.310% |
| 5 | 1036 | 2.252 | -2.252 | 128 | 151 | 279 | 2.790% |
| 6 | 1243 | 2.045 | -2.045 | 245 | 258 | 10000 | 100.000% |
| Total % | | | | 2.450% | 2.580% | | |

Out of 10000 simulated trials 245 crossed the upper stopping boundary and 258 crossed the lower stopping boundary thus confirming (up to Monte Carlo accuracy) that the type-1 error is preserved for this design.

48.6 User Defined R Function

East allows you to customize simulations by inserting user-defined R functions for one or more of the following tasks: generate response, compute test statistic, randomize subjects, generate arrival times, and generate dropout information. The R functionality for arrivals and dropouts will be available only if you have entered such information at the design stage. Although the R functions are also available for all normal and binomial endpoints, we will illustrate this functionality for a time-to-event endpoint. Specifically, we will use an R function to generate Weibull survival responses.

Start East afresh. On the **Design** tab, click **Survival: Two Samples** and then **Logrank Test Given Accrual Duration and Study Duration**.

Choose the design parameters as shown below. In particular, select a one sided test

48 Superiority Trials Given Accrual Duration and Study Duration

with type-1 error of $\alpha = 0.025$.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

Design Type: Superiority Number of Looks: 1

Test Parameters Accrual / Dropouts

Test Type: 1-Sided

Type I Error (α): 0.025

Power: 0.9

Sample Size (n): Computed

No. of Events: Computed

Allocation Ratio: 1 (n_1/n_2)

of Hazard Pieces: 1

Input Method: Hazard Rates

Hazard Ratio (Optional)

Hazard Ratio (λ_1/λ_2) Alternative: 0.5

Log Hazard Ratio $\ln(\lambda_1/\lambda_2)$ Alternative: -0.693

| Hazard Rate | |
|-------------|-----------------|
| Control | Treatment: Alt. |
| 0.035 | 0.017 |

Variance of Log Hazard Ratio

Null Alternative

Click **Compute** and save this design (Des1) to the **Library**. Right-click Des1 in the **Library** and click **Simulate**. In the **Simulation Control Info** tab, check the box for **Suppress All Intermediate Output**. Type 10000 for **Number of Simulations** and select **Clock** for **Random Number Seed**.

Test Parameters Response Generation Accrual / Dropouts Simulation Controls

Number of Simulations: 10000

Refresh Frequency: 1000

Random Number Seed

Clock

Fixed 100

Suppress All Intermediate Output

Output Options

Output Type: Case Data

Save summary statistics for every simulation run

Save subject level data for 1 simulation runs

Note: Max. 100,000 records will be saved.

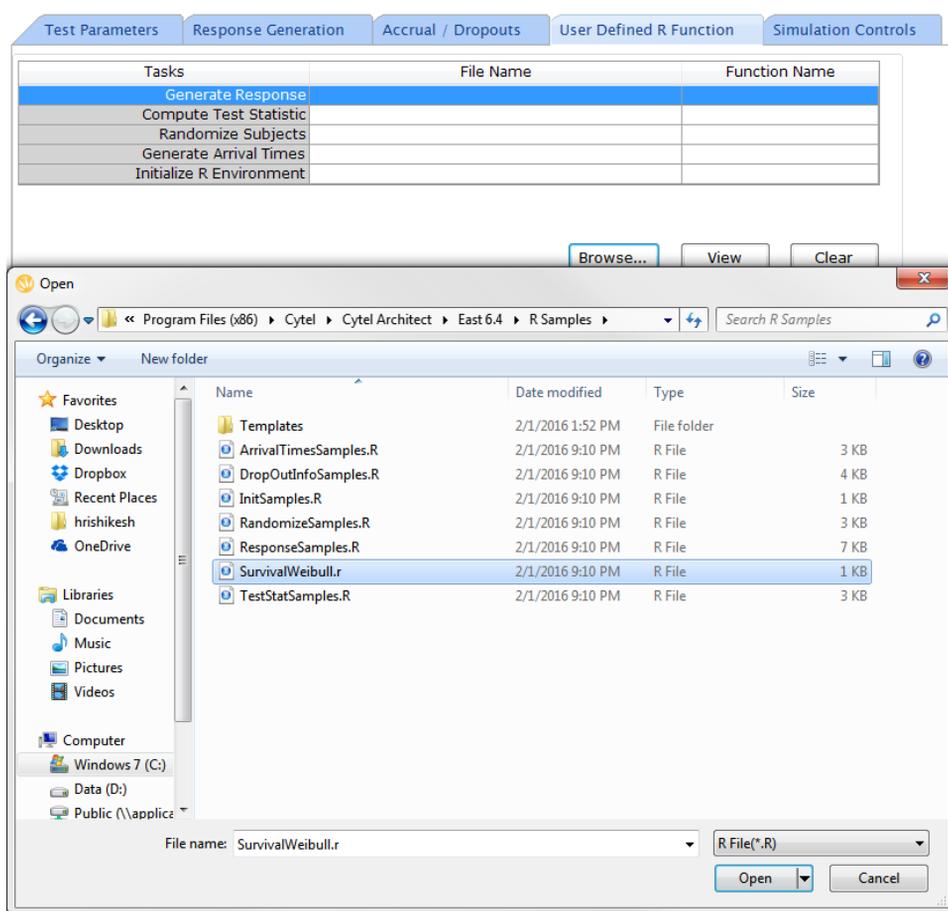
In the top right-hand corner for the input window, click **Include Options**, and then click **User Defined R Function**.

Include Options

- Site Info
- Randomization Info
- User Defined R Function
- Stratification Info

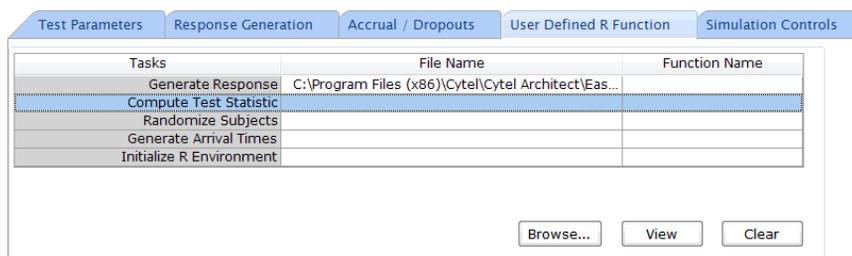
For now, leave the row **Initialize R Environment** blank. This optional task can be useful for loading required libraries, setting seeds for simulations, and initializing global variables.

Select the row for **Generate Response**, click **Browse...**, and navigate to the folder containing your R file.



48 Superiority Trials Given Accrual Duration and Study Duration

Select the file and click **Open**. The path should now be displayed under **File Name**.

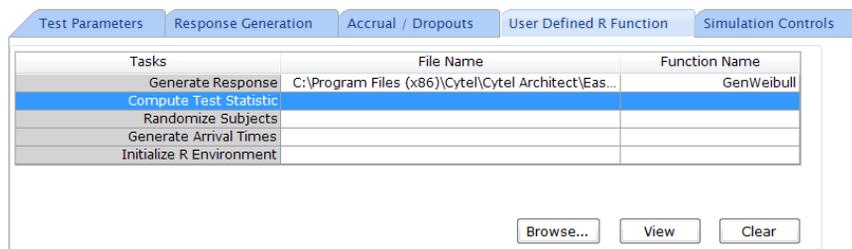


Click on the first row and then click **View** to open a notepad application to view your R file. In this example, I am generating survival responses for both control and treatment arms from a Weibull with shape parameter = 2 (i.e. exponential), with the same hazard rate in both arms.

```

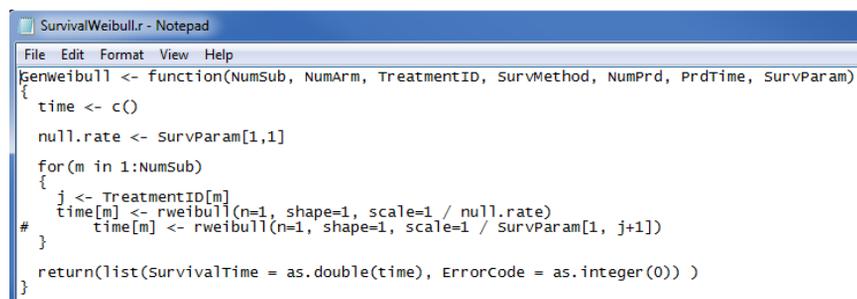
SurvivalWeibull.r - Notepad
File Edit Format View Help
GenWeibull <- function(NumSub, NumArm, TreatmentID, SurvMethod, NumPrd, PrdTime, SurvParam)
{
  time <- c()
  null.rate <- SurvParam[1,1]
  for(m in 1:NumSub)
  {
    j <- TreatmentID[m]
    time[m] <- rweibull(n=1, shape=2, scale=1 / null.rate)
  }
  return(list(SurvivalTime = as.double(time), ErrorCode = as.integer(0)) )
}
    
```

Copy the function name (in this case *GenWeibull*).



Close the R file and paste the function name in the cell for **Function Name**. Click **Simulate**.

Return to the tab for **User Defined R Function**, select the **Generate Response** row, and click **View**. In the R function, change the shape parameter = 1, to generate responses from a Weibull distribution with decreasing hazards. Save and close the R file. You may not be able to save the file in the C: drive due to administrative privileges. So save the updated file somewhere else, say the Desktop.



```
SurvivalWeibull.r - Notepad
File Edit Format View Help
GenWeibull <- function(NumSub, NumArm, TreatmentID, SurvMethod, NumPrd, PrdTime, SurvParam)
{
  time <- c()
  null.rate <- SurvParam[1,1]
  for(m in 1:NumSub)
  {
    j <- TreatmentID[m]
    time[m] <- rweibull(n=1, shape=1, scale=1 / null.rate)
    # time[m] <- rweibull(n=1, shape=1, scale=1 / SurvParam[1, j+1])
  }
  return(list(SurvivalTime = as.double(time), ErrorCode = as.integer(0)) )
}
```

Browse to the new file on the Desktop. The function name is same so no need to change that. Click **Simulate**.

Select both simulations (Sim1 and Sim2) from the **Output Preview**, and on the

48 Superiority Trials Given Accrual Duration and Study Duration

toolbar, click  to display in the **Output Summary**.

| | Sim1 | Sim2 |
|-------------------------------------|--------------------|--------------------|
| Mnemonic | SU-2S-LRSD | SU-2S-LRSD |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| Test Type | 1-Sided | 1-Sided |
| Power | 0.027 | 0.026 |
| No. of Looks | 1 | 1 |
| Test Statistic | Logrank | Logrank |
| Model Parameters | | |
| No. of Hazard Pieces | 1 | 1 |
| Hazard Ratio | 0.5 | 0.5 |
| Sample Size | | |
| Maximum | 182 | 182 |
| Events | | |
| Maximum | 88 | 88 |
| Simulation Results (Overall) | | |
| Average Study Duration | 34.624 | 30.708 |
| Average Sample Size | 182 | 182 |
| Other Parameters | | |
| Subjects Followed-up | Until End of Study | Until End of Study |

Notice that the type-1 error appears to be controlled in both cases. When we simulated from the exponential (Sim2), the average study duration (30.7 months) was close to what was calculated at Des1 for the expected study duration under the null. However, when we simulated from the Weibull with increasing hazards (Sim1), the average study duration increased to 34.6 months.

Appendix O contains detailed specifications for the required inputs and outputs of R functions for each task and endpoint. The ability to use custom R functions for many simulation tasks allows considerable flexibility in performing sensitivity analyses and assessment of key operating characteristics.

48.7 Assurance for Survival

Assurance, or probability of success, is a Bayesian version of power, which corresponds to the (unconditional) probability that the trial will yield a statistically significant result. Specifically, it is the prior expectation of the power, averaged over a prior distribution for the unknown treatment effect (see O’Hagan et al., 2005). For a given design, East allows you to specify a prior distribution, for which the assurance or probability of success will be computed. In this section, we will replicate and extend

an example from Sabin et al. (2014).

Start East afresh. Click **Survival: Two Samples** on the **Design** tab and then click **Parallel Design: Logrank Test Given Accrual Duration and Study Duration**. Compute the following design: Design Type = Superiority, Test Type = 1-sided, Type-1 error = 0.025, Power = 80%, Hazard ratio = 0.75.

Design Type: Superiority Number of Looks: 1

Test Parameters Accrual / Dropouts

Test Type: 1-Sided # of Hazard Pieces: 1 Input Method: Hazard Rates

Type I Error (α): 0.025 Hazard Ratio (Optional) Alternative 0.75

Power: 0.8 Hazard Ratio (λ_t / λ_c) -0.288

Log Hazard Ratio $\ln(\lambda_t / \lambda_c)$

| Hazard Rate | |
|-------------|-----------------|
| Control | Treatment: Alt. |
| 0.035 | 0.026 |

Sample Size (n): 694

No. of Events: 380

Allocation Ratio: 1 (n_t / n_c)

Variance of Log Hazard Ratio

Null Alternative

Assurance (Probability of Success)

This design requires 380 events to achieve 80% power. However, this value of power depends on the assumption that the HR is precisely 0.75, or that $\delta = \ln(\text{HR}) = -0.288$. Sabin et al. (2014) explored various prior distributions derived from Phase 2 data. In one example, they used a Normal prior distribution for $\ln(\text{HR})$, with a mean of -0.183 , and a standard deviation of 0.135.

Select the **Assurance** checkbox in the Input window. In the **Distribution** list, click **Normal**, and in the **Input Method** list, click **E(δ) and SD(δ)**.

48 Superiority Trials Given Accrual Duration and Study Duration

Assurance (Probability of Success) 0.458

Prior Distribution for: Log Hazard Ratio (δ) Distribution: Normal

Input Method: E(δ) and SD(δ)

User Specified

E(δ): -0.183

SD(δ): 0.135

This replicates their reported *Probability of Ph 3 success* of 0.46.

East also allows you to specify an arbitrary discrete prior distribution through an R function. In the **Distribution** list, click **User Specified-R**, and then click **Browse...** to select the R file where you have constructed a prior. Click **View...** to open the R file.

```

InHR - Notepad
File Edit Format View Help
}nHR = function() {
  Error <- 0

  mean <- (-0.183)
  sd <- 0.135

  min <- mean - 4*sd
  max <- mean + 4*sd
  points.total <- 31
  step <- (max - min)/(points.total - 1)
  points <- seq(from=min, to=max, length=points.total)
  mid.points <- (points + (step/2))[-points.total]
  out.points <- c(-Inf, mid.points, Inf)
  L <- 1 : points.total
  R <- 2 : (points.total+1)

  weights <- pnorm(out.points[R], mean, sd) - pnorm(out.points[L], mean, sd)

  points <- c(0, points)
  weights <- c(1, weights)

  return(list(Delta = as.double(points), Prob = as.double(weights), ErrorCode = as.integer(Error)))
}
  
```

In this R file, we have constructed a discretized Normal distribution with the same mean and standard deviation as above, but added a lump of equal weight at the null hypothesis. Type the function name (in this case, InHR) into the **R Function** field, and click **Compute**. The resulting probability of success (0.241) is even lower due to the prior weight on the null hypothesis.

Assurance (Probability of Success)
Prior Distribution for: Distribution:

49 *Non Inferiority Trials Given Accrual Duration and Study Duration*

This chapter will illustrate through a worked example how to design and simulate a two-sample non inferiority trial with a time to event trial endpoint, when the accrual duration and study duration are fixed.

49.1 *Calculating a Sample Size*

For this design, East obtains the maximum number of events D_{max} from the maximum information I_{max} , as described in Appendix sections B.5 and B.5.3. To calculate the sample size, we first equate the expected number of events $d(S_a + S_f)$ (as calculated in Appendix D which depends on the accrual duration (S_a) and the duration of follow-up (S_f) to the maximum number of events D_{max} .

$$d(S_a + S_f) = D_{max} \tag{49.1}$$

In this type of design the accrual duration S_a and the study duration $S_a + S_f$ are given as input. East iterates between sample sizes, increasing onwards from a minimum value of D_{max} , enrolled over a duration of S_a until D_{max} events are found to occur within a study duration of $S_a + S_f$. The result is the unique sample size required to obtain the proper power for the study.

49.2 *The Non Inferiority Margin*

The first step in designing a non-inferiority trial is to establish a suitable non inferiority margin. This is typically done by performing a meta-analysis on past clinical trials of the active control versus placebo. Regulatory agencies then require the sponsor of the clinical trial to demonstrate that a fixed percentage of the active control effect (usually 50%) is retained by the new treatment. A further complication arises because the active control effect can only be estimated with error. We illustrate below with an example provided by reviewers at the FDA.

Rothman et al. (2003) have discussed a clinical trial to establish the non inferiority of the test drug Xeloda (treatment t) relative to the active control (treatment c) consisting of 5 fluorouracil with leucovorin (5FU+LV) for metastatic colorectal cancer. In order to establish a suitable non inferiority margin for this trial it is necessary to first establish the effect of 5FU+LV relative to the reference therapy of 5FU alone (treatment p , here regarded as placebo). To establish this effect the FDA conducted a ten study random effects meta analysis (FDA Medical Statistical review for Xeloda, NDA 20 896, April 2001) of randomized comparisons of 5-FU alone versus 5-FU+LV.

Letting λ_t, λ_c and λ_p denote the constant hazard rates for the new treatment, the active control and the placebo, respectively, the FDA meta analysis established that

$$\ln(\widehat{\lambda_p/\lambda_c}) = 0.234$$

with standard error

$$\text{se}[\ln(\widehat{\lambda_p/\lambda_c})] = 0.075 .$$

Thus with $100\gamma\%$ confidence the active control effect lies inside the interval

$$[0.234 - 0.075\Phi^{-1}(\frac{1+\gamma}{2}), 0.234 + 0.075\Phi^{-1}(\frac{1+\gamma}{2})] \quad (49.2)$$

The new study is required to demonstrate that some fraction (usually 50%) of the active control effect is retained. Rothman et al. (2003) state that the claim of non inferiority for the new treatment relative to the active control can be demonstrated if the upper limit of a two sided $100(1 - \alpha)\%$ confidence interval for $\ln(\lambda_t/\lambda_c)$ is less than a pre specified fraction of the lower limit of a two sided $100\gamma\%$ confidence interval for the active control effect established by the meta-analysis. This is known as the “two confidence intervals procedure”. Specifically in order to claim non inferiority in the current trial it is necessary to show that

$$\ln(\widehat{\lambda_t/\lambda_c}) + \Phi^{-1}(1 - \alpha/2)\text{se}[\ln(\widehat{\lambda_t/\lambda_c})] < (1 - f_0)\{\ln(\widehat{\lambda_p/\lambda_c}) - \Phi^{-1}(\frac{1+\gamma}{2})\text{se}[\ln(\widehat{\lambda_p/\lambda_c})]\} . \quad (49.3)$$

We may re-write the non inferiority condition (49.3) in terms of a one-sided Wald test of the form

$$\frac{\ln(\widehat{\lambda_t/\lambda_c}) - \delta_0}{\text{se}[\ln(\widehat{\lambda_t/\lambda_c})]} < \Phi^{-1}(1 - \alpha/2) , \quad (49.4)$$

where

$$\delta_0 = (1 - f_0)\{\ln(\widehat{\lambda_p/\lambda_c}) - \Phi^{-1}(\frac{1+\gamma}{2})\text{se}[\ln(\widehat{\lambda_p/\lambda_c})]\} \quad (49.5)$$

is the non inferiority margin.

The choice $f_0 = 1$ implies that the entire active control effect must be retained in the new trial and amounts to running a superiority trial. At the other end of the spectrum, the choice $f_0 = 0$ implies that none of the active control effect need be retained; i.e., the new treatment is only required to demonstrate effectiveness relative to placebo. The usual choice is $f_0 = 0.5$, implying that the new treatment is required to retain at least 50% of the active control effect. The usual choice for α is $\alpha = 0.05$. A conservative choice for the coefficient γ is $\gamma = (1 - \alpha) = 0.95$. Rothman et al. (2003) refer to this method of establishing the non inferiority margin as the “two 95 percent

49 Non Inferiority Trials Given Accrual Duration and Study Duration

two sided confidence interval procedure” or the “95-95 rule”. In general this approach leads to rather tight margins unless the active control effect is substantial. Rothman et al. (2003) have also proposed more lenient margins that vary with the amount of power desired. Fleming (2007), however, argues for the stricter 95-95 rule on the grounds that it offers greater protection against an ineffective medical compound being approved in the event that the results of the previous trials used to establish the active control effect are of questionable relevance to the current setting. Accordingly we evaluate (49.5) with $\gamma = 0.95$, $f_0 = 0.5$, $\ln(\widehat{\lambda_p/\lambda_c}) = 0.234$ and $\text{se}[\ln(\widehat{\lambda_p/\lambda_c})] = 0.075$ thereby obtaining the non inferiority margin to be $\delta_0 = 0.044$ for the log hazard ratio and $\exp(0.044) = 1.045$ for the hazard ratio.

49.3 Design of Metastatic Colorectal Cancer Trial

In this section we will use East to design a single-look non inferiority trial comparing the test drug Xeloda (treatment t) to the active control 5FU+LV (treatment c) for the treatment of metastatic colorectal cancer. On the basis of a meta analysis of ten previous studies of the active control versus placebo (Rothman et. al. 2003), a non inferiority margin of 1.045 for λ_t/λ_c has been established. Thus we are interested in testing the null hypothesis of inferiority $H_0: \lambda_t/\lambda_c \geq 1.045$ versus the one-sided alternative hypothesis that $\lambda_t/\lambda_c < 1.045$. Suppose the trial is planned to enroll for 30 months and finish within 70 months of the last patient enrolled.

49.3.1 Single-Look Design

We will use East to create an initial single-look design having 80% power to detect the alternative hypothesis $H_1: \lambda_t/\lambda_c = 1$ with a one sided level-0.025 non-inferiority test.

To begin click **Survival: Two Samples** on the **Design** tab and then click **Parallel Design: Logrank Test Given Accrual Duration and Study Duration** as shown below.

A new screen will appear. Enter the appropriate design parameters into the dialog box

as shown below.

Design Type: **Noninferiority** Number of Looks: **1**

Test Parameters | **Accrual / Dropouts**

Test Type: **1-Sided** # of Hazard Pieces: **1** Input Method: **Median Survival Times**

Type I Error (α): **0.025**

Power: **0.8**

Sample Size (n): **Computed**

No. of Events: **Computed**

Allocation Ratio: **1**
 (n_t/n_c)

Hazard Ratio (Optional)

| | Null | Alternative |
|---|--------------|-------------|
| <input checked="" type="radio"/> Hazard Ratio (λ_t/λ_c) | 1.045 | 1 |
| <input type="radio"/> Ratio of Medians (m_t/m_c) | 0.957 | 1 |

| | Med.Surv.Time | |
|-----------|----------------|-----------------|
| Control | Treatment:Null | Treatment: Alt. |
| 18 | 17.225 | 18 |

Variance of Log Hazard Ratio

Null Alternative

The box labeled **Variance of Log Hazard Ratio** specifies whether the calculation of the required number of events is to be based on the variance estimate of the log hazard ratio under the null hypothesis or the alternative hypothesis. The default choice in East is **Null**. Most textbooks recommend this choice as well (see, for example Collett, 1994, equation (2.21) specialized to no ties). It will usually not be necessary to change this default. For a technical discussion of this issue refer to Appendix B, Section B.5.3.

Next click on the **Accrual/Dropout** tab. Here we will specify the accrual information and dropout rates. Set the accrual duration to **30** months and the study duration to **100** months in the **Accrual** box. Also, suppose that there are **5%** drop-outs per year in each arm. Enter these values as shown below.

Test Parameters | **Accrual / Dropouts**

Subjects are followed: **Until End of Study**

Accrual Info

Accrual Duration: **30** Study Duration: **100**

of Accrual Periods: **1**

| Period # | By Time | Cum. % Accrued |
|----------|---------------|----------------|
| 1 | 30.000 | 100.000 |

Piecewise Dropout Information

of Pieces: **1** Input Method: **Prob. of Dropout**

By Time: **12**

Prob. of Dropout (Control) **0.05**

Prob. of Dropout (Treatment) **0.05**

Note: Period 1 hazard rates apply after time 12.

Click on **Compute** to complete the design. The design is shown as a row in the **Output Preview** located in the lower pane of this window. You can select this design

49 Non Inferiority Trials Given Accrual Duration and Study Duration

by clicking anywhere along the row in the **Output Preview**. With Des1 selected, click the  icon to display the details of this design in the upper pane, which are shown below. You may also wish to save this design. Select Des1 in the **Output Preview** window and click the  to save this design to Workbook1 in the **Library**.

| Des 1 | |
|---|--------------------|
| Mnemonic | SU-25-LRSD |
| Test Parameters | |
| Design Type | Noninferiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.8 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Null) | 1.045 |
| Hazard Ratio (Alt.) | 1 |
| Var (Log HR) | Null |
| Accrual & Dropout Parameters | |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 1 |
| Sample Size | |
| Maximum | 18527 |
| Expected Under H0 | 18527 |
| Expected Under H1 | 18527 |
| Events | |
| Maximum | 16205 |
| Expected Under H0 | 16205 |
| Expected Under H1 | 16205 |
| Study Duration | |
| Maximum | 100 |
| Expected Under H0 | 96.743 |
| Expected Under H1 | 99.956 |
| Accrual Duration | |
| Maximum | 30 |
| Expected Under H0 | 30 |
| Expected Under H1 | 30 |

It is immediately evident that Des1 is untenable. It requires 16,205 events to be fully powered and 18,527 subjects to obtain those events within the course of the study. The problem lies with trying to power the trial to detect a hazard ratio of 1 under the alternative hypothesis. Suppose instead that the investigators actually believe that the treatment is slightly superior to the active control, but the difference is too small to be detected in a superiority trial. In that case a non-inferiority design powered at a hazard ratio less than 1 (0.95, say) would be a better option because such a trial would require fewer events.

To see this create a new design by selecting Des1 in the **Library**, and clicking the  icon on the **Library** toolbar. Then edit this design by specifying a hazard ratio of 0.95 under the alternative hypothesis as shown below.

Test Type: 1-Sided

Type I Error (α): 0.025

Power: 0.8

Sample Size (n): Computed

No. of Events: Computed

Allocation Ratio: 1
 (n_t/n_c)

of Hazard Pieces: 1

Input Method: Median Survival Times

Hazard Ratio (Optional)

Hazard Ratio (λ_t/λ_c)

Ratio of Medians (m_t/m_c)

| | Null | Alternative |
|------------------|-------|-------------|
| Hazard Ratio | 1.045 | 0.95 |
| Ratio of Medians | 0.957 | 1.053 |

| | Med.Surv.Time | |
|---------|----------------|-----------------|
| Control | Treatment:Null | Treatment: Alt. |
| 18 | 17.225 | 18.947 |

Click the **Compute** button to generate output for Des2. With Des2 selected in the **Output Preview**, click the  icon to save Des2 to the **Library**. In the **Library**, select the rows for Des1 and Des2, by holding the Ctrl key, and then click the .

49 Non Inferiority Trials Given Accrual Duration and Study Duration

icon. The upper pane will display the details of the two designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 |
|---|--------------------|--------------------|
| Mnemonic | SU-2S-LRSD | SU-2S-LRSD |
| Test Parameters | | |
| Design Type | Noninferiority | Noninferiority |
| No. of Looks | 1 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.8 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (nt / nc) | 1 | 1 |
| Hazard Ratio (Null) | 1.045 | 1.045 |
| Hazard Ratio (Alt.) | 1 | 0.95 |
| Var (Log HR) | Null | Null |
| Accrual & Dropout Parameters | | |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 |
| Sample Size | | |
| Maximum | 18527 | 3973 |
| Expected Under H0 | 18527 | 3973 |
| Expected Under H1 | 18527 | 3973 |
| Events | | |
| Maximum | 16205 | 3457 |
| Expected Under H0 | 16205 | 3457 |
| Expected Under H1 | 16205 | 3457 |
| Study Duration | | |
| Maximum | 100 | 100 |
| Expected Under H0 | 96.743 | 93.218 |
| Expected Under H1 | 99.956 | 99.87 |
| Accrual Duration | | |
| Maximum | 30 | 30 |
| Expected Under H0 | 30 | 30 |
| Expected Under H1 | 30 | 30 |

Des2 is clearly easier to implement than Des1. It requires only 3,457 events to be fully powered. This can be achieved with only 3,973 patients enrolled in the study.

49.3.2 Early Stopping for Futility

Under the null hypothesis, Des2, with 3,457 events, has expected study duration of 93.2 months. This is a very long time commitment for a trial that is unlikely to be successful. Therefore it would be a good idea to introduce a futility boundary for possible early stopping. Since we wish to be fairly aggressive about early stopping for futility we will generate the futility boundary from the Gamma(-1) β spending function. On the other hand since there no interest in early stopping for efficacy we will not use an efficacy boundary.

Create a new design by selecting Des2 in the **Library**, and clicking the  icon on the **Library** toolbar. Change the number of looks from 1 to 3.

Next, click on the **Boundary** tab. Enter the parameters as shown below. Be sure to select the **Non Binding** option. This choice gives us the flexibility to continue the trial even if a futility boundary has been crossed. Data monitoring committees usually want this flexibility; for example, to follow a secondary endpoint.

Efficacy

Boundary Family: None

Futility

Boundary Family: Spending Functions

Spending Function: Gamma Family

Parameter (γ): -1

Type II Error (β): 0.2

Futility Boundary: Z Scale

Spacing of Looks Equal Unequal

| Look # | Info. Fraction | Cum. β Spent | Futility Boundary |
|--------|----------------|--------------------|-------------------|
| 1 | 0.333 | 0.046 | -0.007 |
| 2 | 0.667 | 0.110 | -1.056 |
| 3 | 1.000 | 0.200 | -1.960 |

Click the **Compute** button to generate output for Des3. With Des3 selected in the **Output Preview**, click the icon to save Des3 to the **Library**. In the **Library**, select the rows for Des1, Des2, and Des3 by holding the Ctrl key, and then click the icon. The upper pane will display the details of the three designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 | Wbk1:Des3 |
|---|--------------------|--------------------|--------------------|
| Test Parameters | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 1 | 3 |
| Test Type | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 | 0.025 |
| Attained α | | | 0.022 |
| Power | 0.8 | 0.8 | 0.8 |
| Model Parameters | | | |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Hazard Ratio (Null) | 1.045 | 1.045 | 1.045 |
| Hazard Ratio (Alt.) | 1 | 0.95 | 0.95 |
| Var (Log HR) | Null | Null | Null |
| Boundary Parameters | | | |
| Spacing of Looks | | | Equal |
| Futility Boundary | | | Gm (-1) (NB) |
| Accrual & Dropout Parameters | | | |
| Subjects are Followed | Until End of Study | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 | 1 |
| Sample Size | | | |
| Maximum | 18527 | 3973 | 4344 |
| Expected Under H0 | 18527 | 3973 | 3965.032 |
| Expected Under H1 | 18527 | 3973 | 4312.721 |
| Events | | | |
| Maximum | 16205 | 3457 | 3780 |
| Expected Under H0 | 16205 | 3457 | 2056.327 |
| Expected Under H1 | 16205 | 3457 | 3583.036 |
| Study Duration | | | |
| Maximum | 100 | 100 | 100 |
| Expected Under H0 | 96.743 | 93.218 | 39.614 |
| Expected Under H1 | 99.956 | 99.87 | 92.714 |
| Accrual Duration | | | |
| Maximum | 30 | 30 | 30 |
| Expected Under H0 | 30 | 30 | 27.383 |
| Expected Under H1 | 30 | 30 | 29.784 |

49 Non Inferiority Trials Given Accrual Duration and Study Duration

Observe that while the sample size has been inflated to 4,344 subjects compared to Des2, the expected study duration under H_0 has been cut down to 39.6 months and the expected sample size under H_0 is 3,965. It would also be useful to simulate Des3 under a variety of scenarios for the hazard ratio. Select Des3 in the **Library** and click the  icon. You will be taken to the following simulation worksheet.

| Test Parameters | | Response Generation | | Accrual / Dropouts | | Simulation Controls | |
|----------------------|---------------------|---------------------|--|---------------------------|--|---------------------|--|
| Trial Type: | Noninferiority | | | Noninf. Margin (ln(HR0)): | | 0.044 | |
| Test Type: | 1-Sided | | | Test Statistic: | | Logrank | |
| Fix at Each Look: | Total No. of Events | | | | | | |
| Total No. of Events: | 3780 | | | | | | |
| Look # | Info. Fraction | Futility Z | | | | | |
| 1 | 0.333 | -0.007 | | | | | |
| 2 | 0.667 | -1.056 | | | | | |
| 3 | 1.000 | -1.960 | | | | | |

We wish to simulate this trial under the null hypothesis that the hazard ratio is $\exp(0.044) = 1.045$. To do this go to the **Response Generation** tab in the simulation worksheet. In this tab change the hazard ratio to 1.045 as shown below.

| Survival Information | | |
|----------------------|-----------|-------------------------------------|
| # of Hazard Pieces | 1 | Input Method: Median Survival Times |
| Med.Surv.Time | | Hazard Ratio |
| Control | Treatment | |
| 18 | 18.947 | 1.045 |

Next, click the **Simulate** button to simulate 10000 trials. A new row labeled Sim1 will appear in the **Output Preview** window. Select Sim1 in the **Output Preview** and click the  icon to save it to the **Library**. In the **Library**, double-click Sim1. A portion of the output is displayed below. (The actual values may differ, depending on the

starting seed used).

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

| Test Parameters | |
|-------------------------------------|---------------------|
| Simulation ID | Sim1 |
| Design Type | Noninferiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 4344 |
| Fix at Each Look | Total No. of Events |
| Noninferiority Margin (ln(HR0)) | 0.044 |
| Test Statistic | Logrank |
| Average Events | 2059.344 |
| Total Accrual Duration | 30 |
| Avg. Power at Termination | 0.024 |
| Response Generation Parameters | |
| HR = λ_1/λ_2 | 1.045 |
| Med. Surv. Time Control (m_1) | 18 |
| Med. Surv. Time Treatment (m_2) | 17.225 |
| Accrual / Dropouts Parameters | |
| Sample Size | 4344 |
| Subject are followed | Until End of Study |
| Accrual Duration | 30 |
| Accr. % | 100 |
| Prob. of Dropout by Time = | 12 |
| Control | 0.05 |
| Treatment | 0.05 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

☰ Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 3590.615 | 619.759 | 640.241 | 68.946 | 67.971 | 24.791 | 8.915 |
| 2 | 4344 | 1249.953 | 1270.047 | 137.441 | 136.392 | 40.215 | 14.735 |
| 3 | 4344 | 1887.31 | 1892.69 | 204.75 | 205.133 | 93.265 | 22.086 |
| Average | 3967.956 | 1013.133 | 1046.211 | 112.606 | 111.139 | 39.618 | 12.815 |

☰ Simulation Boundaries and Cumulative Boundary Crossing Probabilities

| Look # | Events | Boundaries | | Reject H0 / Unable to Reject H1 | Total Simulations | |
|---------|--------|------------|-----------------------|---------------------------------|-------------------|----------|
| | | Futility | Stopping For Futility | | Count | % |
| 1 | 1260 | -0.007 | 4994 | 0 | 4994 | 49.940% |
| 2 | 2520 | -1.056 | 8662 | 0 | 8662 | 86.620% |
| 3 | 3780 | -1.96 | 9762 | 238 | 10000 | 100.000% |
| Total % | | | 97.620% | 2.380% | | |

Simulation Seed and Elapsed Time

Starting Seed: 988853
 Total Number of Simulations: 10000
 Elapsed Time: 00.01.02

Note that 238 out of the 10000 simulations rejected the null hypothesis when it was true, thus confirming (up to Monte Carlo accuracy) that this design achieves a type-1 error of 2.5%. Also, observe that 50% of these trials have crossed the futility boundary at the very first interim look after only 24.7 months of study duration.

50 *A Note on Specifying Dropout parameters in Survival Studies*

This note gives details on specifying dropout parameters for survival studies in East. Dropout in a survival study is a competing risk. You may specify dropout rate as a hazard rate or as a probability of a subject dropping out within a specific period after entering the study. Very often, people, based on their past experience in a particular therapeutic area, are in a position to estimate likely dropout rates in a future study in the same therapeutic area. Their past experience may be that a specific percentage of subjects like 5% or 10% drop out of a study. We will explain with an example, how such estimates can be used in specifying input parameters for dropout rates in East.

Example 1: Logrank Test Given Accrual Duration and Study Duration Suppose we are designing a survival study with the following parameters:

Design Type: Superiority
Number of Looks: 3
Test Type: 1-sided
Type I Error: 0.025
Power: 0.9
Allocation Ratio: 1
Hazard Rate (Control): 0.03466 (default)
Hazard Ratio: 0.7
Hazard Rate (Treatment): 0.024 (this is computed by East given the above two inputs)
Variance of Log Hazard Ratio: Null
Boundary specification: Spending Function -Lan-DeMets (OF)
Accrual Duration: 20 months
Study Duration: 40 months

Further, it is expected that about 10% of the subjects are likely to drop out by end of the study. Now the problem is how to translate this estimate to either a hazard rate or a probability of dropout in a specific period, in the light of the facts that subjects accrue over a time period and the risk set for dropouts will be diminishing due to subjects leaving the study because of events. One way to find the right specification for dropout rate is by trial and error method. We make an initial guess and compute the design. The detailed output for the design will show estimates for sample size and maximum dropouts. If the estimated dropouts is closer to 10% of the sample size, then we can stop there. Otherwise, we have to increase or decrease the input specification for dropout rate and try again till we are able to see the estimated maximum dropouts is about 10% of the estimated maximum sample size.

We can try to create a design with the above input parameters, by entering the input values in the dialog box, in the usual way. For dropout specification, suppose, we

specify the probability of dropout as 0.1 by time 40, the study duration. This implies that the probability of a subject dropping out of the study within 40 months after entering the study is 0.1. The input dialog box for dropout information will be as shown below.

The screenshot shows a dialog box titled "Piecewise Dropout Information". It contains the following fields:

- # of Pieces: 1 (dropdown)
- Input Method: Prob. of Dropout (dropdown)
- By Time: 40 (text input)
- Prob. of Dropout (Control): 0.1 (text input)
- Prob. of Dropout (Treatment): 0.1 (text input)

Note: Period 1 hazard rates apply after time 40.

The equivalent specification in terms of hazard rate can be seen by choosing the item 'Hazard Rates' in the Input Method drop down box, which is shown below. Please apply **Increase Decimal** precision available at the top right corner of the Input dialog to get the exact results.

The screenshot shows a dialog box titled "Piecewise Dropout Information". It contains the following fields:

- # of Pieces: 1 (dropdown)
- Input Method: Hazard Rates (dropdown)
- Hazard Rate (Control): 0.00263 (text input)
- Hazard Rate (Treatment): 0.00263 (text input)

Now Compute this design and save it to a library node. If you double-click on this

50 A Note on Specifying Dropout parameters in Survival Studies

node, part of the detailed output will appear as shown below.

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|-------------------------------|-------------|---------------|---------|
| Sample Size (n) | | | |
| Maximum | 300 | 302 | 602 |
| Expected H1 | 299.65 | 299.65 | 599.3 |
| Expected H0 | 300.994 | 300.994 | 601.988 |
| Events (s) | | | |
| Maximum | 186 | 149 | 335 |
| Expected H1 | 154.934 | 120.51 | 268.466 |
| Expected H0 | 167.223 | 167.223 | 334.315 |
| Dropouts (d) | | | |
| Maximum | 14 | 17 | 31 |
| Expected H1 | 11.588 | 13.468 | 25.056 |
| Expected H0 | 12.97 | 12.976 | 25.945 |
| Maximum Information (I):83.75 | | | |

The above results show that the the maximum dropouts is only 5.1% (31/602) of maximum sample size and not the desired value of 10%. Since all the subjects are not accrued at the beginning of this 40 month duration study, specifying a subject's probability of dropping out as 0.1 within 40 months may not be appropriate. As the accrual duration is 20 months and if we assume that the average accrual duration of the subjects is 10 months, a subject may be in the study on the average for a maximum of 30 months since the maximum study duration is 40 months. So let us specify the probability of dropout in 30 months period as 0.1 as shown below.

Piecewise Dropout Information

of Pieces: Input Method:

By Time:

Prob. of Dropout (Control)

Prob. of Dropout (Treatment)

For this design, the detailed output shows the following results.

⊖ **Sample Size Information**

| | Control Arm | Treatment Arm | Total |
|-------------------------------|-------------|---------------|---------|
| Sample Size (n) | | | |
| Maximum | 304 | 305 | 609 |
| Expected H1 | 303.099 | 303.099 | 606.198 |
| Expected H0 | 304.494 | 304.494 | 608.987 |
| Events (s) | | | |
| Maximum | 186 | 149 | 335 |
| Expected H1 | 155.103 | 120.559 | 268.466 |
| Expected H0 | 167.224 | 167.224 | 334.315 |
| Dropouts (d) | | | |
| Maximum | 19 | 22 | 41 |
| Expected H1 | 15.434 | 17.28 | 32.714 |
| Expected H0 | 16.963 | 16.969 | 33.933 |
| Maximum Information (I):83.75 | | | |

Now the maximum dropouts observed is 6.7% (41/609) of the maximum sample size. So we need to increase the dropout probability to a suitable value. Let us try out 0.15 as the probability of dropout by time 30.

Piecewise Dropout Information

of Pieces: Input Method:

By Time:

Prob. of Dropout (Control)

Prob. of Dropout (Treatment)

The design obtained with the above specification for dropout rate gives the following

50 A Note on Specifying Dropout parameters in Survival Studies

results.

⊖ Sample Size Information

| | Control Arm | Treatment Arm | Total |
|--------------------------------|-------------|---------------|---------|
| Sample Size (n) | | | |
| Maximum | 311 | 313 | 624 |
| Expected H1 | 310.491 | 310.491 | 620.982 |
| Expected H0 | 311.993 | 311.993 | 623.987 |
| Events (s) | | | |
| Maximum | 186 | 149 | 335 |
| Expected H1 | 155.473 | 120.676 | 268.466 |
| Expected H0 | 167.228 | 167.228 | 334.315 |
| Dropouts (d) | | | |
| Maximum | 29 | 34 | 63 |
| Expected H1 | 23.616 | 26.902 | 50.518 |
| Expected H0 | 25.945 | 26.945 | 52.89 |
| Maximum Information (I): 83.75 | | | |

Now the percentage of maximum dropouts to maximum sample size is 10.1% and it satisfies our aim.

Note: Some users may prefer to specify dropout rates upfront in terms of dropout hazard rates instead of probability of dropout. In either case, they may want to carry out the trial and error process, described above, in terms of dropout hazard rates instead of using probability of dropout.

51 Multiple Comparison Procedures for Survival Data

As with both continuous and discrete data, it is often desired to address multiple objectives during one single trial for a survival analysis. Here, the outcome of interest is typically the time from entry until a specific event is observed (i.e. death, recurrence, medical event). As with other data outcomes, formal statistical hypothesis tests are used to support or disprove clinical claims for survival data. When objectives are formulated into a family of hypotheses, as is the case with multiple comparison procedures, type I error is inflated. Failure to compensate for this can have adverse consequences. For example, a drug could be approved even when it is no better than placebo. Multiple comparison (MC) procedures guard against this inflation of type I error due to multiple testing.

East supports the calculation of power from simulated survival data using multiple different MC procedures. The user can choose the most relevant MC procedure that provides maximum power while maintaining the FWER. East maintains strong control of FWER, which refers to the preservation of the probability of incorrectly claiming at least one null hypothesis. The difference between strong control and weak control of FWER is that weak control of FWER assumes that all hypotheses are true.

The following MC procedures are available for survival endpoints in East.

| Category | Procedure | Reference |
|---------------|-------------------------------|------------------------------------|
| P-value Based | Bonferroni | Bonferroni CE (1935, 1936) |
| | Sidak | Sidak Z (1967) |
| | Weighted Bonferroni | Benjamini Y and Hochberg Y (1997) |
| | Holm’s Step Down | Holm S (1979) |
| | Hochberg’s Step Up | Hochberg Y (1988) |
| | Hommel’s Step Up | Hommel G (1988) |
| | Fixed Sequence | Westfall PH, Krishen A (2001) |
| Fallback | Wiens B, Dimitrienko A (2005) | |

P-value based procedures strongly control the FWER regardless of the joint distribution of the raw p-values as long as the individual raw p-values are legitimate p-values. A thorough discussion on calculating the expected number of events $d(l)$ in a time-to-event trial can be found in the Appendix D.

51.0.3 Single step MC procedures

East provides p-value based single step MC procedures to compute power for a

51 Multiple Comparison Procedures for Survival Data

survival data analysis. As with continuous outcomes, these include the Bonferroni procedure, the Sidak procedure, and the weighted Bonferroni procedure.

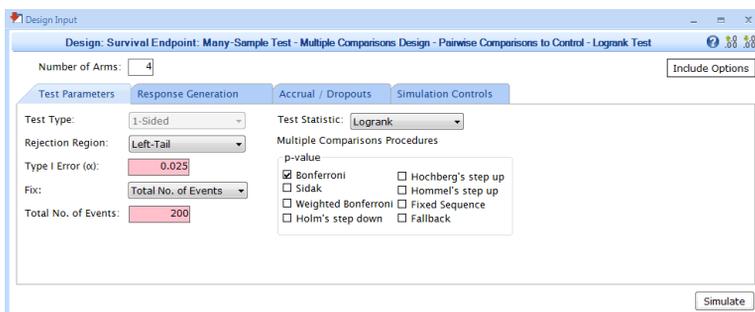
Example: STAMPEDE study

The STAMPEDE study is an ongoing, open-label, 5-stage, 6-arm randomized controlled trial using multi-arm, multi-stage (MAMS) methodology for men with prostate cancer. Started in 2005, it was the first trial of this design to use multiple arms and stages synchronously. The study population consists of men with high-risk localized or metastatic prostate cancer, who are being treated for the first time with long-term androgen deprivation therapy (ADT) or androgen suppression. The study started with 5 treatment groups:

- Standard of care (SOC) = ADT
- SOC + zoledronic acid (IV)
- SOC + docetaxel (IV)
- SOC + celecoxib, an orally administered cox-2 inhibitor
- SOC + zoledronic acid + docetaxel
- SOC + zoledronic acid + celecoxib

MAMS trials allow for the simultaneous assessment of a number of research treatments against a single control arm. By assessing several treatments in one trial, information can be acquired more quickly and with smaller numbers of patients. By combining multiple stages, this adaptive design allows continuing investments to be focused on treatments that show promise. Any therapy with insufficient evidence of activities is discontinued.

The Bonferroni and Sidak procedures in East are presented using relevant data from the STAMPEDE trial for a fixed-sample design. Under the **Design** tab in the **Survival** group, select **Many Samples - Pairwise Comparisons to Control - Logrank Test**. The following screen is displayed.



Change the default **Number of Arms:** to 6. Under the current tab, **Test Parameters**, keep the **Rejection Region:** assigned to Left-Tail, keep the default **Type I Error (α)** to 0.025, ensure that **Fix:** is set to the total number of events, and enter the value 1200. The type of **Test Statistic** used to calculate power can be identified as either the Logrank, Wilcoxon-Gehan, or Harrington-Fleming. Keep the default value of Logrank. Select both **Bonferroni** and **Sidak** for the choice of **Multiple Comparisons Procedures**.

Design: Survival Endpoint: Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Logrank Test

Number of Arms: 6

Test Parameters | Response Generation | Accrual / Dropouts | Randomization | Simulation Controls

Test Type: 1-Sided
 Rejection Region: Left-Tail
 Type I Error (α): 0.025
 Fix: Total No. of Events
 Total No. of Events: 1200

Test Statistic: Logrank

Multiple Comparisons Procedures

p-value

Bonferroni
 Sidak
 Hochberg's step up
 Hommel's step up
 Weighted Bonferroni
 Fixed Sequence
 Holm's step down
 Fallback

Select the **Response Generation** tab:

Design: Survival Endpoint: Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Logrank Test

Number of Arms: 4

Test Parameters | Response Generation | Accrual / Dropouts | Simulation Controls

Response Distribution: Exponential
 Input Method: Median Survival Times
 Time Unit: Weeks

Survival Information

| Arm | Med. Surv. Time (Weeks) | Hazard Ratio |
|---------|-------------------------|--------------|
| Control | 20.000 | |
| Arm1 | 25.000 | 0.800 |
| Arm2 | 35.000 | 0.571 |
| Arm3 | 40.000 | 0.500 |

Simulate

This is where the user can specify the **Response Distribution:** to be either Exponential, Weibull, Lognormal, or R function. The **Input Method** can be set to either Median Survival Times, Cum. % Survival, or Hazard Rates. In addition the **Time Unit** can be selected to be either days, weeks, months or years. Keep the **Response Distribution:** as Exponential, set the **Input Method** to Hazard Rates, and the **Time Unit** to years. Enter the following information into the **Survival**

51 Multiple Comparison Procedures for Survival Data

Information table:

| Arm | Hazard Rates (Per Year) | Hazard Ratio |
|---------|-------------------------|--------------|
| Control | 0.139 | |
| Arm1 | 0.104 | 0.750 |
| Arm2 | 0.104 | 0.750 |
| Arm3 | 0.104 | 0.750 |
| Arm4 | 0.104 | 0.750 |
| Arm5 | 0.104 | 0.750 |

In the next tab, **Accrual/Dropouts**, the user sees the following input dialog box which allows the specification of sample size, duration of follow-up, as well as **Accrual info** and **Piecewise Dropout Information**.

Design: Survival Endpoint: Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Logrank Test

Number of Arms: 4 Include Options

Test Parameters | Response Generation | **Accrual / Dropouts** | Simulation Controls

Sample Size: 400
 Total No. of Events: 200
 Subjects are followed: Until End of Study
 Distribution of Accrual Time: Uniform

Accrual Info

Accrual Duration Time Unit: Weeks
 # of Accrual Periods: 1 Input Method: Accrual Rates

| Period # | Starting at Time (Weeks) | Accr. Rate (Per Week) |
|----------|--------------------------|-----------------------|
| 1 | 0.0000 | 8.000 |

Piecewise Dropout Information

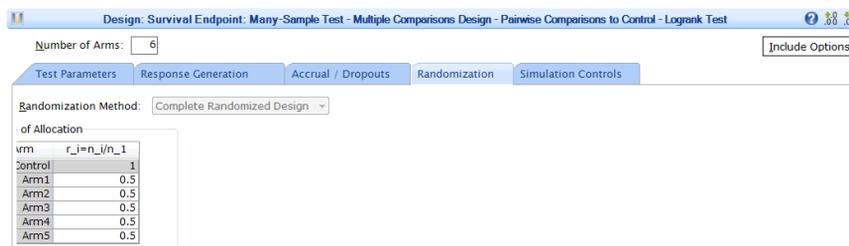
of Pieces: 0

Simulate

Set the **Sample Size** to 3400 and ensure that the **Subjects are followed**: dropdown is selected to be “Until End of Study”. The **Accrual Duration Time Unit**: is “Years”, the number of Accrual Periods is 1, and the Input Method is “Accrual Rates”. The Accrual Rate per Year is 500, starting at time 0. There is no **Piecewise Dropout Information** therefore keep the **Number of Pieces**: set to 0.

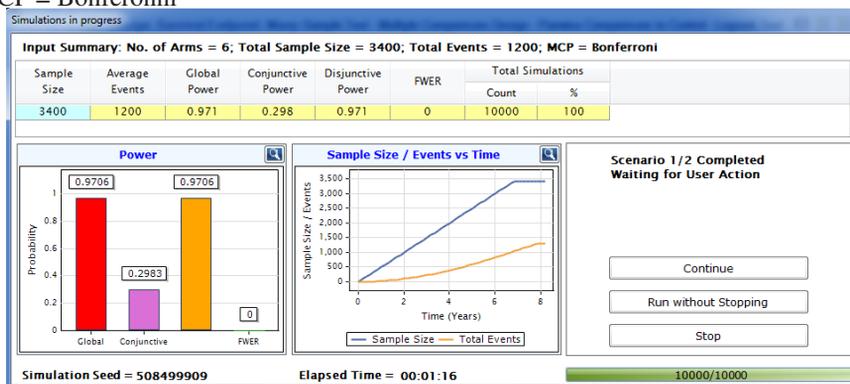
In the upper right hand of the Simulation Window, click the **Include Options** button,

and select “Randomization”. In the now new **Randomization** tab, the second column of the **Table of Allocation** table displays the allocation ratio of each treatment arm to that of control arm. The cell for control arm is always one and is not editable. Only those cells for treatment arms other than control need to be entered. The default value for each treatment arm is 1, which represents a balanced design. For the STAMPEDE, change the allocation ratio of the treatment arms to all be 0.5.



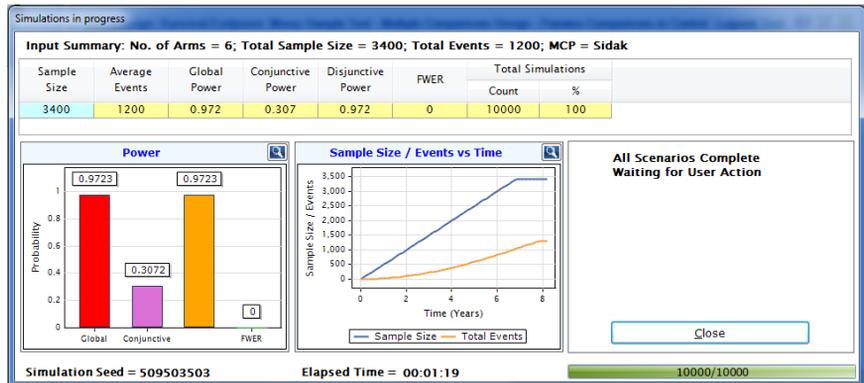
The last tab is the **Simulation Controls**. For this example, all simulation defaults can be maintained. The **Output Options** box is where the user can choose to save summary statistics for each simulation run or to save subject level data for a specific number of runs. Click **Simulate** to start the simulations. Once completed, East will add an additional row to the **Output Preview**, labeled as Sim 1.

MCP = Bonferonni



MCP = Sidak

51 Multiple Comparison Procedures for Survival Data



Note that two new simulations are displayed in the **Output Preview** window. Select the corresponding rows and save to the **Library**. Again select the two simulations and

click the **Output Summary** icon:

| | Stampede:Sim2 | Stampede:Sim3 |
|---|---------------------|---------------------|
| Mnemonic | SU-nS-MC | SU-nS-MC |
| Test Parameters | | |
| Multiple Comparison Procedure | Bonferroni | Sidak |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| No. Pairwise Comps. | 5 | 5 |
| Rejection Region | Left-Tail | Left-Tail |
| Fix | Total No. of Events | Total No. of Events |
| Test Statistic | Logrank | Logrank |
| Model Parameters | | |
| No. of Hazard Pieces | 1 | 1 |
| Accrual & Dropout Parameters | | |
| Subject Followup | Until End of Study | Until End of Study |
| Accrual Rate (Per Year) | 500 | 500 |
| No. of Accrual Periods | 1 | 1 |
| MCP Results | | |
| Overall FWER | 0 | 0 |
| Global Power | 0.971 | 0.972 |
| Conjunctive Power | 0.298 | 0.307 |
| Disjunctive Power | 0.971 | 0.972 |
| Sample Size | | |
| Maximum | 3400 | 3400 |
| Events | | |
| Maximum | 1200 | 1200 |
| Simulation Results (Overall) | | |
| Average Sample Size | 3400 | 3400 |
| Average Events | 1200 | 1200 |
| Average Accrual Duration (Years) | 6.798 | 6.798 |
| Average Study Duration (Years) | 7.448 | 7.449 |

Bonferroni and Sidak procedures have high disjunctive and global powers of about 97% and conjunctive power of about 3%.

Weighted Bonferroni procedure

The same example based on the STAMPEDE study will be used to illustrate the weighted Bonferroni procedure. Select Sim 1 in **Library** and click . In the **Design Parameters** tab, under the **Multiple Comparison Procedures** box, uncheck the

51 Multiple Comparison Procedures for Survival Data

Bonferroni box and check the **Weighted Bonferroni** box.

Number of Arms:

Test Parameters

Test Type:

Rejection Region:

Type I Error (α):

Fix:

Total No. of Events:

Response Generation

Test Statistic:

Multiple Comparisons Procedures

p-value

Bonferroni Hochberg's step up

Sidak Hommel's step up

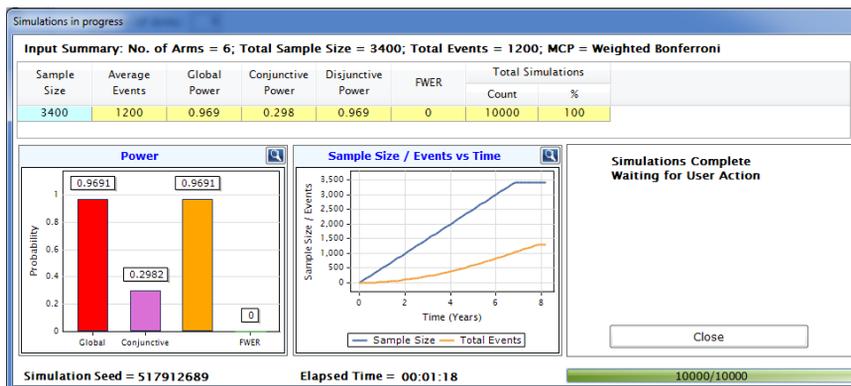
Weighted Bonferroni Fixed Sequence

Holm's step down Fallback

Treatment Arms

| Arm | Proportion of Alpha |
|---------|---------------------|
| Control | |
| Arm1 | 0.2 |
| Arm2 | 0.2 |
| Arm3 | 0.2 |
| Arm4 | 0.2 |
| Arm5 | 0.2 |

An additional table **Treatment Arms** has been added which includes a column labeled **Proportion of Alpha**. This is where to specify the proportion of total alpha to be spent in each test. If necessary, East will normalize the column total to add up to 1, and the default is to distribute the total alpha equally among all tests. Here we have 5 tests in total, therefore each of the tests have proportion of alpha as $1/5$ or 0.2 . Other proportions can be specified as well. For this example, keep the equal proportion of alpha for each test. All other values can remain the same as in the previous example. Click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional row to the **Output Preview**.



The weighted Bonferroni MC procedure has global and disjunctive power of 96.9% and conjunctive power of 29.8%. Note that, the powers in the weighted Bonferroni procedure is quite close to the Bonferroni procedure. This is because the weighted Bonferroni procedure with equal proportion is equivalent to the simple Bonferroni procedure. The exact result of the simulations may differ slightly, depending on the seed. Select the simulation in the **Output Preview** and click  icon. This will save the simulation to the workbook in the **Library**.

51.0.4 Data-driven step-down MC procedure

In the single step MC procedures, the decision to reject any hypothesis does not depend on the decision to reject other hypotheses. On the other hand, in the stepwise procedures the decision of one hypothesis test can influence the decisions on the other tests. There are two types of stepwise procedures. The first proceeds in data-driven order. The other type follows a pre-defined fixed order. Stepwise tests that are in data-driven order can proceed in either a step-down or step-up manner. East supports the Holm step-down MC procedure, which starts with the most significant comparison and continues until the test for a certain hypothesis fails. The testing procedure stops at the first non-significant comparison, and all remaining hypotheses are retained.

Holm's step-down

The STAMPEDE example will be used to illustrate Holm's step-down procedure.

Select Sim 1 in **Library** and click . In the **Design Parameters** tab under the **Multiple Comparison Procedures** box, uncheck the **Weighted Bonferonni** box and check the **Holm's Step-down** box.

Number of Arms:

Test Parameters | Response Generation | Accrual / Dropouts | Randomization | Simulation Controls

Test Type: Test Statistic:

Rejection Region:

Type I Error (α):

Fix:

Total No. of Events:

Multiple Comparisons Procedures

p-value

Bonferroni Hochberg's step up

Sidak Hommel's step up

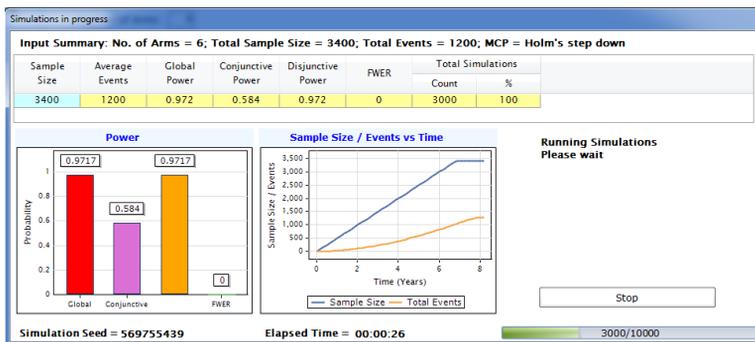
Weighted Bonferonni Fixed Sequence

Holm's step down Fallback

All other previously inputs can stay the same. To calculate the power, click **Simulate**.

51 Multiple Comparison Procedures for Survival Data

Once completed, East will add an additional row to the **Output Preview**.



Holm's step-down procedure has global and disjunctive power of 97.1% and conjunctive power of 58.6%. The exact result of the simulations may differ slightly, depending on the seed. Now select the current simulation **Output Preview** and click

 icon to save it to the workbook in the **Library**.

| Stampede:Sim4 | |
|---|---------------------|
| Mnemonic | SU-nS-MC |
| Test Parameters | |
| Multiple Comparison Procedure | Holm's step down |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| No. Pairwise Comps. | 5 |
| Rejection Region | Left-Tail |
| Fix | Total No. of Events |
| Test Statistic | Logrank |
| Model Parameters | |
| No. of Hazard Pieces | 1 |
| Accrual & Dropout Parameters | |
| Subject Followup | Until End of Study |
| Accrual Rate (Per Year) | 500 |
| No. of Accrual Periods | 1 |
| MCP Results | |
| Overall FWER | 0 |
| Global Power | 0.971 |
| Conjunctive Power | 0.586 |
| Disjunctive Power | 0.971 |
| Sample Size | |
| Maximum | 3400 |
| Events | |
| Maximum | 1200 |
| Simulation Results (Overall) | |
| Average Sample Size | 3400 |
| Average Events | 1200 |
| Average Accrual Duration (Years) | 6.798 |
| Average Study Duration (Years) | 7.446 |

51.0.5 Data-driven step-up MC procedures

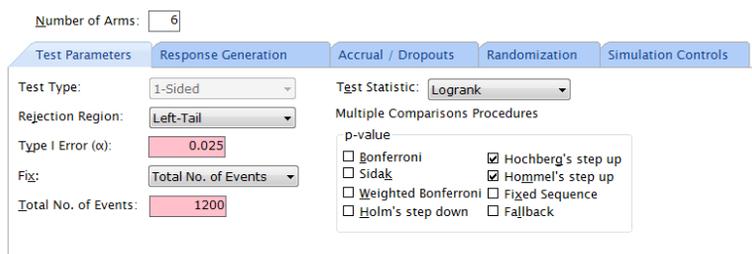
Step-up tests start with the least significant comparison and continue as long as tests are not significant until the first time when a significant comparison occurs and all remaining hypotheses will be rejected. East supports two such MC procedures for time to event data, the Hochberg step-up and the Hommel step-up. In the Hochberg step-up procedure, in i -th step $H_{(k-i)}$ is retained if $p_{(k-i)} > \frac{\alpha}{i}$. In the Hommel step-up procedure, in i -th step $H_{(k-i)}$ is retained if $p_{(k-j)} > \frac{i-j+1}{i} \alpha$ for $j = 1, \dots, i$. Fixed sequence test and fallback test are the types of tests which proceed in a predetermined

51 Multiple Comparison Procedures for Survival Data

order.

Hochberg's and Hommel's step-up procedures

Hochberg's and Hommel's step-up procedures are described below using the STAMPEDE example from the previous sections. All other design specification remains same except that we are using Hocheberg and Hommel step-up procedures in place of Holm's Step Down. Select Sim 1 in **Library** and click . In the **Design Parameters** tab, under the **Multiple Comparison Procedures** box, uncheck the **Holm's Step Down** box and check the **Hochberg's step-up** and **Hommel's step-up** boxes.



Number of Arms: 6

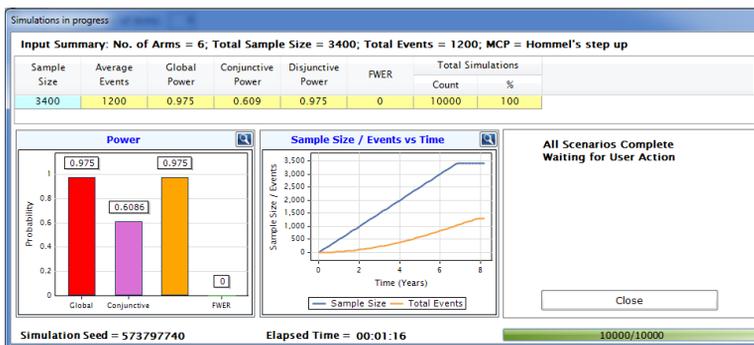
Test Parameters: Test Type: 1-Sided, Rejection Region: Left-Tail, Type I Error (α): 0.025, Fix: Total No. of Events, Total No. of Events: 1200

Response Generation: Test Statistic: Logrank

Multiple Comparisons Procedures (p-value):

- Bonferroni
- Sidak
- Weighted Bonferroni
- Holm's step down
- Hochberg's step up
- Hommel's step up
- Fixed Sequence
- Fallback

Now click **Simulate** to obtain power. Once the simulation run has completed, East will add two additional rows to the **Output Preview** window.



Hocheberg and Hommel procedures both have disjunctive and global powers of about 75% and conjunctive power about 6%. The exact result of the simulations may differ slightly, depending on the seed. Select these simulations in the **Output Preview** using Ctrl key and click  icon. This will save them to the corresponding workbook in the

Library.

| Output Summary: Stampede | | |
|---|---------------------|---------------------|
| | Stampede:Sim5 | Stampede:Sim6 |
| Mnemonic | SU-nS-MC | SU-nS-MC |
| Test Parameters | | |
| Multiple Comparison Procedure | Hochberg's step up | Hommel's step up |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| No. Pairwise Comps. | 5 | 5 |
| Rejection Region | Left-Tail | Left-Tail |
| Fix | Total No. of Events | Total No. of Events |
| Test Statistic | Logrank | Logrank |
| Model Parameters | | |
| No. of Hazard Pieces | 1 | 1 |
| Accrual & Dropout Parameters | | |
| Subject Followup | Until End of Study | Until End of Study |
| Accrual Rate (Per Year) | 500 | 500 |
| No. of Accrual Periods | 1 | 1 |
| MCP Results | | |
| Overall FWER | 0 | 0 |
| Global Power | 0.974 | 0.975 |
| Conjunctive Power | 0.613 | 0.609 |
| Disjunctive Power | 0.974 | 0.975 |
| Sample Size | | |
| Maximum | 3400 | 3400 |
| Events | | |
| Maximum | 1200 | 1200 |
| Simulation Results (Overall) | | |
| Average Sample Size | 3400 | 3400 |
| Average Events | 1200 | 1200 |
| Average Accrual Duration (Years) | 6.798 | 6.798 |
| Average Study Duration (Years) | 7.448 | 7.448 |

51.0.6 Fixed-sequence stepwise MC procedures

In data-driven stepwise procedures, we don't have any control on the order of the hypotheses to be tested. However, sometimes based on our preference or prior knowledge we might want to fix the order of tests a priori. Fixed sequence test and fallback test are the types of tests which proceed in a pre-specified order. East supports both of these procedures for survival, or time to event data.

Assume that H_1, H_2, \dots, H_{k-1} are ordered hypotheses and the order is pre-specified so that H_1 is tested first followed by H_2 and so on. Let p_1, p_2, \dots, p_{k-1} be the associated raw marginal p-values. In the fixed sequence testing procedure, for $i = 1, \dots, k - 1$, in i -th step, if $p_i < \alpha$, reject H_i and go to the next step; otherwise retain H_i, \dots, H_{k-1} and stop.

Fixed sequence testing strategy is optimal when early tests in the sequence have largest treatment effect and performs poorly when early hypotheses have small treatment effect or are nearly true (Westfall and Krishen 2001). The drawback of fixed sequence test is that once a hypothesis is not rejected no further testing is permitted. This will

51 Multiple Comparison Procedures for Survival Data

lead to lower power to reject hypotheses tested later in the sequence.

Fallback test alleviates the above undesirable feature for fixed sequence test. Let w_i be the proportion of α for testing H_i such that $\sum_{i=1}^{k-1} w_i = 1$. In the fixed sequence testing procedure, in i -th step ($i = 1, \dots, k - 1$), test H_i at $\alpha_i = \alpha_{i-1} + \alpha w_i$ if H_{i-1} is rejected and at $\alpha_i = \alpha w_i$ if H_{i-1} is retained. If $p_i < \alpha_i$, reject H_i ; otherwise retain it. Unlike the fixed sequence testing approach, the fallback procedure can continue testing even if a non-significant outcome is encountered by utilizing the fallback strategy. If a hypothesis in the sequence is retained, the next hypothesis in the sequence is tested at the level that would have been used by the weighted Bonferroni procedure. With $w_1 = 1$ and $w_2 = \dots = w_{k-1} = 0$, the fallback procedure simplifies to fixed sequence procedure.

Fixed sequence testing procedure

The STAMPEDE example is used to illustrate fixed sequence testing procedure. Select Sim 1 in **Library** and click . Under the **Design Parameters** tab in the **Multiple Comparison Procedures** box, uncheck the **Bonferonni** box and check the **Fixed Sequence** box.

Number of Arms:

Test Parameters
Response Generation
Accrual / Dropouts
Randomization
Simulation Controls

Test Type:

Rejection Region:

Type I Error (α):

Fix:

Total No. of Events:

Test Statistic:

Multiple Comparisons Procedures

p-value

Bonferroni Hochberg's step up

Sidak Hommel's step up

Weighted Bonferroni Fixed Sequence

Holm's step down Fallback

Treatment Arms

| Arm | Test Sequence |
|---------|---------------|
| Control | |
| Arm1 | 1 |
| Arm2 | 2 |
| Arm3 | 3 |
| Arm4 | 4 |
| Arm5 | 5 |

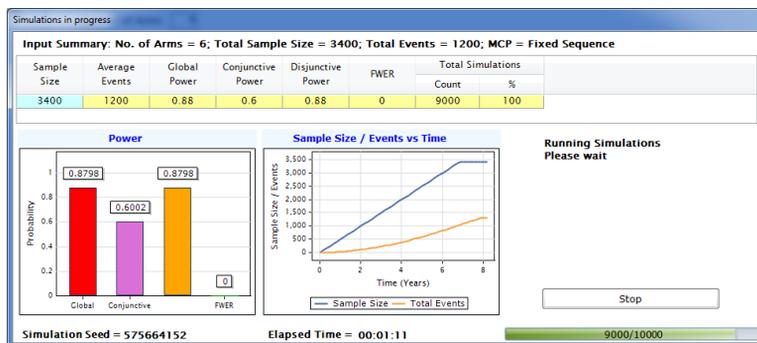
Notice that in the **Test Parameters** window a table called **Treatment Arms** has been added, which includes a column labeled **Test Sequence**. This is where the order of hypothesis tests are determined. Specify 1 for the test that will be tested first, 2 for the test that will be tested next and so on. By default East specifies 1 to the first test, 2 to the second test and so on. For optimal power in fixed sequence procedure, the early tests in the sequence should have larger treatment effects. For now we will keep the default which means that H_1 will be tested first followed by H_2 and finally H_3 will be

tested.

Number of Arms:

| Test Parameters | Response Generation | Accrual / Dropouts | Randomization | Simulation Controls | | | | | | | | | | | | | | |
|--|---------------------|--------------------|---|---------------------|-----|---------------|---------|--|------|---|------|---|------|---|------|---|------|---|
| Test Type: <input type="text" value="1-Sided"/> Rejection Region: <input type="text" value="Left-Tail"/> Type I Error (α): <input type="text" value="0.025"/> Fix: <input type="text" value="Total No. of Events"/> Total No. of Events: <input type="text" value="1200"/> | | | Test Statistic: <input type="text" value="Logrank"/> Multiple Comparisons Procedures p-value <input type="checkbox"/> Bonferroni <input type="checkbox"/> Hochberg's step up <input type="checkbox"/> Sidak <input type="checkbox"/> Hommel's step up <input type="checkbox"/> Weighted Bonferroni <input checked="" type="checkbox"/> Fixed Sequence <input type="checkbox"/> Holm's step down <input type="checkbox"/> Fallback | | | | | | | | | | | | | | | |
| Treatment Arms <table border="1"> <thead> <tr> <th>Arm</th> <th>Test Sequence</th> </tr> </thead> <tbody> <tr> <td>Control</td> <td></td> </tr> <tr> <td>Arm1</td> <td>1</td> </tr> <tr> <td>Arm2</td> <td>2</td> </tr> <tr> <td>Arm3</td> <td>3</td> </tr> <tr> <td>Arm4</td> <td>4</td> </tr> <tr> <td>Arm5</td> <td>5</td> </tr> </tbody> </table> | | | | | Arm | Test Sequence | Control | | Arm1 | 1 | Arm2 | 2 | Arm3 | 3 | Arm4 | 4 | Arm5 | 5 |
| Arm | Test Sequence | | | | | | | | | | | | | | | | | |
| Control | | | | | | | | | | | | | | | | | | |
| Arm1 | 1 | | | | | | | | | | | | | | | | | |
| Arm2 | 2 | | | | | | | | | | | | | | | | | |
| Arm3 | 3 | | | | | | | | | | | | | | | | | |
| Arm4 | 4 | | | | | | | | | | | | | | | | | |
| Arm5 | 5 | | | | | | | | | | | | | | | | | |

Now click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional row to the **Output Preview**.



The fixed sequence procedure with the specified sequence has global and disjunctive power of 87.8% and conjunctive power of 60.1%. Select the simulation in the **Output**

51 Multiple Comparison Procedures for Survival Data

Preview and click  icon.

| Stampede:Sim7 | |
|---|---------------------|
| Mnemonic | SU-nS-MC |
| Test Parameters | |
| Multiple Comparison Procedure | Fixed Sequence |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| No. Pairwise Comps. | 5 |
| Rejection Region | Left-Tail |
| Fix | Total No. of Events |
| Test Statistic | Logrank |
| Model Parameters | |
| No. of Hazard Pieces | 1 |
| Accrual & Dropout Parameters | |
| Subject Followup | Until End of Study |
| Accrual Rate (Per Year) | 500 |
| No. of Accrual Periods | 1 |
| MCP Results | |
| Overall FWER | 0 |
| Global Power | 0.878 |
| Conjunctive Power | 0.601 |
| Disjunctive Power | 0.878 |
| Sample Size | |
| Maximum | 3400 |
| Events | |
| Maximum | 1200 |
| Simulation Results (Overall) | |
| Average Sample Size | 3400 |
| Average Events | 1200 |
| Average Accrual Duration (Years) | 6.798 |
| Average Study Duration (Years) | 7.449 |

It is worthwhile to note that the fixed sequence procedure is powerful provided the hypotheses are tested in a sequence of descending treatment effects. Fixed sequence procedure controls the FWER because for each hypothesis, testing is conditional upon rejecting all hypotheses earlier in sequence. As usual, the exact result of the simulations may differ slightly, depending on the seed.

Fallback procedure

The STAMPEDE example is used to illustrate the fallback procedure. Select Sim 1 in

Library and click . Under the **Design Parameters** tab in the **Multiple Comparison Procedures** box, uncheck the **Bonferonni** box and select the **Fallback** box.

Number of Arms:

Test Parameters
Response Generation
Accrual / Dropouts
Randomization
Simulation Controls

Test Type:

Rejection Region:

Type I Error (α):

Fix:

Total No. of Events:

Test Statistic:

Multiple Comparisons Procedures

p-value

Bonferonni Hochberg's step up

Sidak Hommel's step up

Weighted Bonferonni Fixed Sequence

Holm's step down Fallback

Treatment Arms

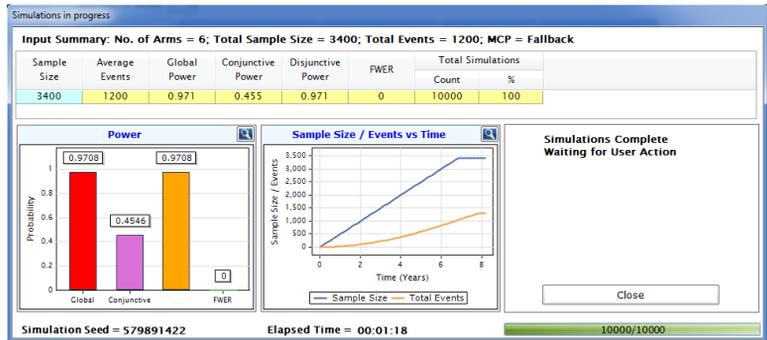
| Arm | Test Sequence | Proportion of Alpha |
|---------|---------------|---------------------|
| Control | | |
| Arm1 | 1 | 0.2 |
| Arm2 | 2 | 0.2 |
| Arm3 | 3 | 0.2 |
| Arm4 | 4 | 0.2 |
| Arm5 | 5 | 0.2 |

Notice that in the **Test Parameters** window a table called **Treatment Arms** has been added, which includes a columns labeled **Test Sequence** and **Proportion of Alpha**. In the column **Test Sequence**, the user specifies the order in which the hypotheses will be tested. Specify 1 for the test that will be tested first, 2 for the test that will be tested next and so on. By default East specifies 1 to the first test, 2 to the second test and so on. Keep the default, which means that H_1 will be tested first followed by H_2 and so on until H_5 is tested.

In the column **Proportions of Alpha**, the user specifies the proportion of total alpha to spend in each test. Ideally, the values in this column should add up to 1; if not, then East will normalize to add to 1. By default East, distributes the total alpha equally among the all tests. There are 5 tests in total, therefore each of the tests have proportion of alpha as $1/5$ or 0.2. Other proportions can be specified, however for this example, keep the equal proportion of alpha for each test. Click **Simulate** to obtain power. Once the simulation run has completed, East will add an additional row to the

51 Multiple Comparison Procedures for Survival Data

Output Preview.



The fallback procedure with the specified sequence has global and disjunctive power of 97.1% and conjunctive power of 45.5%. Select the simulation in the **Output Preview**

and click  icon to save to the workbook in the **Library**.

| Stampede:Sim8 | |
|---|---------------------|
| Mnemonic | SU-nS-MC |
| Test Parameters | |
| Multiple Comparison Procedure | Fallback |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| No. Pairwise Comps. | 5 |
| Rejection Region | Left-Tail |
| Fix | Total No. of Events |
| Test Statistic | Logrank |
| Model Parameters | |
| No. of Hazard Pieces | 1 |
| Accrual & Dropout Parameters | |
| Subject Followup | Until End of Study |
| Accrual Rate (Per Year) | 500 |
| No. of Accrual Periods | 1 |
| MCP Results | |
| Overall FWER | 0 |
| Global Power | 0.971 |
| Conjunctive Power | 0.455 |
| Disjunctive Power | 0.971 |
| Sample Size | |
| Maximum | 3400 |
| Events | |
| Maximum | 1200 |
| Simulation Results (Overall) | |
| Average Sample Size | 3400 |
| Average Events | 1200 |
| Average Accrual Duration (Years) | 6.798 |
| Average Study Duration (Years) | 7.45 |

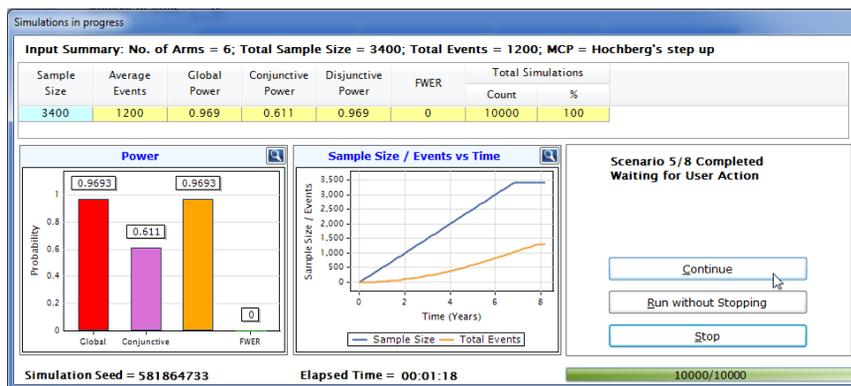
It is worthy to note that the fallback test is more robust to the misspecification of the test sequence while fixed sequence test is very sensitive to the test sequence. If the test order is incorrectly specified, fixed sequence test has very poor performance.

51.1 Comparison of MC procedures

East allows the capability of running all simulations at once in order to choose the most appropriate MC procedure. For the STAMPEDE example, Select Sim 1 in **Library** and click . Under the **Design Parameters** tab in the **Multiple**

51 Multiple Comparison Procedures for Survival Data

Comparison Procedures box, check the all boxes. Select **Simulate** and choose **Continue** as each simulation completes.



Following output displays the powers under different MC procedures.

| Multiple Comparison Procedure | Test Type | Specified α | Overall FWER | No. of Pairwise Comps. | Sample Size | Maximum Events | Rejection Region | Global Power | Conjunctive Power | Disjunctive Power | Average Sample Size | Average Events |
|-------------------------------|-----------|--------------------|--------------|------------------------|-------------|----------------|------------------|--------------|-------------------|-------------------|---------------------|----------------|
| Bonferroni | 1-Sided | 0.025 | 0 | 5 | 3400 | 1200 | Left-Tail | 0.974 | 0.311 | 0.974 | 3400 | 1200 |
| Sidak | 1-Sided | 0.025 | 0 | 5 | 3400 | 1200 | Left-Tail | 0.973 | 0.305 | 0.973 | 3400 | 1200 |
| Weighted Bonferroni | 1-Sided | 0.025 | 0 | 5 | 3400 | 1200 | Left-Tail | 0.971 | 0.301 | 0.971 | 3400 | 1200 |
| Holm's step down | 1-Sided | 0.025 | 0 | 5 | 3400 | 1200 | Left-Tail | 0.971 | 0.583 | 0.971 | 3400 | 1200 |
| Hochberg's step up | 1-Sided | 0.025 | 0 | 5 | 3400 | 1200 | Left-Tail | 0.969 | 0.611 | 0.969 | 3400 | 1200 |
| Hommel's step up | 1-Sided | 0.025 | 0 | 5 | 3400 | 1200 | Left-Tail | 0.977 | 0.619 | 0.977 | 3400 | 1200 |
| Fixed Sequence | 1-Sided | 0.025 | 0 | 5 | 3400 | 1200 | Left-Tail | 0.882 | 0.615 | 0.882 | 3400 | 1200 |
| Fallback | 1-Sided | 0.025 | 0 | 5 | 3400 | 1200 | Left-Tail | 0.973 | 0.459 | 0.973 | 3400 | 1200 |

Here we have used equal proportions for weighted Bonferroni and Fallback procedures. For the two fixed sequence testing procedures (fixed sequence and fallback), just one sequence has been used: the default (H_1, H_2, H_3). The fixed sequence procedure results in the lowest power at 88.2%. Therefore, the fixed sequence procedure easily may not be considered as most appropriate. For this example, most all procedures result in approximately 97% global and disjunctive powers. The step-up and fixed sequence procedures produce the highest conjunctive power at approximately 62% each.

Volume 7 *Adaptive Designs*

| | |
|---|-------------|
| <i>52 Introduction To Adaptive Features</i> | <i>1020</i> |
| <i>53 The Motivation for Adaptive Sample Size Changes</i> | <i>1027</i> |
| <i>54 The Cui, Hung and Wang Method</i> | <i>1055</i> |
| <i>55 The Chen, DeMets and Lan Method</i> | <i>1160</i> |
| <i>56 Muller and Schafer Method</i> | <i>1221</i> |
| <i>57 Conditional Power for Decision Making</i> | <i>1350</i> |

52 *Introduction To Adaptive Features*

This volume describes the adaptive features that can be used in the design of late stage adaptive clinical trials. The adaptive features are fully integrated into East and are invoked through simulation and calculation tools that will be described in the chapters of this volume.

The PhRMA Adaptive Design Working Group defines an adaptive trial as any clinical trial which uses accumulating data, possibly combined with external information, to modify aspects of the design without undermining the validity and integrity of the trial (see Gallo et. al.,2006). This definition is too broad for our purposes. It covers a very wide range of adaptations including dose response strategies in phase I trials, randomized play the winner rules for dose selection in early phase II trials, combination phase II/III designs, and mid-course data-dependent alterations to the later stage phase II and phase III designs. Adaptive features in East deal mainly with the last case. They extend the group sequential methodology of East in a natural way toward data-dependent changes in sample size, number of events (for event-driven trials). These adaptive extensions of group sequential designs are included in the list of adaptive methods discussed in the newly released FDA Guidance For Industry on Adaptive Design Clinical Trials for Drugs and Biologics (2010).

This volume contains Chapters 52 through 57. Chapter 52, the current chapter, describes the availability of adaptive features in East and contents of the remaining chapters in this volume. Chapter 53 provides the motivation for making adaptive changes to a late phase II or phase III trial. Three examples of actual case studies are included in this chapter; for continuous, discrete and survival endpoints, respectively. East provides two different methods for controlling type-1 error after an adaptive design change. These methods are described in the Chapter 54 and Chapter 55 respectively. They may be used to make sample size modifications for trials with normal or binomial endpoints and to make sample size and event modifications for trials with time-to-event or survival endpoints.

The third method, described in Chapter 56, offers considerable additional flexibility. All three methods are able to preserve the type-1 error in the face of data dependent changes to the study design. Each of these chapters is self-contained with a discussion of the statistical methodology followed by one or more worked examples.

A common feature in all these adaptive methods is their reliance on conditional power for making the adaptive modifications. We have developed special conditional power calculators for this purpose. The worked examples within each chapter illustrate the use of these calculators. Additionally, Chapter 57 is devoted entirely to describing how

to invoke and use the conditional power calculators.

The first adaptive method in East is the "weighted combinations" method due to Cui, Hung and Wang (1999), and Lehmacher and Wassmer (1999). In East this is referred to as the **CHW** method. In this method, the test statistic used to determine statistical significance at each interim look is a weighted combination of independent Wald statistics with pre-specified weights. This method is available for designs under Continuous, Discrete and Survival endpoints. The CHW method can be implemented at any interim look in a group sequential trial and can also be implemented multiple times. We provide simulation tools for evaluating the operating characteristics of the CHW design. This tool for the CHW method only permit sample adaptive size increases, not decreases. A special CHW Interim Monitoring Worksheet is provided to facilitate the interim monitoring and final analysis of such a trial.

The second adaptive method was proposed initially by Chen, DeMets and Lan (2004) and has now been extended by Gao, Ware and Mehta (2008) and Mehta and Pocock (2010). It is referred to as the **CDL** method. this method can be used to make sample size modifications for trials with normal or binomial endpoints and to make sample size and event modifications for trials with time-to-event or survival endpoints. The main advantage of the CDL method over the CHW method is that it permits data dependent sample size changes and event changes without the need to adjust the final test statistic with pre-specified weights. This is an attractive feature because the trial results can be presented in a conventional manner without artificially weighting the data from the two stages in ways that are difficult to explain to investigators who might be unfamiliar with the technical details of adaptive methodology. The method is, however, only applicable to two-stage adaptive designs or to multi-stage adaptive designs in which the sample size or number of events is changed at the penultimate stage. Furthermore the simulation tools for the CDL method only permit sample adaptive size increases, not decreases. This is in keeping with the recommendation of the FDA Guidance Document on Adaptive Design (2010).

The third adaptive method is referred to as the **Müller and Schäfer method**. It is based on preserving the conditional type-1 error computed at the time of the adaptation. Many authors have arrived independently at this key idea for making adaptive changes to a clinical trial. For example, it is central to the two-stage designs of Proschan and Hunsberger (1995), and the recursive combination tests of Brannath, Posch and Bauer (2002). Jennison and Turnbull (2003) claim that any fully flexible adaptive approach must respect this principle. The most general application of this principle is due to Müller and Schäfer (2001). These authors have shown explicitly that it is permissible to make any desired data dependent change to an ongoing group sequential clinical trial, possibly more than once, by the simple process of preserving the conditional type-1 error of the remainder of the trial after each change. When the

52 Introduction To Adaptive Features

only adaptive change is a change in sample size, the Müller and Schäfer method can be shown to be equivalent to the CHW method. However, the Müller and Schäfer method is not restricted to sample size changes exclusively.

The following table displays the types of designs for which adaptive methods are available in East with indications of their limitations.

Table 52.1: Adaptive Methods - Designs Applicable

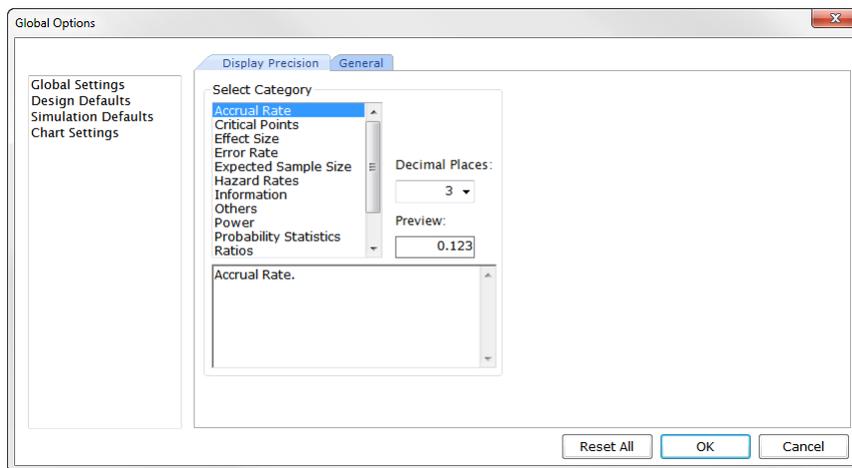
| Multi-look Design | <i>Adaptive Method^a</i> | | |
|--|------------------------------------|------------|------------|
| | CHW | CDL | MS |
| Continuous-Two Samples-Difference of Means-Superiority | <i>yes</i> | <i>yes</i> | <i>yes</i> |
| Discrete-Two Samples-Difference of Proportions-Superiority | <i>yes</i> | <i>yes</i> | <i>yes</i> |
| Discrete-Two Samples-Ratio of Proportions-Superiority | <i>yes</i> | <i>yes</i> | <i>yes</i> |
| Survival-Two Samples-Both Designs-Superiority | <i>yes</i> | <i>yes</i> | <i>yes</i> |

(a) H_0 only (1-sided); H_0 or H_1 (1-sided, binding/non-binding)

52.1 Settings



Click the **Global Options** icon in the **Home** menu to adjust default values in East 6.



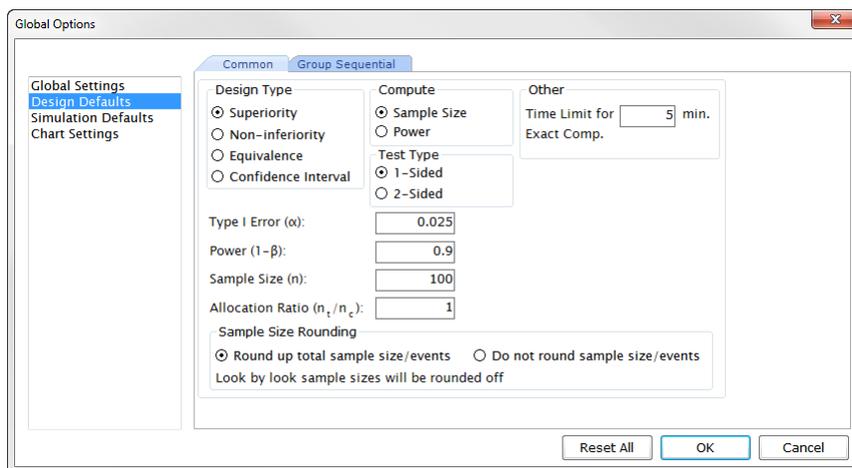
The options provided in the **Display Precision** tab are used to set the decimal places of numerical quantities. The settings indicated here will be applicable to all tests in East 6 under the **Design** and **Analysis** menus.

All these numerical quantities are grouped in different categories depending upon their usage. For example, all the average and expected sample sizes computed at simulation or design stage are grouped together under the category "Expected Sample Size". So to view any of these quantities with greater or lesser precision, select the corresponding category and change the decimal places to any value between 0 to 9.

The **General** tab has the provision of adjusting the paths for storing workbooks, files, and temporary files. These paths will remain throughout the current and future sessions even after East is closed. This is the place where we need to specify the installation directory of the R software in order to use the feature of R Integration in East 6.

52 Introduction To Adaptive Features

The **Design Defaults** is where the user can change the settings for trial design:

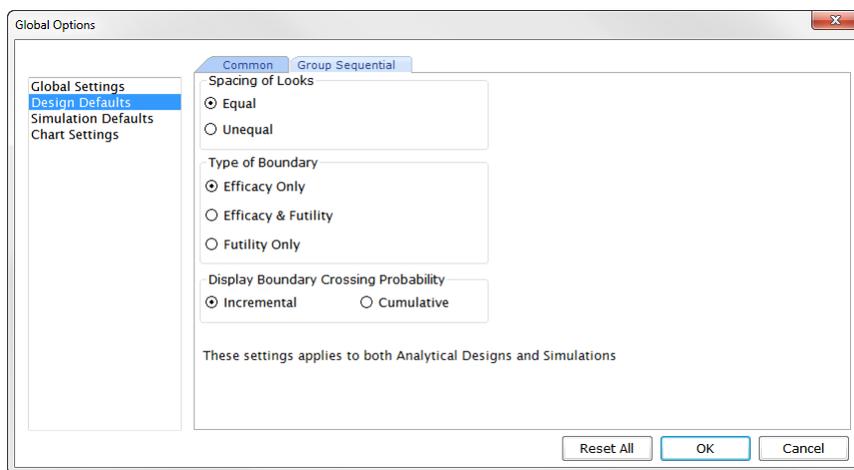


Under the **Common** tab, default values can be set for input design parameters.

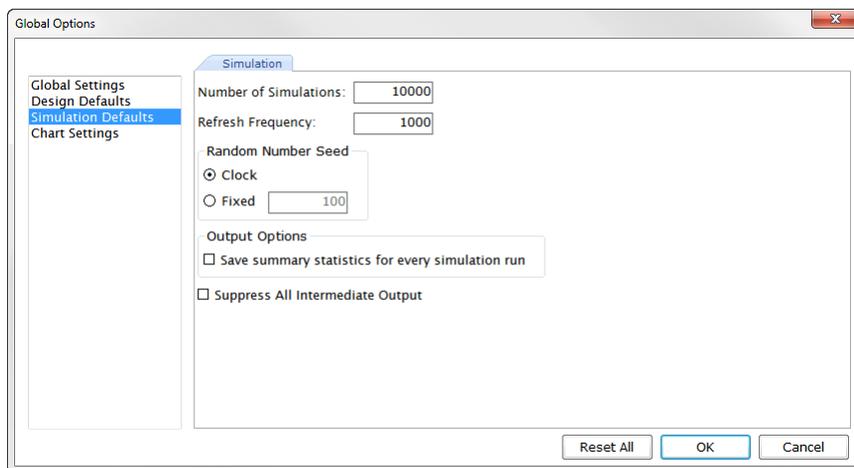
You can set up the default choices for the design type, computation type, test type and the default values for type-I error, power, sample size and allocation ratio. When a new design is invoked, the input window will show these default choices.

- **Time Limit for Exact Computation**
This time limit is applicable only to exact designs and charts. Exact methods are computationally intensive and can easily consume several hours of computation time if the likely sample sizes are very large. You can set the maximum time available for any exact test in terms of minutes. If the time limit is reached, the test is terminated and no exact results are provided. Minimum and default value is 5 minutes.
- **Type I Error for MCP**
If user has selected 2-sided test as default in global settings, then any MCP will use half of the alpha from settings as default since MCP is always a 1-sided test.
- **Sample Size Rounding**
Notice that by default, East displays the integer sample size (events) by rounding up the actual number computed by the East algorithm. In this case, the look-by-look sample size is rounded off to the nearest integer. One can also see the original floating point sample size by selecting the option "Do not round sample size/events".

Under the **Group Sequential** tab, defaults are set for boundary information. When a new design is invoked, input fields will contain these specified defaults. We can also set the option to view the Boundary Crossing Probabilities in the detailed output. It can be either Incremental or Cumulative.



Simulation Defaults is where we can change the settings for simulation:



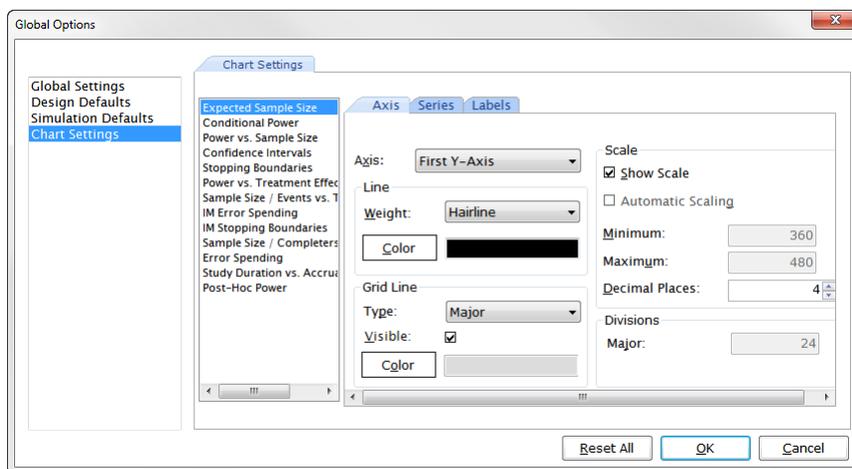
If the checkbox for "Save summary statistics for every simulation" is checked, then East simulations will by default save the per simulation summary data for all the

52 Introduction To Adaptive Features

simulations in the form of a case data.

If the checkbox for "Suppress All Intermediate Output" is checked, the intermediate simulation output window will be always suppressed and you will be directed to the **Output Preview** area.

The **Chart Settings** allows defaults to be set for the following quantities on East6 charts:



We suggest that you do not alter the defaults until you are quite familiar with the software.

53 *The Motivation for Adaptive Sample Size Changes*

In this chapter, we will highlight, through some prototypical examples, the motivation for making adaptive changes to the **sample size** in an on-going clinical trial.

Sample size is a key design input for any randomized clinical trial. Unfortunately, it is often computed in the face of inadequate knowledge about σ^2 the inter-subject variance, and δ the effect size. Economic pressures, possibly combined with competition for patients, then encourage trial investigators to make optimistic decisions about these two design parameters, a tendency that frequently results in underpowered studies. An underpowered trial is extremely undesirable, for it places human subjects at risk with a low probability of reaching a positive scientific conclusion and diverts resources that could be better utilized elsewhere. Therefore, in recent years there has been a considerable amount of research on more flexible clinical trials where the sample size is re-estimated after the clinical trial is underway, on the basis of updated information about σ^2 and δ . The updated information may arise either from external sources, from interim results of the on-going trial, or from a combination of the two. Sample size re-estimation based exclusively on updated information about σ^2 is covered in Chapter 59 in Special Topics volume of the East Manual, dealing with information based design. Here we are concerned primarily with sample size re-estimation due to updated information about δ after the study is activated.

Although statistical methods are available to make data dependent mid-course changes to sample size, the appropriateness of such sample size re-estimation has generated some debate. Critics of this type of design revision argue that the same end – ensuring adequate power at the appropriate value of δ – can be achieved more efficiently through a group sequential design (Tsiatis and Mehta, 2003; Jennison and Turnbull, 2003). This is a valid argument in settings where one is prepared to pre-specify a minimum clinically meaningful value of δ , commit a large maximum sample size to the trial up-front, and forgo the option to make data driven design changes as the trial progresses. There may be situations, however, where the flexibility to learn from the interim data and adapt the future course of the trial offsets the improved efficiency of the group sequential approach. Furthermore if the primary endpoint of the study is only measured after a lengthy follow-up, the sample size saving available through a group sequential design might be rather small. Finally, for two-stage design in which the sample size is only increased if the interim results fall in a promising zone, there may be no loss of efficiency whatsoever. We provide an example of this type at the end of Chapter 55.

53 The Motivation for Adaptive Sample Size Changes

53.1 The Benefits of Adaptive Designs

- 53.1.1 Rescuing an Underpowered On-Going Study
- 53.1.2 Designing a Study
- 53.1.3 Availing of Data from External Sources After the Study is Activated
- 53.1.4 Reducing the Sponsor's Risk

There are several reasons why it might be beneficial to allow for the possibility of a sample size increase in the middle of a group sequential clinical trial. Below we present a few real examples that we have encountered either in publications or in our consulting practice.

53.1.1 Rescuing an Underpowered On-Going Study

Cui, Hung and Wang (1999) discuss a phase III group sequential clinical trial for evaluating the effect of a new drug for prevention of myocardial infarction in patients undergoing coronary artery bypass graft surgery. The study was planned to detect a reduction in the incidence rate from 22% for placebo to 11% for the new drug with 95% power on a 1-sided level 0.025 test. The study was planned for one interim and one final look. On this basis the maximum sample size was computed to be 591 patients. There was, considerable uncertainty about the incidence rates at which the study was powered because, at that time, very little data were available on the new drug. The interim analysis results were less optimistic than was hoped at the design stage. The incidence rate in the placebo group was close to the rate specified at design, but the incidence rate in the treatment group was only 16.5%. The drop in the incidence of myocardial infarction due to the treatment was only half of what was expected. At that time, there was no valid method in the literature for increasing the sample size in mid-stream based on the observed efficacy outcome at the interim analysis. Thus the sample size was not increased and the trial eventually failed.

53.1.2 Designing a Study Given Limited Data About the Efficacy Endpoint

Consider the design of a two-arm schizophrenia trial for subjects with negative symptoms. For reasons of confidentiality, we will not reveal the names of the two drugs being tested, but will simply refer to them as the control and treatment arms, respectively. The primary clinical endpoint is the change from baseline in the Negative Symptom Assessment (NSA) at six months. There is, however, a second regulatory requirement that the new treatment must also show benefit in functional outcome as measured by a 21-item clinician rated Quality of Life Scale (QLS) measuring psychosocial functioning. The design of this trial poses inherent difficulties for the sponsor because there is very limited data from previous trials on NSA, and no data whatsoever on QLS, for patients with negative symptoms. Therefore it is not clear what values of δ one should use for calculating sample size. A pure group sequential strategy would be powered at the smallest clinically meaningful value of δ . This option is impractical in the current situation because there exists no previous experience with QLS in negative symptoms patients, and hence no notion of what constitutes a clinically meaningful effect. A second option is to run a preliminary phase II study and then follow up with a separate phase III study using the results of the earlier study, possibly combined with newly available external information, as inputs for specifying

δ and σ . This is a safe conservative choice but it does delay the time taken to reach a final conclusion about the new product. Also, with this option, the data from the phase II study cannot be combined with the data from the phase III study. A third option is to combine the phase II and phase III designs into a single integrated trial using one of the three adaptive methods provided in EastAdapt. In this option, one would start out with an initial group sequential design, powered using a sample size that reflects a compromise between the **scientific goal** of detecting the smallest clinically meaningful value of δ and the **pragmatic goal** of staying within budgetary constraints. This compromise is justified because there is still considerable uncertainty about the precise value of δ that should be used to perform the sample size calculation. Therefore the study is activated with the understanding that the current sample size assessment is preliminary and will be re-visited at a future interim analysis time point, when reliable data on the NSA and QLS endpoints become available.

53.1.3 Availing of Data from External Sources After the Study is Activated

A long-term clinical trial was activated comparing adjuvant chemotherapy to placebo in an oncology trial where the primary endpoint was survival. A retrospective analysis of historical data conducted at the design stage suggested that the study should be powered to detect a hazard ratio of 0.7. However, two years into the trial, a publication in a peer reviewed medical journal suggested that the quality of care in this disease had greatly improved, suggesting a decline in the hazard rate for the placebo arm. The investigators were very concerned by this report because it suggested that their study might now be underpowered. Although enrollment had been completed, there remained the option to adaptively extend the study duration, to see a larger number of events than had been planned at the design stage.

53.1.4 Reducing the Sponsor's Risk

From the sponsor's perspective a very attractive feature of an adaptive design is the opportunity it gives to invest in the trial in stages and thereby reduce risk. Under this scenario, the initial (first stage) investment of sample size resources might be small. The second stage investment would then be contingent on seeing promising results from the first stage. The sponsor risk is thereby reduced since the request for additional sample size resources, if made, would imply that the trial has a good chance of success. Many small biotechnology companies rely on outside investors to finance their trials. Creative adaptive designs of this type might make the financing easier.

In the remainder of the Chapter we illustrate all the above points through three case studies of actual phase 3 adaptive trials. In Section 53.2 we discuss a normal endpoint clinical trial of schizophrenia. In Section 53.3 we discuss a binomial endpoint clinical trial of acute coronary syndromes. In Section 53.4 we discuss a time-to-event

53 The Motivation for Adaptive Sample Size Changes

(survival) endpoint trial of lung cancer. These three examples will be carried forward to Chapters 54 and 55 where they will be used to demonstrate trial design and interim monitoring in East.

53.2 Normal Endpoint: Schizophrenia Trial

53.2.1 Fixed Sample Design

53.2.2 Group Sequential Design

53.2.3 The Problem of Overruns

53.2.4 Adaptive Design

53.2.4 Adaptive Sample Size Increase

53.2.5 Adaptive Design

Consider a two-arm trial to determine if there is an efficacy gain for an experimental drug relative to the industry standard treatment for negative symptoms schizophrenia. The primary endpoint is the improvement from baseline to week 26 in the Negative Symptoms Assessment (NSA), a 16-item clinician-rated instrument for measuring the negative symptomatology of schizophrenia. Let μ_t denote the difference between the mean NSA at baseline and the mean NSA at week 26 for the treatment arm and let μ_c denote the corresponding difference of means for the control arm. Denote the efficacy gain by $\delta = \mu_t - \mu_c$. The trial will be designed to test the null hypothesis $H_0: \delta = 0$ versus the one-sided alternative hypothesis that $\delta > 0$. It is expected from limited data on related studies that $\delta \geq 2$ and σ , the between-subject standard deviation, is believed to be about 7.5. In the discussion that follows we shall focus our attention on adaptive sample size adjustments due to uncertainty surrounding the true value of δ . Even though the statistical methods discussed here are applicable when there is uncertainty about either δ or σ , the adaptive approach requires careful justification primarily when δ is involved. Adaptive sample size adjustments relating to uncertainty about σ are fairly routine and non-controversial.

We shall consider fixed-sample, group sequential and adaptive design options for this study. There are advantages and disadvantages to each option with no single approach dominating over the others. We are interested, however, in exploring whether the adaptive methodology can add value to the better established fixed sample and group sequential approaches to trial design. We will see that an adaptive design alleviates to some extent the problem of “overruns” encountered by group sequential designs when the primary endpoint is observed after a lengthy follow-up period as is the case here. Additionally, we will see that an adaptive design may, in certain settings, have a more favorable risk versus benefit trade-off.

53.2.1 Fixed Sample Design

Since it is believed a priori that $\delta \geq 2$, we first create Des 1, a single-look design with 80% power to detect $\delta = 2$ using a one-sided level 0.025 test, given $\sigma = 7.5$. With these design parameters we can show that Des 1 will be fully powered if a total of 442 subjects are enrolled (221/arm). There is, however, considerable uncertainty about the true value of δ , and to a lesser extent about σ . Nevertheless it is believed that even if the true value of δ were as low as 1.6 on the NSA scale, that would constitute a clinically meaningful effect. We therefore also create Des 2, having 80% power to detect $\delta = 1.6$ using a one-sided level-0.025 test, given $\sigma = 7.5$. Des 2 requires a total

sample size of 690 subjects.

We have now proposed two design options. Under Des 1 we would enroll 442 subjects and hope that the study is adequately powered, which it will be if $\delta = 2$ and $\sigma = 7.5$. If, however $\delta = 1.6$ the power drops from 80% to 61%. There is thus a risk of launching an underpowered study for an effective drug under Des 1. Under Des 2 we will enroll 690 subjects, thereby ensuring 80% power at the smallest clinically meaningful value, $\delta = 1.6$, and rising to 94% power at $\delta = 2$. The operating characteristics of Des 1 and Des 2 are displayed side by side in Table 53.1 for values of δ between 1.6 and 2.0.

Table 53.1: Operating Characteristics of Des 1 and Des 2

| δ | Des 1 | | Des 2 | |
|----------|-------------|-------|-------------|-------|
| | Sample Size | Power | Sample Size | Power |
| 1.6 | 442 | 61% | 690 | 80% |
| 1.7 | 442 | 66% | 690 | 84% |
| 1.8 | 442 | 71% | 690 | 88% |
| 1.9 | 442 | 76% | 690 | 91% |
| 2.0 | 442 | 80% | 690 | 94% |

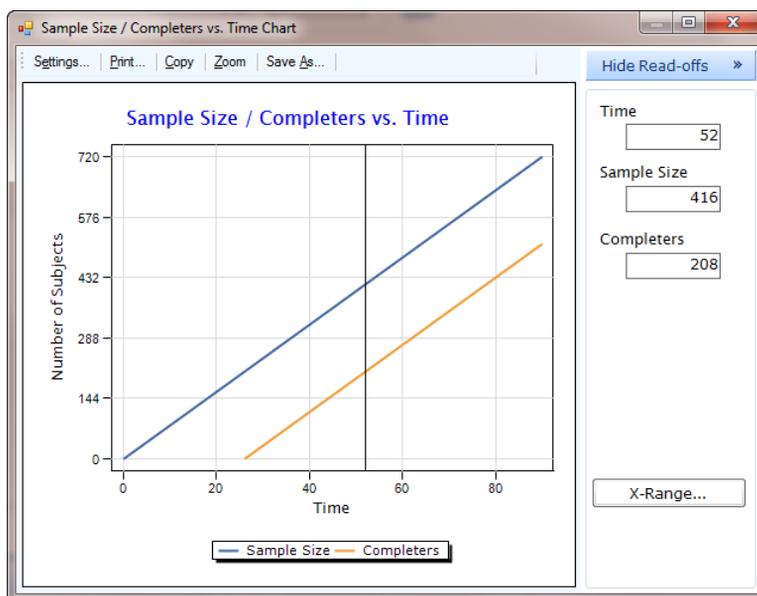
If resources were plentiful, Des 2 would clearly be the preferred option. The sponsor must, however, allocate scarce resources over a number of studies and in any case is not in favor of designing an overpowered trial. This leads naturally to considering a design that might be more flexible with respect to sample size than either of the above two single-look fixed sample designs. We will consider two types of flexible designs; group sequential and adaptive.

53.2.2 Group Sequential Design

When sample size flexibility is desired for late-stage trials, it is often appropriate to first explore the group sequential option. Let us then construct a group sequential design with one interim look and 80% power to detect $\delta = 1.6$ such that if in fact $\delta = 2$, the trial will stop early. While this would appear to be an attractive option, it is important to consider not just the saving in study duration but also the saving in the actual number of subjects randomized to the study. Since the efficacy endpoint for this trial will only be observed at week 26, the actual saving in sample size will be affected by the enrollment rate. In the current study it is anticipated that subjects will enroll at an average rate of 8 per week. The number of subjects enrolled and the number of completers over time are displayed graphically in Figure 53.1

53 The Motivation for Adaptive Sample Size Changes

Figure 53.1: Impact of Enrollment Rate and Length of Follow-Up on Trial Completion

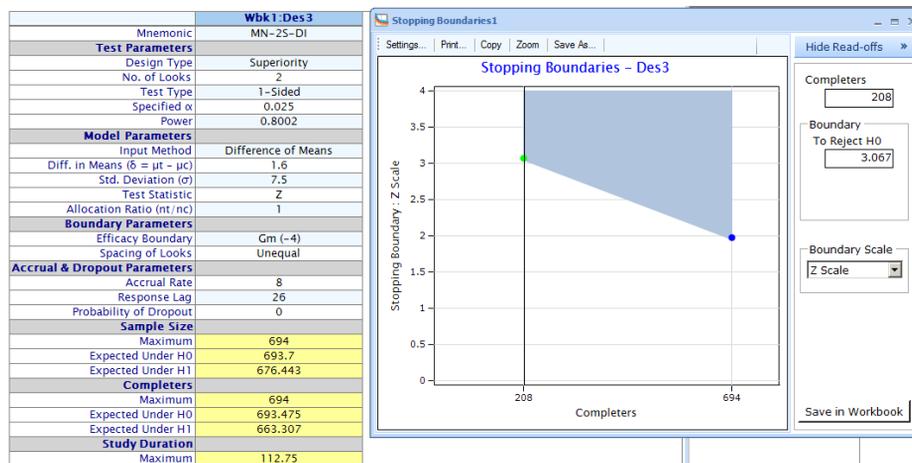


Observe that there is a 26-week horizontal separation between the two parallel lines depicting, respectively, the graph for enrollment and the graph for study completion. This 26-week gap must be taken into consideration when evaluating the savings achieved by utilizing a group sequential design.

The two major design parameters to be specified for a two-look group sequential design are the timing of the interim analysis and the amount of type-1 error to be spent. We will assume that data must be available for at least 200 completers before the trial can be terminated for efficacy so that an adequate safety profile may be developed for the study drugs. Therefore a suitable time point for the interim analysis is week 52, when we will have enrolled 416 subjects with data on 208 completers. Next we must decide on the amount of type-1 error to spend (see Lan and DeMets, 1983) for the early stopping boundary. It is generally held that the type-1 error should be spent conservatively in the early stages of a trial so as to ensure that results based on premature termination will be compelling and have the capacity to change medical practice (see Pocock, 2005). Suppose then that we use the $\gamma(-4)$ error spending function proposed by Hwang, Shih and DeCani (1990) to obtain the early stopping

boundary. The boundary thus produced resembles the conservative O’Brien-Fleming (1979) boundary. The corresponding group sequential design, having a sample size of 694, is displayed in Figure 53.2 as Des 3.

Figure 53.2: Group Sequential Design Denoted as Des 3



In Des 3 the nominal critical point for early stopping is 3.067 standard deviations. The one sided p-value corresponding to this early stopping boundary is $1 - \Phi(3.067) = 0.0011$ which, if met, would indeed be compelling enough to justify premature termination. Both Des 2 and Des 3 have 80% power to detect $\delta = 1.6$ with a one-sided level-0.025 test. Their sample size commitments too are almost the same. However, under Des 2 there is no possibility of early stopping whereas under Des 3, it is possible to stop early and thereby save on sample size. Figure 53.2 shows that the expected number of completers if in truth $\delta = 1.6$, is 663 subjects, a saving of 61 subjects compared to the maximum sample size of 694. The saving will be even more if the true value of δ is greater than 1.6. These expected savings in sample size are discussed next along with the problem of "overruns".

53.2.3 The Problem of Overruns

Care must be taken when estimating the actual sample size savings of a group sequential design. Even if the early stopping boundary is crossed at week 52 on the basis of the data from the 208 completers, we must still take into account the additional 208 randomized subjects who enrolled between week 26 and week 52 for whom the week 26 endpoint will not yet have been attained. These additional 208 subjects are

53 The Motivation for Adaptive Sample Size Changes

referred to as the "overruns". When the overruns are accounted for, the saving in sample size due to early stopping is only $694 - 416 = 278$ subjects, rather than $694 - 208 = 486$ subjects. The power and expected sample size values of the group sequential Des 3 for different choices of δ are displayed in Table 53.2. The table shows the impact of overruns on the expected sample size. For comparison we have also included corresponding power and sample size values for the fixed sample Des 2 in Table 53.2.

Table 53.2: Operating Characteristics of Plan3 (Group Sequential) and Plan2 (Fixed Sample)

| δ | Plan3 (Group Sequential) | | | | Plan2 (Fixed Sample) | |
|----------|-------------------------------|----------------------|---------------|-------|----------------------|-------|
| | Probability of Early Stopping | Expected Sample Size | | Power | SampSiz | Power |
| | | No Overruns | With Overruns | | | |
| 1.6 | 6.6% | 662 | 676 | 80% | 690 | 80% |
| 1.7 | 7.9% | 656 | 672 | 84% | 690 | 85% |
| 1.8 | 9.3% | 649 | 668 | 88% | 690 | 88% |
| 1.9 | 11.0% | 640 | 663 | 91% | 690 | 91% |
| 2.0 | 13.0% | 631 | 658 | 94% | 690 | 94% |

It is seen from Table 53.2 that Des 3 offers a modest benefit relative to Des 2. After accounting for the overruns, the expected sample sizes under Des 3 range between 658 and 676 for corresponding values of δ between 2 and 1.6, as compared to a fixed sample size of 690 under Plan2. In terms of power, Des 2 and Des 3 are practically identical. For the current trial a group sequential design with conservative error spending offers no substantial advantage over a conventional single look design with a fixed sample size. One is still faced with the dilemma of committing excessive sample size resources up front in order to ensure adequate power at $\delta = 1.6$, with limited prospects of saving on sample size in the event that $\delta = 2$.

Although in general group sequential designs do offer savings in expected sample size, their actual benefit may be diminished if a study enrolls subjects very rapidly but the primary endpoint can only be observed after a lengthy follow-up. In the current example we assumed that subjects are enrolled at the rate of 8 per week and the endpoint is observed after 26 weeks of follow-up for each subject. This resulted in 208 additional subjects being on-study who were not yet followed for 26 weeks at the time of the interim analysis. The efficiency loss due to an overrun of this magnitude was difficult to overcome. If instead the enrollment rate were to be halved to 4 subjects per week, and the endpoint were to be observed after only 12 weeks instead of 26 weeks,

there would only be an overrun of 48 subjects, and the resulting operating characteristics of the two group sequential designs would be more favorable relative to the corresponding fixed sample design. The accrual rate and the duration of follow-up are thus two extremely important design parameters for a group sequential trial.

We next consider adopting an adaptive design for this study. This is a radically different approach to trial design in which the difficulties encountered by group sequential designs – rapid accrual, delayed endpoint, and large up-front commitment of patient resources – can to some extent be mitigated.

53.2.4 Adaptive Design

To motivate the adaptive design let us recall that although the actual value of δ is unknown, the investigators believe that $\delta \geq 2$. For this reason Des 1 was constructed to have 80% power to detect $\delta = 2$. Des 2 on the other hand was constructed to have 80% power to detect $\delta = 1.6$, the smallest clinically meaningful treatment effect. If there were no resource constraints one would of course prefer to design the study for 80% power at $\delta = 1.6$ since that would imply even more power at $\delta = 2$. However, as we saw in Table 53.1, this conservative strategy carries as its price a substantially larger up-front sample size commitment which is, moreover, unnecessary if in truth $\delta = 2$. Des 3 was therefore constructed as a group sequential alternative to Des 2. Des 3 also has 80% power to detect $\delta = 1.6$ but there is a possibility of early stopping. We have seen, however, that due to the overruns problem, the expected sample size savings realized by Des 3 is small while the up-front sample size commitment is large.

The above difficulties lead us to consider whether Des 1, which was intended to detect $\delta = 2$ with 80% power and hence does not have such a large up-front sample size commitment, might be improved so as to provide some insurance against substantial power loss in the event that $\delta = 1.6$. The adaptive approach is suited to this purpose. In this approach we start out with a sample size of 442 subjects as in Des 1, but take an interim look after data are available on 208 completers. The purpose of the interim look is not to stop the trial early but rather to examine the interim data and continue enrolling past the planned 442 subjects if the interim results are promising enough to warrant the additional investment of sample size. This strategy has the advantage that the sample size is finalized only after a thorough examination of data from the actual study rather than through making a large up-front sample size commitment before any data are available. Furthermore if the sample size may only be increased but never decreased from the originally planned 442 subjects, there is no loss of efficiency due to overruns. The technical problem of avoiding inflating the type-1 error despite increasing the sample size in a data dependent manner has been solved by, among others, Cui, Hung and Wang (1999).

53 The Motivation for Adaptive Sample Size Changes

Selecting the Criteria for an Adaptive Sample Size Increase

The operating characteristics of an adaptive design depend in a complicated way on the criteria for increasing the sample size after observing the interim data. These criteria may combine objective information such as the current estimate of δ or the current conditional power with assessments of safety and with information available from other clinical trials that was not available at the start of the study. The adaptive approach provides complete flexibility to modify the sample size without having to pre-specify a precise mathematical formula for computing the new sample size based on the interim data. Therefore the full benefit of the flexibility offered by an adaptive design cannot be quantified ahead of time. Nevertheless it is instructive to investigate power and expected sample size by simulating the trial under different values of δ and applying precise pre-specified rules for increasing the sample size on the basis of the observed interim results. This will provide at least some idea, at the design stage, of the trade-off between the fixed sample or group sequential approaches and the adaptive approach. To this end we create Des 4, a design with 80% power to detect $\delta = 2$ with a one-sided level-0.025 test, based on a planned enrollment of 442 subjects. Des 4 specifies, in addition, that there will be one interim analysis after 26 weeks of follow-up data are available on the first 208 subjects enrolled. The purpose of the interim analysis is not to stop the trial early but rather to examine the interim data and decide whether a sample size increase is warranted. If no action were taken at the interim look, Des 4 would be identical to Des 1. The timing of the interim look reflects a preference for performing the interim analysis as late as possible but nevertheless while the trial is still enrolling subjects since, once the enrollment sites have closed down, it will be difficult to start them up again. Under the assumption that subjects enroll at the rate of 8 per week we will have enrolled 416 subjects by week 52; 208 of them will have completed the required 26 weeks of follow-up for the primary endpoint, and an additional 208 subjects will comprise the overruns. Only the data from the 208 completers will be used in making the decision to increase the sample size. After this decision is taken, enrollment will continue until the desired sample size is attained. The primary efficacy analysis will be based on the full 26 weeks of follow-up data from all enrolled subjects. It should be noted that unlike the group sequential setting, where the 208 overruns played no role in the early stopping decision but were still added to the final sample size, here the data from the 208 overruns will be fully utilized in the primary efficacy analysis which will only occur when all enrolled subjects have completed 26 weeks of follow-up. This is one of the advantages of the adaptive approach relative to the group sequential approach for trials with lengthy follow-up.

It remains to specify the criteria for increasing the sample size at the interim look. A well planned trial should pre-specify as far as possible the decision rules to be adopted for increasing the sample size once the interim data are available. Thereby the operating characteristics of the trial can be studied through simulation and if they are

unsatisfactory, the rules for sample size adaptation can be modified. It should be stressed, however, that in practice there is flexibility to overrule these pre-specified rules should unexpected results, either internal or external to the trial, be encountered at the time of the interim analysis. Nevertheless a precise formula for increasing the sample size must be pre-specified for purposes of simulation. While there are an infinite number of ways to construct such a formula it must address the following three questions:

- For what range of interim outcomes should a sample size increase be contemplated?
- How should the magnitude of the new sample size be calculated?
- What should be the upper limit to the sample size increase?

The answers to these questions might be driven by both clinical and business concerns, and will depend on the importance the investigators place on avoiding a false negative outcome for the current trial.

Range of Interim Outcomes for a Sample Size Increase

It is convenient to partition the sample space of possible interim outcomes into three zones; *unfavorable, promising and favorable*. An adaptive strategy is built on the premise that if the interim outcome lies in either the unfavorable or favorable zones, it is unnecessary to alter the sample size. In one case it would be risky to invest further in what appears to be a failed trial, while in the other case the trial appears slated to succeed anyway, without an additional sample size investment. Thus an adaptive sample size increase is only intended to help studies whose interim results fall in a promising zone, between these two extremes. How might these three zones be identified? One could use the interim estimate $\hat{\delta}$ or its standardized version $z = \hat{\delta}/se(\hat{\delta})$ to partition the sample space into the three zones. Alternatively one could rely on the conditional power or *probability of obtaining a positive outcome at the end of the trial, given the data already observed*. The conditional power approach is favored by most practitioners because it has a meaningful interpretation that is independent of the type of endpoint being measured, and incorporates both the current estimate of treatment effect as well as its standard error. Accordingly for the present trial we pre-specify that a sample size increase will only be contemplated if the conditional power at the interim look lies between 30% and 80%. That is, the unfavorable zone is characterized by conditional power values at most equal to 30%, the promising zone by conditional power values between 30% and 80% and the favorable zone by conditional power values at least equal to 80%.

Computing the Required Sample Size Increase

Just as at the design stage of a trial the sample size is determined by the desired power

53 The Motivation for Adaptive Sample Size Changes

(80%, say) to detect an anticipated value of δ , so also at the time of the interim analysis the new sample size may be determined by the desired conditional power (also 80%, say) to detect an anticipated value of δ . Now, however, data from the actual trial are available, and may be used to update the anticipated value of δ at which to power the trial. One could, if desired, incorporate prior beliefs, external information and current data into a value of δ at which to power the study. For simplicity however, we shall use the estimate of δ obtained at the interim analysis to recompute the sample size needed to hit the target of 80% conditional power. It is possible that this calculation could result in a reduction in the total sample size. This is permitted by the statistical methodology of adaptive designs. For the current example, however, we do not wish to decrease the sample size. Therefore if the recomputed sample size constitutes a decrease, the original sample size of 442 subjects will be used.

Specifying an Upper Limit to the Sample Size Increase

Since resources are limited, there must be an upper limit to the sample size increase, no matter what sample size is required to attain 80% conditional power. This upper limit is usually restricted to between 150% and 200% of the original sample size and is pre-specified at the start of the trial. Larger sample size increases are undesirable since they could yield statistically significant outcomes that are clinically non-significant. For the current trial we pre-specify an upper limit of 884 subjects. That is, we are prepared to double our investment in the trial, but only if the interim estimate of conditional power falls in the promising zone.

Finally, the design specifications of the adaptive Des 4 are as follows:

1. The initial sample size is 442 subjects, and has 80% power to detect $\delta = 2$ with a one-sided level-0.025 test.
2. An interim analysis is performed after data are available on 208 completers with 26 weeks of follow-up data.
3. At the interim analysis the conditional power is computed using the estimated value $\hat{\delta}$ as though it were the true value of δ . If the conditional power lies between 30% and 80%, the interim outcome is deemed to be promising.
4. If the interim outcome is promising, the sample size is re-computed so as to achieve 80% conditional power at the estimated value, $\hat{\delta}$. The original sample size is then updated to the re-computed sample size, subject to the constraint in item 5 shown below.
5. If the re-computed sample size is less than 442, the original sample size of 442 subjects is used. If the re-computed sample size exceeds 884, the sample size is curtailed to 884 subjects.

53.2.5 Operating Characteristics of Adaptive Design

Due to the complex adaptive scheme for re-computing sample size, the operating characteristics of Des 4 can best be evaluated by simulation. Table 53.3 displays power and expected sample sizes for selected values of δ between 1.6 and 2.0, based on 100,000 simulations of Des 4. For comparative purposes, corresponding power and sample size values for Des 1 are also displayed. The power of the adaptive Des 4 has

Table 53.3: Operating Characteristics of Des 1 (Fixed Sample) and Des 4 (Adaptive)

| Value of δ | Des 1(Fixed Sample) | | Des 4 (Adaptive) | |
|-------------------|---------------------|----------------------|------------------|----------------------|
| | Power | Expected Sample Size | Power | Expected Sample Size |
| 1.6 | 61% | 442 | 67% | 507 |
| 1.7 | 66% | 442 | 72% | 503 |
| 1.8 | 71% | 442 | 76% | 501 |
| 1.9 | 76% | 442 | 81% | 498 |
| 2.0 | 80% | 442 | 84% | 495 |

All Des 4 results are based on 100,000 simulated trials

increased by 6% at $\delta = 1.6$ and by 4% at $\delta = 2$ compared to Des 1. These power gains were obtained at the cost of corresponding average sample size increases of 67 subjects at $\delta = 1.6$ and 57 subjects at $\delta = 2$. The gains in power appear to be fairly modest, especially as they are offset by corresponding sample size increases. However, Des 4 offers a significant benefit in terms of risk reduction, not reflected in Table 53.3. To see this it is important to note that the sample size under Des 4 is only increased when the interim results are promising; i.e., when the conditional power at the interim analysis is greater than 30% but less than 80%. This is the very situation in which it is advantageous to increase the sample size and thereby avoid an underpowered trial. When the interim results are unfavorable (conditional power $\leq 30\%$) or favorable (conditional power $\geq 80\%$), a sample size increase is not warranted and hence the sample size is unchanged at 442 subjects for both Des 1 and Des 4. But when the interim results are promising (conditional power between 30% and 80%) the sample size is increased under Des 4 in an attempt to boost the conditional power back to 80%. It is this feature of the adaptive design that makes it more attractive than the simpler fixed sample design.

Table 53.4 displays the probability of falling into the following zones: unfavorable+futility, promising and Fav + Eff at the interim look, along with the power and expected sample size, conditional on falling into each zone, under both Des 1 and Des 4. The table highlights the key advantage of the adaptive Des 4 compared to the fixed sample Des 1; i.e., the ability to invest in the trial in stages, with the second stage

53 The Motivation for Adaptive Sample Size Changes

Table 53.4: Operating Characteristics of Des 1 and Des 4 Conditional on Interim Outcome

| δ | Interim Outcome | Probability of Interim Outcome | Power Conditional on Interim Outcome | | Expected Sample Size | |
|---|-----------------|--------------------------------|--------------------------------------|-------|----------------------|-------|
| | | | Des 1 | Des 4 | Des 1 | Des 4 |
| 1.6 | Unfav + Fut | 33% | 28% | 28% | 442 | 442 |
| | Promising | 27% | 61% | 83% | 442 | 696 |
| | Fav + Eff | 40% | 87% | 88% | 442 | 435 |
| 1.7 | Unfav + Fut | 30% | 32% | 32% | 442 | 442 |
| | Promising | 26% | 65% | 86% | 442 | 693 |
| | Fav + Eff | 45% | 89% | 90% | 442 | 435 |
| 1.8 | Unfav + Fut | 26% | 36% | 35% | 442 | 442 |
| | Promising | 25% | 69% | 89% | 442 | 691 |
| | Fav + Eff | 48% | 91% | 92% | 442 | 434 |
| 1.9 | Unfav + Fut | 23% | 41% | 39% | 442 | 442 |
| | Promising | 25% | 72% | 91% | 442 | 688 |
| | Fav + Eff | 52% | 93% | 93% | 442 | 434 |
| 2.0 | Unfav + Fut | 21% | 45% | 46% | 442 | 442 |
| | Promising | 24% | 76% | 92% | 442 | 685 |
| | Fav + Eff | 56% | 94% | 95% | 442 | 433 |
| All results are based on 100,000 simulated trials | | | | | | |

of the investment being required only if promising results are obtained at the first stage. This feature of Des 4 makes it far more attractive as an investment strategy than Des 1 which has no provision for increasing the sample size if a promising interim outcome is obtained. Suppose, for example that $\delta = 1.6$, the smallest clinically meaningful treatment effect. The trial sponsor only commits the resources needed for 442 subjects at the start of the trial, at which point the chance of success is 61%, as shown in Table 53.3. The additional sample size commitment is forthcoming only if promising results are obtained at the interim analysis, and in that case the sponsor's risk is substantially reduced because the chance of success jumps to 83%, as shown in Table 53.4. Similar results are observed for the other values of δ .

The probabilities of entering the unfavorable, promising and favorable zones at the interim analysis, displayed in Table 53.4, are instructive. Consider again the case $\delta = 1.6$. At this value of δ there is a 26% chance of landing in the promising zone and

thereby obtaining a substantial power boost under Des 4 as compared to Des 1. That is, 27% of the time the adaptive strategy can rescue a trial that is underpowered at the interim look. The chance of entering the favorable zone is 40%. That is, 40% of the time the sponsor will be lucky and have a well powered trial at the interim look without the need to increase the sample size. The remaining 33% of the time the sponsor will be unlucky and will enter the unfavorable zone from which also there is no sample size increase, and the chance of success is only 28%. These odds improve with larger values of δ .

53.3 Binomial Endpoint: Acute Coronary Syndromes Trial

53.3.1 Group Sequential Design

53.3.2 Adaptive Group Sequential Design

53.3.3 Adaptive Group Sequential Design

53.3.4 Adding a Futility Boundary

Consider a two-arm, placebo controlled randomized clinical trial for subjects with acute cardiovascular disease undergoing percutaneous coronary intervention (PCI). The primary endpoint is a composite of death, myocardial infarction or ischemia-driven revascularization during the first 48 hours after randomization. We assume on the basis of prior knowledge that the event rate for the placebo arm is 8.7%. The investigational drug is expected to reduce the event rate by at least 20%. The investigators are planning to randomize a total of 8000 subjects in equal proportions to the two arms of the study. It is easy to show that a conventional fixed sample design enrolling a total of 8000 subjects will have 83% power to detect a 20% risk reduction with a one-sided level-0.025 test of significance. The actual risk reduction is expected to be larger, but could also be as low as 15%, a treatment effect that would still be of clinical interest given the severity and importance of the outcomes. In addition, there is some uncertainty about the magnitude of the placebo event rate. For these reasons the investigators wish to build into the trial design some flexibility for adjusting the sample size. Two options under consideration are, a group sequential design with the possibility of early stopping in case the risk reduction is large, and an adaptive design with the possibility of increasing the sample size in case the risk reduction is small. In the remainder of this section we shall discuss these two options and show how they may be combined into a single design that captures the benefits of both.

53.3.1 Group Sequential Design

We first transform the fixed sample design into an 8000 person group sequential design with two interim looks, one after 4000 subjects are enrolled (50% of total information) and the second after 5600 subjects are enrolled (70% of total information). Early stopping efficacy boundaries are derived from the Lan and DeMets (1983) O'Brien-Fleming type error spending function. Let us denote this group sequential design as GSD1. The operating characteristics of GSD1 are displayed in Table 53.5. The first column of Table 53.5 is a list of potential risk reductions, defined as $100 \times (1 - \rho)\%$ where $\rho = \pi_t / \pi_c$, π_t is the event rate for the treatment arm, and π_c is the event rate for the control arm. The remaining columns display early stopping probabilities, power and expected sample size. Since the endpoint is observed within

53 The Motivation for Adaptive Sample Size Changes

48 hours, the problem of overruns that we encountered in the schizophrenia trial is negligible and may be ignored.

Table 53.5: Operating Characteristics of GSD1, a Three-Look 8000-Person Group Sequential Design

| Risk Reduction $100 \times (1 - \rho)$ | Probability of Crossing Efficacy Boundary | | | Overall Power | Expected Sample Size |
|---|---|-----------------------------|---------------------------------|---------------|----------------------|
| | At Look 1 ($N = 4000$) | At Look 2 ($N = 5600$) | At Final Look ($N = 8000$) | | |
| 15% | 0.074 | 0.183 | 0.309 | 57% | 7264 |
| 17% | 0.109 | 0.235 | 0.335 | 68% | 7002 |
| 20% | 0.181 | 0.310 | 0.330 | 82% | 6535 |
| 23% | 0.279 | 0.362 | 0.275 | 92% | 6017 |
| 25% | 0.357 | 0.376 | 0.222 | 96% | 5671 |

Table 53.5 shows that GSD1 is well powered, with large savings of expected sample size for risk reductions of 20% or more. It is thus a satisfactory design if, as is initially believed, the magnitude of the risk reduction is in the range 20% to 25%. This design does not, however, offer as good protection against a false negative conclusion for smaller risk reductions. In particular, even though 15% is still a clinically meaningful risk reduction, GSD1 offers only 57% power to detect this treatment effect. One possibility then is to increase the up-front sample size commitment of the group sequential design so that it has 80% power if the risk reduction is 15%. This leads to GSD2, a three-look group sequential design with a maximum sample size commitment of 13,853 subjects, one interim look after 6926 subjects (50% of total information) and a second interim look after 9697 subjects (70% of total information). GSD2 has 80% power to detect a risk reduction of 15% with a one-sided level-0.025 test.

Table 53.6 displays operating characteristics of GSD2 for risk reductions between 15%, and 25%. Notice that by attempting to provide adequate power at 15% risk reduction, the low end of clinically meaningful treatment effects, we have significantly over-powered the trial for values of risk reduction in the expected range of risk reductions, 20% to 25%. If, as expected, the risk reduction exceeds 20%, the large up-front sample size commitment of 13,853 subjects under GSD2 is unnecessary. GSD1 with an up-front commitment of only 8000 subjects will provide sufficient power in this setting.

From this point of view, GSD2 is not a very satisfactory design. It commits the investigators to a very large and expensive trial in order to provide adequate power in

Table 53.6: Operating Characteristics of GSD2, a Three-Look 13,853-Person Grp Sequential Design

| Risk Reduction $100 \times (1 - \rho)$ | Probability of Crossing Efficacy Boundary | | | Overall Power | Expected Sample Size |
|---|---|-----------------------------|-----------------------------------|---------------|----------------------|
| | At Look 1 ($N = 6926$) | At Look 2 ($N = 9697$) | At Final Look ($N = 13,853$) | | |
| 15% | 0.167 | 0.298 | 0.335 | 80% | 11,456 |
| 17% | 0.246 | 0.349 | 0.296 | 89% | 10,699 |
| 20% | 0.395 | 0.375 | 0.196 | 97% | 9558 |
| 23% | 0.565 | 0.329 | 0.099 | 99.3% | 8574 |
| 25% | 0.675 | 0.269 | 0.054 | 99.8% | 8061 |

the pessimistic range of risk reductions, without any evidence that the true risk reduction does indeed lie in the pessimistic range. Evidently a single group sequential design cannot provide adequate power for the "worst-case" scenario, and at the same time avoid overpowering the more optimistic range of scenarios. This leads us to consider building an adaptive sample size re-estimation option into the group sequential design GSD1, such that the adaptive component will provide the necessary insurance for the worst-case scenario, and thereby free the group sequential component to provide adequate power for the expected scenario, without a large and unnecessary up-front sample size commitment.

53.3.2 Adaptive Group Sequential Design

We convert the three-look group sequential design GSD1 into an adaptive group sequential design by inserting into it the option to increase the sample size at look 2, when 5600 subjects have been enrolled. Denote the modified design by A-GSD1. The rules governing the sample size increase for A-GSD1 are similar to the rules specified in Section 53.2.4 for the schizophrenia trial, but tailored to the needs of the current trial. The idea is to identify unfavorable, promising and favorable zones for the interim results at look 2, based on the attained conditional power. Sample size should only be increased if the interim results fall in the promising zone. Subject to an upper limit, the sample size should be increased by just the right amount to boost the current conditional power to some desired level (say 80%). The following are the design specifications for A-GSD1:

1. The starting design is GSD1 with a sample size of 8000 subjects, one interim look after enrolling 4000 subjects and a second interim look after enrolling 5600 subjects. The efficacy stopping boundaries at these two interim looks are derived from the Lan and DeMets (1983) error spending function of the

53 The Motivation for Adaptive Sample Size Changes

O'Brien-Fleming type.

2. At the second interim analysis, with data available on 5600 subjects, the conditional power is computed using the estimated value $\hat{\rho}$ as though it were the true relative risk ρ . If the conditional power is no greater than 30% the outcome is deemed to be unfavorable. If the conditional power is between 30% and 80%, the outcome is deemed to be promising. If the conditional power is at least 80%, the outcome is deemed to be favorable.
3. If the interim outcome is promising, the sample size is re-computed so as to achieve 80% conditional power at the estimated value $\hat{\rho}$. The original sample size is then updated to the re-computed sample size, subject to the constraint in item 4 shown below.
4. If the re-computed sample size is less than 8000, the original sample size of 8000 subjects is used. If the re-computed sample size exceeds 16,000, the sample size is curtailed at 16,000 subjects .

Some features of this adaptive strategy are worth pointing out. First, the sample size is re-computed on the basis of data from 5600 subjects from the trial itself. Therefore the estimate of ρ available at the interim analysis is substantially more reliable than the estimate that was used at the start of the trial to compute an initial sample size of 8000 subjects. The latter estimate is typically derived from smaller pilot studies or from other phase 3 studies in which the patient population might not be exactly the same as that of the current trial. Second, a sample size increase is only requested if the interim results are promising, in which case the trial sponsor should be willing to invest the additional resources needed to power the trial adequately. In contrast GSD2 increases the sample size substantially at the very beginning of the trial, before any data are available to determine if the large sample size is justified.

53.3.3 Operating Characteristics of Adaptive Group Sequential Design

Table 53.7 displays the power and expected sample size of the adaptive group sequential design A-GSD1. For comparative purposes corresponding power and sample size values of GSD1 are also provided.

If there is a 15% risk reduction, A-GSD1 has 6% more power than GSD1 but utilizes an additional 1093 subjects on average. It is seen that as the risk reduction parameter increases the power advantage and additional sample size requirement of A-GSD1 are reduced relative to GSD1.

The power and sample size entries in Table 53.7 were computed unconditionally, and for that reason do not reveal the real benefit that design A-GSD1 offers compared to design GSD1. As discussed previously in the schizophrenia example, the real benefit of an adaptive design is the opportunity it provides to invest in the trial in stages with

Table 53.7: Operating Characteristics of GSD1 (Group Sequential) and A-GSD1 (Adaptive Group Sequential) Designs

| Risk Reduction $100 \times (1 - \rho)$ | GSD1 (Group Sequential) | | A-GSD1 (Adaptive Group Sequential) | |
|--|-------------------------|----------------------|------------------------------------|----------------------|
| | Power | Expected Sample Size | Power | Expected Sample Size |
| 15% | 57% | 7264 | 62% | 8253 |
| 17% | 68% | 7002 | 73% | 7945 |
| 20% | 82% | 6535 | 86% | 7294 |
| 23% | 92% | 6017 | 94% | 6531 |
| 25% | 96% | 5671 | 97% | 6036 |
| All results for A-GSD1 are based on 100,000 simulated trials | | | | |

the second stage investment forthcoming only if promising results are obtained at the first stage. To explain this better it is necessary to display power and expected sample size results conditional on the zone (unfavorable, promising or favorable) into which the results of the trial fall at the second interim analysis. Accordingly Table 53.8 displays the operating characteristics of both GSD1 and A-GSD1 conditional on the zone into which the conditional power falls at the second interim analysis. The table reveals substantial gains in power for A-GSD1 compared to GSD1 at all values of risk reduction if the second interim outcome falls in the promising zone, thereby leading to an increase in the sample size. Outside this zone the two designs have the same operating characteristics since the sample size does not change. If the second interim outcome falls in the unfavorable zone, the trial appears to be headed for failure and an additional sample size investment would be risky. If the second interim outcome falls in the favorable zone, the trial is headed for success without the need to increase the sample size. Thus the adaptive design provides the opportunity to increase the sample size only when the results of the second interim analysis fall in the promising zone. This is precisely when the trial can most benefit from a sample size increase.

53.3.4 Adding a Futility Boundary

One concern with design A-GSD1 is that it lacks a futility boundary. There is thus the risk of proceeding to the end, possibly with a sample size increase, when the magnitude of the risk reduction is small and unlikely to result in a successful trial. In particular, suppose that the null hypothesis is true. In that case we can show that the power (i.e., the type-1 error) is 2.5% and the expected sample size under A-GSD1 is 8253 subjects. It might thus be desirable to include some type of futility stopping rule for the trial. In this trial the investigators proposed the following futility stopping rules at the two interim analysis time points:

53 The Motivation for Adaptive Sample Size Changes

Table 53.8: Operating Characteristics of GSD1 (Group Sequential) and A-GSD1 (Adaptive Group Sequential) Designs Conditional on Second Interim Outcome

| Risk Reduction $100 \times (1 - \rho)$ | Second Interim Outcome | Probability of Interim Outcome | Power Conditional on Second Interim Outcome | | Expected Sample Size | |
|---|------------------------|--------------------------------|---|--------|----------------------|--------|
| | | | GSD1 | A-GSD1 | GSD1 | A-GSD1 |
| 15% | Unfav + Fut | 36% | 15% | 15% | 8000 | 8000 |
| | Promising | 24% | 57% | 81% | 8000 | 12099 |
| | Fav + Eff | 40% | 94% | 94% | 6152 | 6152 |
| 17% | Unfav + Fut | 27% | 19% | 20% | 8000 | 8000 |
| | Promising | 24 % | 64% | 87% | 8000 | 11956 |
| | Fav + Eff | 49 % | 96% | 96% | 5992 | 5992 |
| 20% | Unfav + Fut | 16% | 29% | 30% | 8000 | 8000 |
| | Promising | 20% | 73% | 93% | 8000 | 11780 |
| | Fav + Eff | 64% | 98% | 98 % | 5721 | 5726 |
| 23% | Unfav + Fut | 9% | 40% | 40% | 8000 | 8000 |
| | Promising | 14% | 81% | 96% | 8000 | 11606 |
| | Fav + Eff | 77% | 99% | 99% | 5440 | 5440 |
| 25% | Unfav + Fut | 5% | 48% | 48% | 8000 | 8000 |
| | Promising | 11% | 85% | 98% | 8000 | 11449 |
| | Fav + Eff | 85% | 99.6% | 99.5% | 5250 | 5247 |

All results are based on 100,000 simulated trials

1. Stop for futility at the first interim analysis ($N = 4000$) if the estimated event rate for the experimental arm is at least 1% higher than the estimated event rate for the control arm
2. Stop for futility at the second interim analysis ($N = 5600$) if the conditional power, based on the estimated risk ratio $\hat{\rho}$, is no greater than 20%

The impact of the futility boundary on the unconditional operating characteristics of the A-GSD1 design are displayed in Table 53.9. The inclusion of the futility boundary has resulted in a dramatic saving of nearly 3000 subjects, on average, at the null hypothesis of no risk reduction. Furthermore, notwithstanding a small power loss of 2-3%, the trial continues to have well over 80% power for risk reductions of 20% or more. The trial suffers a power loss of 4% if the magnitude of the risk reduction is 15%, the low end of the range of clinical interest. In this situation, however, the unconditional power is inadequate (only 63%) even without a futility boundary. To

Table 53.9: Operating Characteristics of the A-GSD1 Design with and without a Futility Boundary

| Risk Reduction $100 \times (1 - \rho)$ | A-GSD1 with No Futility Boundary | | A-GSD1 with Futility Boundary | |
|---|----------------------------------|----------------------|-------------------------------|----------------------|
| | Power | Expected Sample Size | Power | Expected Sample Size |
| 0% | 2.4% | 8260 | 2.1% | 4866 |
| 15% | 63% | 8253 | 57% | 7063 |
| 20% | 86% | 7294 | 81% | 6726 |
| 25% | 97% | 6036 | 94% | 5846 |
| All results are based on 100,000 simulated trials | | | | |

fully appreciate the impact of the futility boundary on power and expected sample size, it is necessary to study the operating characteristics of the trial conditional on the results of the second interim analysis. These results are displayed in Table 53.10.

53 The Motivation for Adaptive Sample Size Changes

Table 53.10: Operating Characteristics of A-GSD1 Design with and without a Futility Boundary, Conditional on the Second Interim Outcome

| Risk Reduction $100 \times (1 - \rho)$ | Second Interim Outcome | Probability of Interim Outcome | Power Conditional on Second Interim Outcome | | Expected Sample Size | |
|---|------------------------|--------------------------------|---|----------|----------------------|----------|
| | | | No Fut | With Fut | No Fut | With Fut |
| 0% | Unfav + Fut | 93% | 0.5% | 0.1% | 8000 | 4370 |
| | Promising | 5% | 15% | 15% | 13030 | 12916 |
| | Fav + Eff | 2% | 65% | 64% | 7017 | 6928 |
| 15% | Unfav + Fut | 38% | 15% | 5% | 8000 | 5093 |
| | Promising | 23 % | 81% | 81% | 12099 | 11950 |
| | Fav + Eff | 39 % | 94% | 94% | 5992 | 6152 |
| 20% | Unfav + Fut | 18% | 30% | 9% | 8000 | 5264 |
| | Promising | 18% | 93% | 92% | 11780 | 11670 |
| | Fav + Eff | 64% | 98% | 98% | 5726 | 5711 |
| 25% | Unfav + Fut | 6% | 48% | 14% | 8000 | 5354 |
| | Promising | 10% | 98% | 97% | 11449 | 11370 |
| | Fav + Eff | 84% | 99.5% | 99.5% | 5247 | 5274 |
| All results are based on 100,000 simulated trials | | | | | | |

It is seen that the presence of the futility boundary does not cause any loss of power for trials that enter the promising or favorable zones at the second interim analysis. Additionally the presence of the futility boundary causes the average sample size to be reduced substantially in the unfavorable zone while remaining the same in the other two zones. In effect the futility boundary terminates a proportion of trials that enter the unfavorable zone thereby preventing them from proceeding to conclusion. It has no impact on trials that enter the promising or favorable zones.

53.4 Survival Endpoint: Lung Cancer Trial

A two-arm multi-center randomized clinical trial is planned for subjects with advanced metastatic non-small cell lung cancer with the goal of comparing the current standard second line therapy (docetaxel+cisplatin) to a new docetaxel containing combination regimen. The primary endpoint is overall survival (OS). The study is required to have one-sided $\alpha = 0.025$, and 90% power to detect an improvement in median survival, from 8 months on the control arm to 11.4 months on the experimental arm, which corresponds to a hazard ratio of 0.7. A group sequential design is adopted with an efficacy boundary derived from the Lan and DeMets (1983) O'Brien-Fleming type spending function and a futility boundary derived from the γ -spending function of Hwang, Shih and DeCani (1990) with parameter $\gamma = -5$. It is decided, with the help of the East software, to keep the study open for a maximum of 334 OS events, with one interim analysis after 167 events (50% of the total information), whereby a 1-sided level-0.025 group sequential logrank test will have 90% power to detect a hazard ratio of 0.7. As this is an event-driven trial, sample size does not play a direct role in the above power calculation. Nevertheless the rate of accrual, duration of accrual and duration of follow-up will affect the total study duration or time needed to obtain 334 events. Again, with the help of East, it is determined that by enrolling 483 subjects over a two year period and following them for an additional 6 months, the required 334 OS events can be expected to arrive by the end of the follow-up period.

Now the assumption of 8 months for median survival on the control arm is based on published results from a previously completed large, well-controlled trial. There is less data available on the experimental arm. It is thus possible, either because the new treatment is somewhat less effective than anticipated or because of improved standard of care for patients on the control arm, that the underlying hazard ratio could be larger than 0.7. If this were the case, the study would be underpowered. For example, if the true hazard ratio was 0.77, an effect that is still considered clinically meaningful, the power of a 483-subject study would drop from 90% to 67.2%. Thus one possibility would be to design the trial from the very beginning to have 90% power to detect a hazard ratio of 0.77. East shows that such a trial would require 621 events. In order to complete the trial in 30 months it would be necessary to enroll 878 subjects over 24 months with an additional 6 months of follow-up.

The sponsor is either unable or unwilling to make such a large sample size commitment up-front purely on the basis of the limited prior data available on the new compound. However, since an independent data monitoring committee (DMC) will be reviewing the interim efficacy data in an unblinded fashion at 50% of the total information, the sponsor might be prepared to authorize the investment of additional resources on the recommendation of this committee. In a manner analogous to the pre-specification of group sequential boundaries for early stopping, the sponsor must pre-specify to the DMC the precise data dependent rules for increasing the number of

53 The Motivation for Adaptive Sample Size Changes

events and sample size at the time of the interim analysis.

These rules follow the same basic structure as was adopted in Section 53.2 for the schizophrenia trial and in Section 53.3 for the acute coronary syndromes trial. The sample space of possible interim outcomes is partitioned into three zones; *unfavorable/futility*, *promising*, *favorable/efficacy*. The partitioning utilizes conditional power (CP) evaluated at the current estimate of hazard ratio with the initial specification of 334 events. The promising zone is of the form $CP_{\min} \leq CP < CP_{\max}$. To the left of the promising zone lies the unfavorable/futility zone while to the right of the promising zone lies the favorable/efficacy zone. If the data fall in the promising zone the number of events and sample size are increased by a pre-specified formula that is written into the DMC charter. If the interim data fall in the unfavorable/futility zone there is either no change in the initial design or an early termination because the futility boundary is crossed. Similarly if the interim data fall in the favorable/efficacy zone, there is either no change in the initial design or an early termination because the efficacy boundary is crossed. The choice of CP_{\min} , CP_{\max} and the rules for increasing resources in the promising zone require are best determined with the help of the simulation tools available in East. In Chapter 54, Section 54.5.3 we demonstrate how the EastSurvAdapt module of East may be used to simulate different criteria for increasing event and sample size resources and thereby obtain an adaptive design that best satisfies the goals of the trial within the resource constraints imposed on the sponsor. Based on these simulation results it has been decided to implement an adaptive increase in the number of events by 50% (from 334 to 501) if the interim results fall in the promising zone, here defined as conditional power between $CP_{\min} = 30\%$ and $CP_{\max} = 90\%$. It has further been decided that the sample size will be increased in the same ratio as the increase in events.

The operating characteristics of the lung cancer trial are displayed in Tables 53.11 and 53.12 and 53.13 for underlying hazard ratios of 0.77, 0.75 and 0.70 respectively. In each table the classical group sequential design and the adaptive group sequential design are compared with respect to power, average study duration and average number of subjects.

Table 53.11: Operating Characteristics of Optimistic Design (powered to Detect HR=0.70) under the Pessimistic Scenario (true HR=0.77)

| 10,000 Simulations Under the Pessimistic Scenario that HR = 0.77 | | | | | | | |
|--|---------|---------|-------|-------------------|--------|---------------|-------|
| Zone | P(Zone) | Power | | Duration (months) | | # of Subjects | |
| | | NonAdpt | Adapt | NonAdpt | Adapt | NonAdpt | Adapt |
| Unf+Fut | 30% | 29% | 29% | 27.8 | 27.82 | 469 | 468 |
| Prom | 34% | 69% | 85% | 29.2 | 31.03 | 483 | 712 |
| Fav+Effic | 36% | 92% | 94% | 26.2 | 26.18 | 450 | 451 |
| Total | — | 66% | 71% | 27.7 | 28.713 | 467 | 548 |

Table 53.12: Operating Characteristics of Optimistic Design (powered to Detect HR=0.70) under the Semi-Pessimistic Scenario (true HR=0.75)

| 10,000 Simulations Under the Semi-Pessimistic Scenario that HR = 0.75 | | | | | | | |
|---|---------|---------|-------|-------------------|-------|---------------|-------|
| Zone | P(Zone) | Power | | Duration (months) | | # of Subjects | |
| | | NonAdpt | Adapt | NonAdpt | Adapt | NonAdpt | Adapt |
| Unf+Fut | 24% | 35% | 36% | 28.2 | 28.3 | 471 | 471 |
| Prom | 34% | 73% | 89% | 29.4 | 32.1 | 483 | 712 |
| Fav+Effic | 42% | 96% | 95% | 25.8 | 25.9 | 446 | 446 |
| Total | — | 74% | 79% | 27.6 | 28.6 | 465 | 542 |

The results follow a similar pattern to what was observed in the previous two examples. Let us focus first on the simulation results when the underlying hazard ratio is 0.77. This is the setting where the adaptive design can play an important role since a hazard ratio of 0.77 is still clinically meaningful, and yet the sponsor is unable to command the resources that would be required to guarantee 90% power with a non-adaptive design. Row 4 of Table 53.11 shows that the adaptive design produces about a 6% gain in overall power, from 66% to 71%, at an average cost of about 1 additional month of study duration and 81 additional subjects. The real appeal of the adaptive design, however, is more evident when the overall simulation results are partitioned into the three zones. It is then seen from Table 53.11 that the interim outcome will fall in the unfavorable/futility zone 30% of the time, in which case the prospects for a successful trial are equally bleak for both the classical and adaptive designs, but no additional resources are committed to the adaptive trial. The interim outcome will fall in the favorable/efficacy zone 36% of the time, in which case the

53 The Motivation for Adaptive Sample Size Changes

Table 53.13: Operating Characteristics of Optimistic Design (powered to Detect HR=0.70) under the Optimistic Scenario (true HR=0.70)

| 10,000 Simulations Under the Optimistic Scenario that HR = 0.70 | | | | | | | |
|---|---------|---------|-------|-------------------|-------|---------------|-------|
| Zone | P(Zone) | Power | | Duration (months) | | # of Subjects | |
| | | NonAdpt | Adapt | NonAdpt | Adapt | NonAdpt | Adapt |
| Unf+Fut | 12% | 54% | 56% | 28.9 | 29.1 | 475 | 476 |
| Prom | 27% | 87% | 97% | 29.8 | 32.4 | 483 | 712 |
| Fav+Effic | 61% | 98% | 98% | 25.1 | 25.1 | 436 | 436 |
| Total | — | 90% | 93% | 26.9 | 27.6 | 454 | 516 |

prospects are excellent for both the classical and adaptive designs, and again, no additional resources are committed to the adaptive trial. The remaining 34% of the time the interim outcome will fall in the promising zone and this is where the adaptive design will help by boosting up the power from 69% to 85%. To be sure the study duration and sample size will also increased in the promising zone. Presumably, however, the power gain justifies the use of these additional resources. In summary, the additional resources will be called up to boost power only if they can make a difference to the chance of a successful outcome for the trial. Table 53.12 demonstrates that these results are similarly compelling if the true hazard ratio is 0.75. If, however, the true hazard ratio is 0.7 Table 53.13 shows that the trial as initially designed has adequate power without the need for any adaptation of events or sample size. There is now a 27% chance of landing in the promising zone and adding resources in order to boost power from 87% to 97%. In this setting the trial would be overpowered and some of the additional resources might not have been needed. The sponsor cannot of course know what the true hazard ratio is, and must weigh the likelihood of incurring these additional costs against the possibility of a loss to the patient population, and also a financial loss to the sponsor, if the study should fail despite the treatment difference being clinically meaningful.

53.5 Concluding Remarks

Many small companies with new molecules or technologies under development often rely on outside investors or large pharmaceutical companies for financing their phase 3 trials. The two-stage nature of the investment, with the second installment being obligated only if the interim results have significantly increased the odds of success, might make the adaptive design more attractive to outside investors than a conventional design requiring a fixed investment up-front, even when the two designs have an equivalent unconditional risk profiles. Simulations, performed prior to starting the

trial, are necessary to quantify the risks and benefits involved in selecting an adaptive design in preference to a conventional fixed sample or group sequential design, and to enable the sponsor to make an informed decision.

A major additional benefit of the adaptive approach is flexibility. The adaptive methodology controls the type-1 error even if the pre-specified criteria for increasing the sample size are overruled at the interim analysis. This might be desirable for a variety of reasons both internal and external to the current trial. For example, in addition to observing a promising outcome at the interim time analysis, the safety profile for the test drug might turn out to be far superior to what was originally anticipated, and this might make the new drug more competitive in the marketplace. One could therefore justify increasing the sample size by a larger amount than that determined by the pre-specified rules, and thereby further reduce the chances of a false negative outcome. Another possible situation in which one might overrule the pre-specified criteria for sample size change would be if compelling results from other clinical trials on comparable populations, treated with the same class of drugs became available and caused the sponsor to revise the value of δ at which to power the current study. Ideally one would wish to adhere strictly to the pre-specified criteria for sample size change since the operating characteristics of the design would change if they were overruled. This would certainly be the preference of regulatory authorities. As a practical matter, however, it is not possible to anticipate every contingency under which a sample size change is desirable. It is a strength of the adaptive approach that the validity of the statistical test at the end of the trial is not affected by unanticipated developments arising over the course of the clinical trial that necessitate making changes to the pre-specified criteria for sample size adaptation.

Adaptive trials require very careful up-front planning. An independent interim analysis review committee (IARC) must be appointed with the responsibility to actually implement the adaptive decision rules. A charter listing the members of the IARC, describing their roles and responsibilities, and providing the details of the proposed adaptations must be created. The charter should also discuss the steps that will be taken to ensure that the interim results remain confidential, as premature disclosure of interim results to the trial investigators could compromise the trial. Finally, regulatory approval must be secured in advance through a special protocol assessment (SPA). For this purpose the sponsor is required to submit the protocol, the charter and the simulations backing up the statistical validity of the proposed adaptive approach in good time.

Logistical and operational issues must also be considered. In a fixed sample study the total sample size is determined in advance. In a traditional group sequential study, the

53 *The Motivation for Adaptive Sample Size Changes*

maximum sample size is determined in advance. In an adaptive study, however, the maximum sample size might be increased at an interim look thereby further complicating the management of the trial, especially as it relates to patient recruitment and drug supply. Because of all these complexities an adaptive design might not always be the right choice. The more established fixed sample and group sequential designs should always be evaluated alongside an adaptive design. Simulations play a crucial role in understanding the operating characteristics of an adaptive design and deciding whether it is an appropriate choice for the trial under consideration. There should be a tangible, quantifiable benefit arising from the decision to take the adaptive route.

54

The Cui, Hung and Wang Method

This chapter discusses the Cui, Hung and Wang (1999) (CHW) method for adaptive sample size modification of an on-going two-arm, K -look group sequential clinical trial. The method is based on making a sample size modification, if required, each time that an interim analysis is performed. The interim monitoring continues in this way until either a boundary is crossed or the K looks are exhausted. Since the changes to the sample size may be based on unblinded analyses of the accruing data, the test statistic is not the usual Wald statistic utilized for monitoring a conventional group sequential design. Instead the test statistic comprises of a weighted sum of incremental Wald statistics with weights that are pre-specified and chosen appropriately so as to preserve the type-1 error. This test statistic was proposed independently by Cui, Hung and Wang (1999) and by Lehmacher and Wassmer (1999). We shall refer to this test statistic as the CHW statistic and to this method of making adaptive sample size modifications as the CHW method. The CHW method is only valid for adaptive designs involving data dependent alterations in the sample size. The operating characteristics of any CHW design are obtained through simulation using a special **Sample Size Re-estimation** tab. Interim monitoring is performed through a special **CHW Interim Monitoring** dashboard.

In Section 54.1, we provide a quick review of the underlying theory for normal and binomial endpoints. In Section 54.2 we show how these same results can be extended for trials with survival or time-to-event endpoints. (Hereafter we shall use the terms "survival" or "time-to-event" synonymously.) In Section 54.3, we illustrate the method for a normal endpoint adaptive design. In Section 54.4 we illustrate the method for a binomial endpoint adaptive design. In Section 54.5 we illustrate the method for a time-to-event adaptive design. These three designs were discussed at length in Chapter 53. Here we illustrate how to use the adaptive modules of East to simulate and monitor them. As already stated in the introductory chapter to this volume, we provide two such adaptive packages, East[®] Adapt and East[®] SurvAdapt. The East[®] Adapt package is required for studies with normal or binomial endpoints while the East[®] SurvAdapt package is required for studies with time-to-event endpoints. Since these packages do not function independently of East we will refer to the software as "East" rather than "EastAdapt" or "EastSurvAdapt" throughout this volume. The context will clarify which adaptive module must be available to the core East program in order to run a specific example.

54 The Cui, Hung and Wang Method

54.1 Statistical Method: Normal and Binomial

54.1.1 Hypothesis Testing

54.1.2 RCI's and RPV's

54.1.3 Conditional Power

54.1.4 East Defaults

In this section, we discuss hypothesis testing, confidence interval and p-value estimation, and computation of conditional power for a group sequential trial that permits adaptive sample size changes at the interim looks.

54.1.1 Hypothesis Testing

Consider a level- α test of the null hypothesis

$$H_0: \delta = 0$$

versus the two-sided alternative hypothesis

$$H_1: \delta \neq 0$$

for a two-arm randomized clinical trial. Before any data are obtained we pre-specify that this hypothesis will be tested by a group sequential trial designed for up to K looks, at cumulative sample sizes n_1, n_2, \dots, n_K , with corresponding level- α stopping boundaries b_1, b_2, \dots, b_K derived from some spending function. Although we have pre-specified the initial sample sizes look by look, there is full flexibility to adapt based on either external information, information from the trial itself, or a combination of the two. Accordingly let $n_1^*, n_2^*, \dots, n_K^*$ denote the altered cumulative sample sizes at the K looks after sample size adaptation. For adaptive designs it is convenient to express sample sizes, and parameter estimates that depend on sample size, in terms of incremental quantities as well as cumulative ones. Thus, for $j = 1, 2, \dots, K$ we define $n^{(j)} = n_j - n_{j-1}$ and $n^{*(j)} = n_j^* - n_{j-1}^*$ to be the incremental sample sizes for the pre-specified and altered designs, respectively, with $n_0 = n_0^* = 0$. In keeping with this notation, we will hereafter index all statistics computed from cumulative sample sizes with subscripts and all statistics computed from incremental sample sizes with superscripts. Additionally we will assign a superscript “*” to all statistics that are computed with the altered sample sizes $n_j^*, j = 1, 2, \dots, K$, rather than the pre-specified sample sizes $n_j, j = 1, 2, \dots, K$.

Suppose we are at look j . Denote the j incremental Wald statistics by

$$Z^{*(l)} = \frac{\hat{\delta}^{*(l)}}{\text{se}(\hat{\delta}^{*(l)})} = \hat{\delta}^{*(l)} \sqrt{I^{*(l)}}, \quad l = 1, 2, \dots, j, \quad (54.1)$$

where $\hat{\delta}^{*(l)}$ and $I^{*(l)} = [\text{se}(\hat{\delta}^{*(l)})]^{-2}$ are, respectively, the point estimate and Fisher information about δ based only on data from the incremental $n^{*(l)}$ observations obtained between look $(l - 1)$ and look l . The CHW statistic at look j , sometimes referred to as the *weighted* statistic, is constructed by combining these incremental Wald statistics with the *pre-specified* weights

$$w^{(l)} = \frac{n^{(l)}}{n_K} \quad l = 1, 2, \dots, j$$

as shown below:

$$Z_{j,\text{chw}}^* = \frac{\sqrt{w^{(1)}}Z^{*(1)} + \sqrt{w^{(2)}}Z^{*(2)} + \dots + \sqrt{w^{(j)}}Z^{*(j)}}{\sqrt{w^{(1)} + w^{(2)} + \dots + w^{(j)}}} . \quad (54.2)$$

This statistic is asymptotically normally distributed with mean

$$E(Z_{j,\text{chw}}^*) = \frac{\delta \sum_{l=1}^j \sqrt{w^{(l)} I^{*(l)}}}{\sqrt{\sum_{l=1}^j w^{(l)}}}$$

and unit variance. Interim monitoring proceeds just as it would in a conventional group sequential trial, and with the same stopping boundaries. The null hypothesis is rejected at the first look j which is such that $|Z_{j,\text{chw}}^*| \geq b_j$.

Both Cui, Hung and Wang (1999) and Lehman and Wassmer (1999) have shown that the CHW statistic preserves the type-1 error despite the data dependent changes in the sample sizes at the interim looks. That is,

$$P_0\left(\bigcup_{j=1}^K |Z_{j,\text{chw}}^*| \geq b_j\right) = \alpha .$$

Now consider the conventional Wald statistic

$$Z_{j,\text{wald}}^* = \frac{\hat{\delta}_j^*}{\text{se}(\hat{\delta}_j^*)} = \hat{\delta}_j \sqrt{I_j^*} , \quad (54.3)$$

where $\hat{\delta}_j^*$ and $I_j^* = [\text{se}(\hat{\delta}_j^*)]^{-2}$ are, respectively, the point estimate and Fisher information about δ based on data from all the n_j^* observation obtained up to and including look j . Because of the data dependent changes in sample size at each stage of the trial, the type-1 error may not be preserved; in general,

$$P_0\left(\bigcup_{j=1}^K |Z_{j,\text{wald}}^*| \geq b_j\right) \neq \alpha .$$

The conventional Wald statistic (54.3) is sometimes referred to as the *unweighted* statistic. This is really a misnomer because we can represent (54.3) at any look j as a weighted sum of j incremental Wald statistics (54.1) using weights

$$w^{*(l)} = \frac{n^{*(l)}}{n_K^*}, \quad l = 1, 2, \dots, j,$$

54 The Cui, Hung and Wang Method

that depend on the actual rather than the pre-specified sample sizes as shown below:

$$Z_{j,\text{wald}}^* = \frac{\sqrt{w^{*(1)}}Z^{*(1)} + \sqrt{w^{*(2)}}Z^{*(2)} + \dots + \sqrt{w^{*(j)}}Z^{*(j)}}{\sqrt{w^{*(1)} + w^{*(2)} + \dots + w^{*(j)}}}. \quad (54.4)$$

The statistics (54.3) and (54.4) are functionally equivalent for the normally distributed data with known variance. In all other settings the two statistics are asymptotically equivalent. It follows that if there is no sample size change one may use either the unweighted or weighted statistic for the interim monitoring since, in that case $w^{(j)} = w^{*(j)}$ for all j , and hence $Z_{j,\text{chw}}^* = Z_{j,\text{wald}}^*$ for all j , and

$$P_0\left(\bigcup_{j=1}^K |Z_{j,\text{chw}}^*| \geq b_j\right) = P_0\left(\bigcup_{j=1}^K |Z_{j,\text{wald}}^*| \geq b_j\right) = \alpha.$$

Although the above hypothesis testing procedure was described for two-sided tests with symmetric boundaries, the modifications to accommodate two-sided tests with asymmetric boundaries and one-sided tests with or without futility boundaries is straightforward.

54.1.2 Repeated Confidence Intervals and Repeated P-Values

The confidence intervals and p-values described in this section are generalizations of the repeated confidence intervals (RCI's) and repeated p-values (RVP's) discussed by Jennison and Turnbull (2000, Chapter 9) for classical group sequential designs. The extension to the adaptive setting is discussed in Lehmacher and Wassmer (1999) and more generally in Mehta, Bauer, Posch and Brannath (2007). All the RCI's and RPV's in this chapter utilize the method of Lehmacher and Wassmer (1999). Like the CHW method with which they are associated, these RCI's and RPV's are only valid for adaptive changes in the sample size. They are not applicable if additional adaptive changes are made to the initial design, such as data dependent changes to the number and spacing of the interim looks, or changes to the error spending function.

Lehmacher and Wassmer (1999) have shown that the K RCI's for δ are given by

$$\frac{(Z_{j,\text{chw}}^* \pm b_j)\sqrt{s_j}}{\sum_{l=1}^j \sqrt{w^{(l)}I^{*(l)}}}, \quad j = 1, 2, \dots, K \quad (54.5)$$

where $s_j = n_j/n_K$ is the information fraction at look j based on the pre-specified sample sizes, are repeated confidence intervals (RCI's). Thus, if δ_0 is the true value of

δ then, for all $j = 1, 2, \dots, K$,

$$P_{\delta_0} \left\{ \bigcap_{i=1}^j \left(\frac{(Z_{i,\text{chw}}^* - b_i)\sqrt{s_i}}{\sum_{l=1}^i \sqrt{w^{(l)}I^{*(l)}}} \leq \delta_0 \leq \frac{(Z_{i,\text{chw}}^* + b_i)\sqrt{s_i}}{\sum_{l=1}^i \sqrt{w^{(l)}I^{*(l)}}} \right) \right\}. \quad (54.6)$$

Following the development in Jennison and Turnbull (2000, page 202) we can use (54.6) to obtain a repeated p-value at any look j . This is accomplished by iteratively altering the significance level of the hypothesis test for δ until a level \tilde{p}_j , say, is obtained such that one of the two extremes of the corresponding RCI (54.5), with confidence coefficient $1 - p_j$, just excludes zero. To be specific, let $b_j(q)$, $j = 1, 2, \dots, K$ represent any level- q two-sided stopping boundaries derived from some spending function. That is,

$$P_0 \left(\bigcup_{j=1}^K |Z_{j,\text{chw}}^*| \geq b_j(q) \right) = q$$

Let $z_{j,\text{chw}}^*$ be the observed value of $Z_{j,\text{chw}}^*$ at look j . Then the two-sided repeated p-value at look j is the probability \tilde{p}_j that satisfies the relationship

$$z_{j,\text{chw}}^* - b_j(\tilde{p}_j) = 0 \quad \text{if } \hat{\delta}_j \geq 0 \quad (54.7)$$

$$z_{j,\text{chw}}^* + b_j(\tilde{p}_j) = 0 \quad \text{if } \hat{\delta}_j < 0. \quad (54.8)$$

These results can be readily modified to accommodate two-sided asymmetric tests and one-sided tests with or without futility boundaries. Suppose, for example that we have obtained the asymmetric two-sided boundaries (a_j, b_j) , $j = 1, 2, \dots, K$ such that

$$P_0 \left\{ \bigcup_{i=1}^{j-1} (a_i < Z_{i,\text{chw}}^* < b_i) \cap (Z_{j,\text{chw}}^* \leq a_j), j = 1, 2, \dots, K \right\} = \alpha_l$$

and

$$P_0 \left\{ \bigcup_{i=1}^{j-1} (a_i < Z_{i,\text{chw}}^* < b_i) \cap (Z_{j,\text{chw}}^* \geq b_j), j = 1, 2, \dots, K \right\} = \alpha_u$$

where $\alpha_l + \alpha_u = \alpha$. Then the K repeated confidence intervals for δ are given by

$$\left[\frac{(Z_{j,\text{chw}}^* - b_j)\sqrt{s_j}}{\sum_{l=1}^j \sqrt{w^{(l)}I^{*(l)}}}, \frac{(Z_{j,\text{chw}}^* - a_j)\sqrt{s_j}}{\sum_{l=1}^j \sqrt{w^{(l)}I^{*(l)}}} \right] \quad j = 1, 2, \dots, K$$

To compute the two-sided repeated p-values let q be any probability and let $(a_j(q), b_j(q))$, $j = 1, 2, \dots, K$, be any level- q asymmetric two-sided stopping

54 The Cui, Hung and Wang Method

boundaries derived from a pair of level- q asymmetric spending functions $(\alpha_l(q), \alpha_u(q))$ which have the same functional form as the spending functions used in the original asymmetric level- α trial design. That is,

$$P_0 \left\{ \bigcup_{i=1}^{j-1} (a_i(q) < Z_{i,\text{chw}}^* < b_i(q)) \cap (Z_{j,\text{chw}}^* \leq a_j(q)), j = 1, 2, \dots, K \right\} = \alpha_l(q)$$

and

$$P_0 \left\{ \bigcup_{i=1}^{j-1} (a_i(q) < Z_{i,\text{chw}}^* < b_i(q)) \cap (Z_{j,\text{chw}}^* \geq b_j(q)), j = 1, 2, \dots, K \right\} = \alpha_u(q)$$

Note: In this notation, if $q = \alpha$, then $\alpha_l(q) = \alpha_l$, $\alpha_u(q) = \alpha_u$, $\alpha_l(q) + \alpha_u(q) = \alpha_l + \alpha_u = \alpha$, and $(a_j(q), b_j(q)) = (a_j, b_j)$, $j = 1, 2, \dots, K$.

Let $z_{j,\text{chw}}^*$ be the observed value of $Z_{j,\text{chw}}^*$ at look j . The two-sided repeated p-value at look j is the probability \tilde{p}_j that satisfies the relationship

$$z_{j,\text{chw}}^* - a_j(\tilde{p}_j) = 0 \quad \text{if } \hat{\delta}_j \geq 0 \tag{54.9}$$

$$z_{j,\text{chw}}^* + b_j(\tilde{p}_j) = 0 \quad \text{if } \hat{\delta}_j < 0. \tag{54.10}$$

54.1.3 Conditional Power

Suppose that an on-going trial has reached some interim look $L < K$, and the observed value of the CHW test statistic is $Z_{L,\text{chw}}^* = z_L$. Having examined the data so far obtained, suppose it is planned to proceed through the remaining stages of the trial with cumulative sample sizes $n_{L+1}^*, n_{L+2}^*, \dots, n_K^*$ that are possibly different than the cumulative sample sizes $n_{L+1}, n_{L+2}, \dots, n_K$ pre-specified at the start of the trial. We define the conditional power at look L as the probability of attaining statistical significance **in the direction of the alternative hypothesis** at any future look, given $z_{(L)}$. Thus, if we are testing the null hypothesis that $\delta = 0$ against the alternative that $\delta > 0$, the conditional power is defined as

$$\text{CP}_\delta(z_L) = P_\delta \left\{ \bigcup_{j=L+1}^K (Z_{j,\text{chw}}^* \geq b_j | z_L) \right\} \tag{54.11}$$

whereas if the alternative hypothesis is that $\delta < 0$, then the conditional power is defined as

$$\text{CP}_\delta(z_L) = P_\delta \left\{ \bigcup_{j=L+1}^K (Z_{j,\text{chw}}^* \leq b_j | z_L) \right\}. \tag{54.12}$$

For two sided tests the conditional power is given by

$$CP_{\delta}(z_L) = P_{\delta}\left\{ \bigcup_{j=L+1}^K (|Z_{j,\text{chw}}^*| \geq b_j | z_L) \right\} \quad (54.13)$$

These probabilities are obtained by recursive integration in East. Special East calculators are available from within the CHW interim monitoring dashboard and the CHW adaptive simulations to compute the conditional power for any specified value of δ . Use of these calculators will be demonstrated in the worked examples that form part of this chapter as well as in a separate chapter of the current user manual.

We conclude this section with some additional remarks about conditional power:

- Although equations (54.11) through (54.13) are expressed in terms of δ , their dependence on σ for normally distributed data is implicit through the expression (54.2) for $Z_{j,\text{chw}}^*$. In fact for the normal case one can show that conditional power depends only on the ratio δ/σ .
- By increasing the sample size of the remainder of the trial after look L one increases the conditional power. The calculators in East can be used to determine the magnitude of the sample size increase that is needed to achieve any desirable conditional power, for any assumed value of δ (and σ).
- Each simulation performed in the CHW adaptive simulations implements a one-time adaptive increase in sample size at a specified look L of the K -look group sequential design. The magnitude of the sample size increase is determined by a pre-specified conditional power, say $1 - \beta$, that the user desires to achieve. In order to speed up the simulations, this sample size is computed by an approximation to equation (54.11) (or (54.12)) that assumes that the next time the data are monitored will be at look K , and all the intermediate looks $L + 1, L + 2, \dots, K - 1$ will be skipped. Specifically the approximate conditional power calculation is given by

$$CP_{\delta}(z_L) = 1 - \Phi \left\{ b_K \sqrt{1 + \frac{n_L}{n_K - n_L}} - z_L \sqrt{\frac{n_L}{n_K - n_L}} - \frac{\delta \sqrt{r(1-r)} \sqrt{n_K^* - n_L}}{\sigma} \right\} \quad (54.14)$$

where r is the fraction randomized to the experimental arm. The approximate sample size needed to achieve conditional power $1 - \beta$ is then obtained by finding the value of n_K^* that satisfies

$$1 - \Phi \left\{ b_K \sqrt{1 + \frac{n_L}{n_K - n_L}} - z_L \sqrt{\frac{n_L}{n_K - n_L}} - \frac{\delta \sqrt{r(1-r)} \sqrt{n_K^* - n_L}}{\sigma} \right\} = 1 - \beta. \quad (54.15)$$

The operating characteristics of the adaptive design under this approximate way of computing conditional power are almost the same as the operating

54 The Cui, Hung and Wang Method

characteristics that would be obtained by using say equation (54.11) to evaluate conditional power at each simulation. The simulations are, however, speeded up substantially thereby.

54.1.4 Adaptive Simulations Defaults

It will be useful to discuss now, the default settings used for adaptive simulation procedures in East, and the options available to you if you want to change them. If you click on the button **Include Options** and check **Sample Size Re-estimation**, you will see the following additional tab called **Sample Size Re-estimation** on the screen.

| | | | |
|--|---------------------------|--|------------|
| Use Adaptation Method | | | |
| <input checked="" type="radio"/> CHW | <input type="radio"/> CDL | <input type="radio"/> Müller and Schäfer | Weights... |
| Adapt at: | Look # | 1 | |
| Max. Sample Size if Adapt (multiplier; total #): | | 1 | 694 |
| Target CP for Re-estimating Sample Size: | | 0.8 | |
| Promising Zone Scale: | | Cond. Power | CP |
| Promising Zone: | Min. CP: | 0.3 | |
| | Max. CP: | 0.8 | |
| CP Computation Based on: | | Estimated δ/σ | |

All the parameters on this tab will explained in detail in the subsequent sections. Here, we explain the default settings on this tab.

Three adaptation methods are implemented in this version of East - Cui-Hung-Wang, Chen-DeMets-Lan, and Müller and Schäfer methods. The default method selected is Cui-Hung-Wang.

By default, the adaptation happens at a specified look number. One can also perform adaptation after a specified sample size or information fraction.

By default, the promising zone is defined on the Conditional Power scale. One can also define it on the Test Statistic scale or Estimated δ scale.

The default settings for CP Computations are:

Estimated δ/σ for Normal Endpoint

Estimated (π_c, π_t) for Binomial Endpoint

Estimated HR for Survival Endpoint

We recommend that you do not alter these settings except for research purposes. The choices you make in this dialog box will determine how the adaptive simulations are conducted. Because changes to these settings can substantially alter the operating characteristics of the adaptive simulations, East will revert back to the default values each time the East session is terminated. For Binomial Endpoint designs, there are only two choices: **Estimated** (π_c, π_t) and **Design** (π_c, π_t). Depending on the selection you make, the conditional power computation at the specified interim look (Or sample size Or information fraction) will be based either on the estimated value of π_c and π_t or the values that have been used for creating the study design. An example of a binomial endpoint design is shown below.

| | |
|--|-------------------|
| | Wbk1:bin |
| Mnemonic | PN-25-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 3 |
| Test Type | 1-Sided |
| Variance | Unpooled Estimate |
| Specified α | 0.025 |
| Power | 0.901 |
| Model Parameters | |
| Proportion under Control (π_c) | 0.25 |
| Proportion under Treatment (π_t) | 0.4 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0.15 |
| Allocation Ratio (n_t/n_c) | 1 |
| Boundary Parameters | |
| Efficacy Boundary | LD (OF) |
| Spacing of Looks | Equal |
| Sample Size | |
| Maximum | 405 |
| Expected Under H0 | 404.169 |
| Expected Under H1 | 324.601 |

The design parameters in the above plan are $\pi_c = 0.25$ and $\pi_t = 0.40$. If you have chosen the setting **Design** (π_c, π_t) for conditional power computation in the **Sample Size Re-estimation** tab, then in any adaptive simulations, the conditional power at the interim analysis will be computed using the design values, $\pi_c = 0.25$ and

54 *The Cui, Hung and Wang Method*

$\pi_t = 0.40$, regardless of the estimated values obtained for these parameters at the time of the interim analysis.

For Normal Endpoint designs, the conditional power depends on δ and σ only through the ratio δ/σ . You may choose either **Design** δ/σ or **Estimated** δ/σ in the **Sample Size Re-estimation** tab and the conditional power at the interim analysis in any adaptive simulation will be computed accordingly. The sigma used for the computing test statistic is determined by the choice of sigma made in the **Simulation Parameters** tab. If this choice is Z then **design** σ is used otherwise **Estimated** σ is used in the test statistic computation.

Survival Endpoint designs are discussed in Section 54.2. For such designs the treatment effect δ is defined to be the log hazard ratio. You may choose either the Design HR or Estimated HR for purposes of computing conditional power at the interim look.

We end this section with an example that shows how the choice of Design or Estimated parameter values for conditional power computation at an interim look can alter the operating characteristics of an adaptive design. Consider the normal endpoint design

shown below.

| | Wbk1:Des1 |
|---|------------------|
| Mnemonic | MN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.901 |
| Model Parameters | |
| Input Method | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 15 |
| Mean Control (μ_c) | 0 |
| Mean Treatment (μ_t) | 15 |
| Std. Deviation (σ) | 30 |
| Test Statistic | Z |
| Allocation Ratio (nt/nc) | 1 |
| Boundary Parameters | |
| Efficacy Boundary | LD (OF) |
| Spacing of Looks | Equal |
| Sample Size | |
| Maximum | 171 |
| Expected Under H0 | 170.649 |
| Expected Under H1 | 136.959 |

The design parameters for the above plan are $\delta = 15$ and $\sigma = 30$. Suppose you have chosen **Design** δ/σ for the conditional power computation on the **Sample Size**

54 The Cui, Hung and Wang Method

Re-estimation tab:

| | | | |
|--|---------------------------|--|-----|
| Use Adaptation Method | | Weights... | |
| <input checked="" type="radio"/> CHW | <input type="radio"/> CDL | <input type="radio"/> Müller and Schäfer | |
| Adapt at: | Look # | 2 | |
| Max. Sample Size if Adapt (multiplier; total #): | | 1 | 171 |
| Target CP for Re-estimating Sample Size: | | 0.9 | |
| Promising Zone Scale: | | Cond. Power | CP |
| Promising Zone: | Min. CP: | 0.3 | |
| | Max. CP: | 0.9 | |
| CP Computation Based on: | | Design δ/σ | |

and the test statistic as Z on **Simulation Parameters** tab:

| | |
|------------------------------|----|
| Test Statistic: | Z |
| Std. Deviation (σ): | 30 |

Then, in every adaptive simulation you carry out for this design, the conditional power at the interim analyses will be based on the ratio $\delta/\sigma = 15/30 = 0.5$ and the test statistic will be computed under the assumption that design $\sigma = 30$, rather than estimating these quantities from the actual data generated at the interim look.

In order to explore the impact of changes to these simulation settings, change the choice of the test statistic to t in **Simulation Parameters** tab as shown below:

| | |
|-----------------|-------|
| Test Statistic: | t |
| Variance: | Equal |

Change the value of δ_1 under **Response Generation Info** to 10 and keep the σ_1 value same as 30.

| Simulation Parameters | Response Generation Info | Sample Size Re-estimation |
|---------------------------------------|--------------------------|---|
| Generate Data Using: Individual Means | | <input checked="" type="checkbox"/> Common Standard Deviation |
| Mean Control (μ_c): | 0 | SD Control (σ_c): 30 |
| Mean Treatment (μ_t): | 10 | SD Treatment (σ_t): 30 |

Also make changes on the **Sample Size Re-estimation** tab as shown below.

| Use Adaptation Method | | Weights... | |
|--|---------------------------|--|-----------------------------------|
| <input checked="" type="radio"/> CHW | <input type="radio"/> CDL | <input type="radio"/> Müller and Schäfer | |
| Adapt at: | Look # | 2 | |
| Max. Sample Size if Adapt (multiplier; total #): | 2 | 342 | |
| Target CP for Re-estimating Sample Size: | 0.9 | | |
| Promising Zone Scale: | | Cond. Power | <input type="button" value="CP"/> |
| Promising Zone: | Min. CP: | 0.3 | |
| | Max. CP: | 0.9 | |
| CP Computation Based on: | | Estimated δ/σ | |

and set the simulation control parameters as shown below:

54 The Cui, Hung and Wang Method

| Number of Simulations: <input type="text" value="100000"/> | Output Options | | | | | | |
|--|---|----------------|---|----|----|----|----|
| Refresh Frequency: <input type="text" value="10000"/> | Output Type: <input type="text" value="Case Data"/> | | | | | | |
| Random Number Seed | <input type="checkbox"/> Save summary statistics for every simulation run | | | | | | |
| <input checked="" type="radio"/> Clock | <input type="checkbox"/> Save subject level data for <input type="text" value="1"/> simulation runs | | | | | | |
| <input type="radio"/> Fixed <input type="text" value="100"/> | Note: Max. 100,000 records will be saved. | | | | | | |
| <input type="checkbox"/> Suppress All Intermediate Output | Output for All Trials | | | | | | |
| <input type="checkbox"/> Pause after Refresh | <table border="1"><thead><tr><th>Percentile (%)</th></tr></thead><tbody><tr><td>5</td></tr><tr><td>25</td></tr><tr><td>50</td></tr><tr><td>75</td></tr><tr><td>95</td></tr></tbody></table> | Percentile (%) | 5 | 25 | 50 | 75 | 95 |
| Percentile (%) | | | | | | | |
| 5 | | | | | | | |
| 25 | | | | | | | |
| 50 | | | | | | | |
| 75 | | | | | | | |
| 95 | | | | | | | |
| <input checked="" type="checkbox"/> Stop At End | | | | | | | |

Click on the button **Simulate** to run the simulations. An entry will be added in the **Output Preview** pane. Save it in the **Library** and click the  icon to see the output summary of these simulations.

| Wbk1:Des1:CHWSim1 | |
|---|---------------------------|
| Mnemonic | MN-25-DI |
| Test Parameters | |
| Design Type | Superiority |
| Test Type | 1-Sided |
| Power | 0.6594 |
| Power (Promising) | 0.8724 |
| No. of Looks | 3 |
| Model Parameters | |
| Test Statistic | Z |
| Mean Control Data (μ c) | 0 |
| Mean Treatment Data (μ t) | 10 |
| Std. Deviation (σ) | 30 |
| Boundary Parameters | |
| Spacing of Looks | User Specified |
| Efficacy Boundary | User Specified |
| Sample Size | |
| Maximum | 171 |
| Sample Size Re-estimation Parameters | |
| Method of Adaptation | Cui-Hung-Wang |
| Adaptation Stage | Look # : 2 |
| Max. Sample Size if Adapt | 342 |
| Target CP | 0.9 |
| CP Computation Based on | Estimated δ/σ |
| Promising Zone | 0.3 <= CP < 0.9 |
| Simulation Results (Overall) | |
| Average Sample Size | 191.238 |
| Simulation Results (Promising) | |
| Average Sample Size | 273.782 |

Remember that these results were obtained when the values of δ/σ and σ were both set to **Estimated** values. Now you can change your settings for these two parameters to **Estimated** and **Design** respectively. It can be done by editing the current simulation node. Select the simulation node in the **Library** and click the  icon. you will be taken **Simulation Parameters** tab. Here, select the Test statistic as Z from its dropdown. Run the simulation and get new results. Similarly you can carry out other two simulations with the combinations **Design-Estimated** and **Design-Design** for the two parameters δ/σ and σ and obtain the results. So now you will have four sets of results for the four assumptions on the two parameters. These results are all obtained by simulating under the values of $\delta_1 = 10$ and $\sigma_1 = 30$. Carry out two more sets of similar analyses using the combinations of $\delta_1 = 8$ & $\sigma_1 = 40$ and $\delta_1 = 12$ & $\sigma_1 = 24$. You may compare the various resulting values from

54 The Cui, Hung and Wang Method

the simulations, like power, average combined sample size, average adapted sample size, etc. As an example, let us compare the values of power, from the different simulations carried out, as tabulated below.

Table 54.1: Results for different assumptions of East Settings for Adaptive Simulation

(Design parameters: $\delta = 15, \sigma = 30$)

| Settings for | | Estimates of Power under different simulation parameters | | | |
|-------------------|----------------|--|------------------------------------|-------------------------------------|------------------------------------|
| Conditional Power | Test Statistic | $\delta_1 = 10,$ $\sigma_1 = 30$ | $\delta_1 = 8,$ $\sigma_1 = 40$ | $\delta_1 = 12,$ $\sigma_1 = 24$ | $\delta_1 = 0,$ $\sigma_1 = 30$ |
| δ/σ | σ | | | | |
| Estimated | Estimated | 66.25% | 30.8% | 93.24% | 2.65% |
| Estimated | Design | 65.93% | 48.96% | 85.81% | 2.5% |
| Design | Estimated | 67.12% | 30.66% | 94.42% | 2.5% |
| Design | Design | 67.27% | 49.21% | 87.78% | 2.44% |

We can also compare multiple simulation scenarios in East itself. It can be by selecting from the **Library** the scenarios to be compared and clicking the  icon to see the comparison. Let us compare first four scenarios from the above table.

| | Wbk1:Des1:CHWSim1 | Wbk1:Des1:CHWSim2 | Wbk1:Des1:CHWSim3 | Wbk1:Des1:CHWSim4 |
|---|----------------------------------|----------------------------------|-------------------------------|-------------------------------|
| Mnemonic | MN-2S-DI | MN-2S-DI | MN-2S-DI | MN-2S-DI |
| Test Parameters | | | | |
| Design Type | Superiority | Superiority | Superiority | Superiority |
| Test Type | 1-Sided | 1-Sided | 1-Sided | 1-Sided |
| Power | 0.663 | 0.659 | 0.671 | 0.673 |
| Power (Promising) | 0.877 | 0.875 | 0.642 | 0.645 |
| No. of Looks | 3 | 3 | 3 | 3 |
| Model Parameters | | | | |
| Test Statistic | t | Z | t | Z |
| Mean Control Data (μc) | 0 | 0 | 0 | 0 |
| Mean Treatment Data (μt) | 10 | 10 | 10 | 10 |
| Std. Deviation (σ) | | 30 | | 30 |
| Boundary Parameters | | | | |
| Spacing of Looks | User Specified | User Specified | User Specified | User Specified |
| Efficacy Boundary | User Specified | User Specified | User Specified | User Specified |
| Sample Size | | | | |
| Maximum | 171 | 171 | 171 | 171 |
| Sample Size Re-estimation Parameters | | | | |
| Method of Adaptation | Cui-Hung-Wang | Cui-Hung-Wang | Cui-Hung-Wang | Cui-Hung-Wang |
| Adaptation Stage | Look # : 2 | Look # : 2 | Look # : 2 | Look # : 2 |
| Max. Sample Size if Adapt | 342 | 342 | 342 | 342 |
| Target CP | 0.9 | 0.9 | 0.9 | 0.9 |
| CP Computation Based on Promising Zone | Estimated δ/σ 0.3 <= CP < 0.9 | Estimated δ/σ 0.3 <= CP < 0.9 | Design δ/σ 0.3 <= CP < 0.9 | Design δ/σ 0.3 <= CP < 0.9 |
| Simulation Results (Promising) | | | | |
| Average Sample Size | 275.034 | 274.842 | 236.222 | 235.744 |

By comparing the power estimates across rows or columns in the above table, you will be able to gauge the effect due to different parameters/computations assumptions. Our recommendation is to leave these parameters at their default values except for exploratory purposes.

54.2 Statistical Method: Survival

For studies involving survival (time-to-event) endpoints the parameter δ denoting the treatment effect is defined to be the logarithm of the hazard ratio of the treatment arm to the control arm; $\delta = \ln(\text{HR})$. Under proportional hazards, $\delta < 0$ implies longer survival times for the treatment arm than for the control arm. In order to test the null hypothesis

$$H_0: \delta = 0$$

versus one and two-sided alternatives we exploit the independent increment structure of the sequentially computed logrank statistic (Tsiatis, 1981; Jennison and Turnbull, 1997). Before any data are obtained we pre-specify that the null hypothesis will be tested by a K -look group sequential design at potential stopping times D_1, D_2, \dots, D_K , where D_j denotes the **cumulative** number of events obtained at look j and b_j is the corresponding level- α stopping boundary derived from some spending function. The CHW method permits data dependent alterations to the cumulative events at which these looks occur. Accordingly let $D_1^*, D_2^*, \dots, D_K^*$ denote the altered cumulative events at the K looks resulting from an adaptation of the original design. Analogous to the notation developed for normal and binomial endpoints let $D^{(j)} = D_j - D_{j-1}$ and $D^{*(j)} = D_j^* - D_{j-1}^*$ be the incremental increase in the number of events between looks $j - 1$ and j .

54 The Cui, Hung and Wang Method

Let $Z_{j,cum}^*$ denote the Z-score based by either a logrank statistic or the treatment effect estimate obtained by fitting the Cox proportional hazard model to the cumulative data available at look j . Then the results of Tsiatis(1981) and Jennison and Turnbull (1997) show that the incremental statistics

$$Z^{*(j)} = \frac{\sqrt{I_j^*} Z_{j,cum}^* - \sqrt{I_{j-1}^*} Z_{j-1,cum}^*}{\sqrt{I_j^* - I_{j-1}^*}}, \text{ for } j = 1, \dots, K, \quad (54.16)$$

are asymptotically independent and normally distributed with mean

$$E \left[Z^{*(j)} \right] \approx \delta \sqrt{I_j^* - I_{j-1}^*} \quad (54.17)$$

and unit variance, where r is the fraction randomized to the treatment arm. In the simulation module $Z_{j,cum}^*$ comes from the log-rank test and we assume that

$$I_j^* \approx r(1-r)D_j^* \quad (54.18)$$

Here r is the proportion of subjects in the active treatment group. In the interim monitoring module we use an approximation (54.18) as default and use a slightly different approximation

$$I_j^* = \frac{1}{\left(\widehat{se}(\hat{\delta}_j) \right)^2} \quad (54.19)$$

if the monitoring relies on the estimates of treatment effect $\hat{\delta}_j$ and $\widehat{se}(\hat{\delta}_j)$ provided by fitting the Cox proportional hazard model to the cumulative data at look j . Let

$$w^{(j)} = \frac{D^{(j)}}{D^{(K)}}, \text{ for } j = 1, 2, \dots, K,$$

be pre-specified weights. Following Wassmer (2006), the CHW statistic for survival designs is constructed by combining the independent incremental statistics (54.16) with these weights:

$$Z_{j,chw}^* = \frac{\sqrt{w^{(1)}} Z^{*(1)} + \sqrt{w^{(2)}} Z^{*(2)} + \dots + \sqrt{w^{(j)}} Z^{*(j)}}{\sqrt{w^{(1)} + w^{(2)} + \dots + w^{(j)}}}. \quad (54.20)$$

The CHW statistic (54.20) for survival endpoints has the same asymptotic distribution as the CHW statistic (54.2) for normal and binomial endpoints. Thus all the distributional results, repeated confidence intervals, p-values, and conditional power calculations derived in Section 54.1.1, Section 54.1.2, and Section 54.1.3 for normal and binomial endpoints also hold for time-to-event endpoints with $\delta = \ln(\text{HR})$, $\sigma = 1$, $D^{*(j)}$ substituting for $n^{*(j)}$.

In particular equation (54.14), depicting the conditional power if $Z_{L,\text{chw}}^* = z_L$ at look L and D_K^* cumulative events are required at the K th look, can be re-expressed in the form

$$\text{CP}_\delta(z_L) = 1 - \Phi \left\{ b_K \sqrt{1 + \frac{D_L}{D_K - D_L}} - z_L \sqrt{\frac{D_L}{D_K - D_L}} - \delta \sqrt{r(1-r)} \sqrt{D_K^* - D_L} \right\}. \tag{54.21}$$

Since the true value of δ is unknown it is customary to substitute either its look L estimate $\hat{\delta}_L$,

$$\hat{\delta}_L^* = \frac{Z_{L,\text{cum}}^*}{\sqrt{r(1-r)} D_L^*} \tag{54.22}$$

or else the value δ_1 specified under the alternative hypothesis at the design stage, in the above expression for conditional power. East provides the user with both options. Note that like equation (54.14), equation (54.21) also involves the simplifying assumption that the next look following look L will be the last look, and all intermediate looks $L + 1, L + 2, \dots, K - 1$ will be skipped. This assumption yields an approximate conditional power that can be computed rapidly and is sufficiently accurate for use in simulation experiments such as those discussed in Section 54.5.3. However, special calculators documented in Chapter 57 are available if a more accurate conditional power computation that respects the actual stopping boundaries at looks $L + 1, L + 2, \dots, K - 1$ is desired.

54.3 Normal Endpoint: Schizophrenia Trial

54.3.1 Fixed Sample Design

54.3.2 Adaptive Design

54.3.3 Interim Monitoring

Consider a two-arm trial to determine if there is an efficacy gain for an experimental drug relative to the industry standard treatment for negative symptoms schizophrenia. The primary endpoint is the improvement from baseline to week 26 in the Negative Symptoms Assessment (NSA), a 16-item clinician-rated instrument for measuring the negative symptomatology of schizophrenia. Let μ_t denote the difference between the mean NSA at baseline and the mean NSA at week 26 for the treatment arm and let μ_c denote the corresponding difference of means for the control arm. Denote the efficacy gain by $\delta = \mu_t - \mu_c$. The trial will be designed to test the null hypothesis $H_0: \delta = 0$ versus the one-sided alternative hypothesis that $\delta > 0$. It is expected, from limited data on related studies, that $\delta \geq 2$ and σ , the between-subject standard deviation, is believed to be about 7.5. In the discussion that follows, we shall focus our attention on

54 The Cui, Hung and Wang Method

the uncertainty about δ . Even though the statistical methods discussed here are applicable when there is uncertainty about either δ or σ , the adaptive approach requires careful justification primarily when δ is involved. Adaptive sample size adjustments relating to uncertainty about σ are fairly routine and non-controversial. One way to eliminate the uncertainty due to σ is to re-parameterize the treatment effect in terms of δ/σ , since it turns out that the sample size, power and conditional power are all dependent on δ and σ only through this ratio. Although we shall not follow that approach here, we wish to point out that it is supported by the EastAdapt software.

This example is discussed in detail in Chapter 53, Section 53.2, where the relative merits of the fixed sample, group sequential and adaptive designs are compared. We have re-introduced this example in the present chapter in order to illustrate how to use the adaptive features in East software to design, simulate and monitor an adaptive clinical trial that will test the null hypothesis $\delta = 0$ and estimate the parameter δ .

54.3.1 Fixed Sample Design

Since it is believed a priori that $\delta \geq 2$, we first create Des 1, a single-look design with 80% power to detect $\delta = 2$ using a one-sided level 0.025 test, given $\sigma = 7.5$.

| | CHW:Des 1 |
|---|---------------------|
| Mnemonic | MN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.8 |
| Model Parameters | |
| Input Method | Difference of Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 2 |
| Std. Deviation (σ) | 7.5 |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | Z |
| Allocation Ratio (nt/nc) | 1 |
| Sample Size | |
| Maximum | 442 |

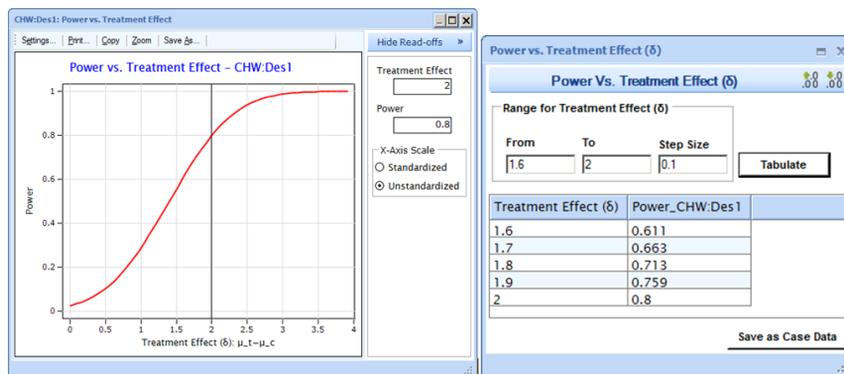
Des 1 shows that if the assumptions about δ and σ are correct, the trial will achieve 80% power with a total sample size of 442 subjects. There is, however, considerable

uncertainty about the true value of δ , and to a lesser extent about σ . Nevertheless it is believed that even if the true value of δ were as low as 1.6 on the NSA scale, that would constitute a clinically meaningful effect. Des 2, displayed below, shows that if 690 subjects are enrolled the power to detect $\delta = 1.6$ is 80%.

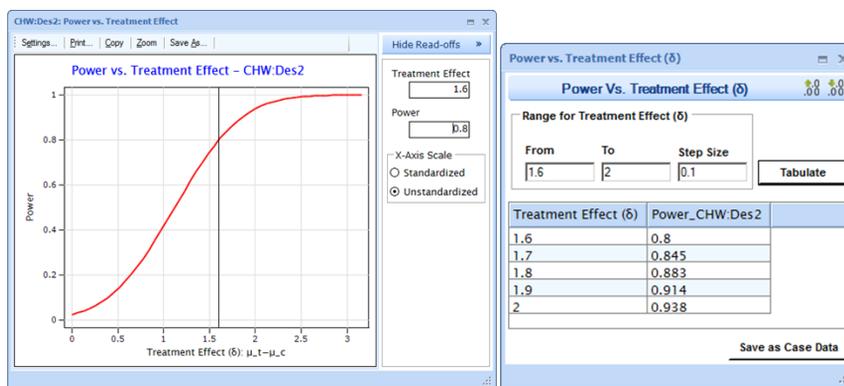
| | CHW:Des1 | CHW:Des2 |
|---|---------------------|---------------------|
| Mnemonic | MN-2S-DI | MN-2S-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.8 | 0.8 |
| Model Parameters | | |
| Input Method | Difference of Means | Difference of Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 2 | 1.6 |
| Std. Deviation (σ) | 7.5 | 7.5 |
| Test Statistic | Z | Z |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Sample Size | | |
| Maximum | 442 | 690 |

So far we have proposed two design options. Under Des 1 we would enroll 442 subjects and hope that the study is adequately powered, which it will be if $\delta = 2$ and $\sigma = 7.5$. If, however $\delta = 1.6$ the power drops from 80% to 61%.

54 The Cui, Hung and Wang Method



There is thus a risk of launching an underpowered study for an effective drug under Des 1, even if $\sigma = 7.5$. Under Des 2 we will enroll 690 subjects, thereby ensuring 80% power at the smallest clinically meaningful value, $\delta = 1.6$, and rising to 94% power at $\delta = 2$.



The operating characteristics of Des 1 and Des 2 are displayed side by side in Table 54.2.

If resources were plentiful, Des 2 would clearly be the preferred option. The sponsor must, however, allocate scarce resources over a number of studies and in any case is not in favor of designing an overpowered trial. This leads naturally to considering a design that might be more flexible with respect to sample size than either of the above

Table 54.2: Operating Characteristics of Des 1 and Des 2

| δ | Des 1 | | Des 2 | |
|----------|-------------|-------|-------------|-------|
| | Sample Size | Power | Sample Size | Power |
| 1.6 | 442 | 61% | 690 | 80% |
| 1.7 | 442 | 66% | 690 | 85% |
| 1.8 | 442 | 71% | 690 | 88% |
| 1.9 | 442 | 76% | 690 | 91% |
| 2.0 | 442 | 80% | 690 | 94% |

two single-look fixed sample designs. Two options for providing this greater flexibility are the group sequential design and the adaptive design. In the group sequential design one starts out with a large up-front commitment by powering the study to detect the smallest clinically meaningful treatment effect $\delta = 1.6$, but the expected sample size is reduced by means of early stopping boundaries. In the adaptive design, one starts out with a smaller initial sample size by powering the study to detect the optimistic treatment effect $\delta = 2$, but reserves the option to increase the sample size on the basis of the data obtained at an interim look, should it appear advantageous to do so.

Group sequential designs are discussed extensively elsewhere in the East manual and hence this option need not be illustrated in the current chapter. We refer the user to Chapter 53, Section 53.2 of this user manual for a thorough discussion of the relative merits of the group sequential and adaptive options as they relate to the present example. It is seen that the relatively long follow-up (26 weeks) before the primary endpoint is observed leads to patient overruns which offset some of the advantages of the group sequential design. We shall accordingly confine our discussion to adaptive design for the remainder of this section.

54.3.2 Adaptive Design

To motivate the adaptive design let us recall that although the actual value of δ is unknown, the investigators believe that $\delta \geq 2$. For this reason Des 1 was constructed to have 80% power to detect $\delta = 2$. Des 2 on the other hand was constructed to have 80% power to detect $\delta = 1.6$, the smallest clinically meaningful treatment effect. If there were no resource constraints one would of course prefer to design the study for 80% power at $\delta = 1.6$ since that would imply even more power at $\delta = 2$. However, as can be seen from Table 54.2, this conservative strategy carries as its price a substantially larger up-front sample size commitment which is, moreover, unnecessary if in truth $\delta = 2$.

54 The Cui, Hung and Wang Method

The above difficulties lead us to consider whether Des 1, which was intended to detect $\delta = 2$ with 80% power and hence does not have such a large up-front sample size commitment, might be improved so as to provide some insurance against substantial power loss in the event that $\delta = 1.6$. The adaptive approach is suited to this purpose. In this approach we start out with a sample size of 442 subjects as in Des 1, but take an interim look after data are available on 208 completers. The purpose of the interim look is not to stop the trial early but rather to examine the interim data and continue enrolling past the planned 442 subjects if the interim results are promising enough to warrant the additional investment of sample size. This strategy has the advantage that the sample size is finalized only after a thorough examination of data from the actual study rather than through making a large up-front sample size commitment before any data are available. Furthermore, if the sample size may only be increased but never decreased from the originally planned 442 subjects, there is no loss of efficiency due to overruns. For the final analysis we adopt the CHW statistic described in Section 54.1, so as to avoid inflating the type-1 error.

Selecting the Criteria for an Adaptive Sample Size Increase

The operating characteristics of an adaptive design depend in a complicated way on the criteria for increasing the sample size after observing the interim data. These criteria may combine objective information such as the current estimate of δ or the current conditional power with assessments of safety and with information available from other clinical trials that was not available at the start of the study. The adaptive approach provides complete flexibility to modify the sample size without having to pre-specify a precise mathematical formula for computing the new sample size based on the interim data. Therefore the full benefit of the flexibility offered by an adaptive design cannot be quantified ahead of time. Nevertheless it is instructive to investigate power and expected sample size by simulating the trial under different values of δ and applying precise pre-specified rules for increasing the sample size on the basis of the observed interim results. This will provide at least some idea, at the design stage, of the trade-off between the fixed sample or group sequential approaches and the adaptive approach.

To this end we create Des 3 as a 2-look design with 80% power to detect $\delta = 2$ with a one-sided level-0.025 test, and one interim analysis utilizing the $\gamma(-24)$ spending function after data are available on 208 completers. The $\gamma(-24)$ early stopping boundary selected for Des 3 is so conservative that for all practical purposes there is no early stopping at all. The specification of this early stopping boundary is simply an artificial device for permitting an interim look at which one may adaptively increase

the sample size. Therefore Des 3 may be viewed as an extension of Des 1.

Design: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means

| Test Parameters | |
|--------------------------------|------------------|
| Design ID | Des3 |
| Design Type | Superiority |
| Number of Looks | 2 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.8 |
| Model Parameters | |
| Test Statistic | Z |
| Input Method | Individual Means |
| Mean Control (μ_c) | 0 |
| Mean Treatment (μ_t) | 2 |
| $\delta = \mu_t - \mu_c$ | |
| Under H0 | 0 |
| Under H1 | 2 |
| Std. Deviation (σ) | 7.5 |
| Allocation Ratio (n_t/n_c) | 1 |
| Boundary Parameters | |
| Spacing of Looks | Unequal |
| Efficacy Boundary | Gm (-24) |

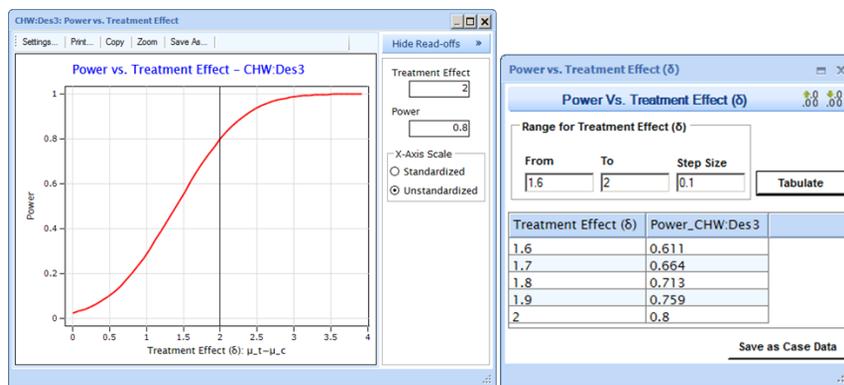
Sample Size Information

| | Control Arm | Treatment Arm | Total |
|--------------------------------|-------------|---------------|---------|
| Sample Size (n) | | | |
| Maximum | 221 | 221 | 442 |
| Expected H1 | 220.949 | 220.949 | 441.898 |
| Expected H0 | 221 | 221 | 442 |
| Maximum Information (I): 1.964 | | | |

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries Efficacy Z | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|-----------------------|---|----------|
| | | | | | Under H0 | Under H1 |
| | | | | | Efficacy | Efficacy |
| 1 | 0.471 | 208 | 7.583E-8 | 5.251 | 7.583E-8 | 4.38E-4 |
| 2 | 1 | 442 | 0.025 | 1.96 | 0.025 | 0.8 |

At the start of the trial, both plans have the same sample size of 442 subjects and 80% power at $\delta = 2$, deteriorating to 61% power at $\delta = 1.6$.

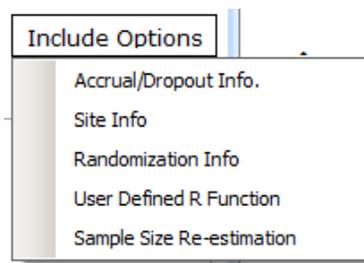


Des 3 stipulates, however, that an interim look will be taken after 26 weeks of follow-up data are available on 208 of the planned 442 subjects. At that interim look the sample size may be increased. The timing of the interim look reflects a preference for performing the interim analysis as late as possible but nevertheless while the trial is still enrolling subjects since, once the enrollment sites have closed down, it will be difficult to start them up again. Under the assumption that subjects enroll at the rate of 8 per week we will have enrolled 416 subjects by week 52; 208 of them will have

54 The Cui, Hung and Wang Method

completed the required 26 weeks of follow-up for the primary endpoint, and an additional 208 subjects will comprise the overruns. Only the data from the 208 completers will be used in making the decision to increase the sample size. After this decision is taken, enrollment will continue until the desired sample size is attained. The primary efficacy analysis will be based on the full 26 weeks of follow-up data from all enrolled subjects and will utilize the CHW test, thereby ensuring that the type-1 error is preserved despite the data dependent sample size change at the interim look. It should be noted that, unlike the group sequential setting where the 208 overruns at the time of the interim look played no role in the early stopping decision, here the data from the 208 overruns will be fully utilized in the primary efficacy analysis which will only occur when all enrolled subjects have completed 26 weeks of follow-up. This is one of the advantages of the adaptive approach relative to the group sequential approach for trials with lengthy follow-up.

The East software provides a simulation tool for studying the consequences of increasing the sample size of Des 3 at the interim look. To implement this tool we must add the sample size re-estimation tab for Des 3. Select Des 3 in the **Library** and click the **S** icon. In addition to the default tabs appearing by default on inserting Simulations, one can add more tabs to enter information available on randomization, stratification and sample size re-estimation. This can be done by clicking the **Include Options** button on this right hand top corner of the screen.



Select Sample Size Re-estimation from the list. This will add a tab named as **Sample Size Re-estimation** as shown below:



Let us focus on these tabs. Several parameters on these four tabs shown below play important role in simulation and adaptation of a design. The three tabs **Simulation Parameters**, **Response Generation Info** and **Simulation Control Info** contain all the

information about the design Des 3 in the absence of any adaptive change. It is a two-look design with a sample size of 442 and an interim look after 208 completers. The early stopping boundary generated by the $\gamma(-24)$ spending function equals 5.251 standard deviations on the Wald statistic scale. With this extremely conservative boundary there is practically no chance of early stopping even at the alternative hypothesis that $\delta = 2$. This design is for all practical purposes the same as Des 1.

The fourth tab **Sample Size Re-estimation** is used to specify the rules of adaptation for modifying the initial sample size of Des 3, based on the data at the interim analysis. Before running the simulations we must input suitable values into the cells of this tab. We have made the following choices in different tabs:

The **Response Generation Info** tab:

| | |
|---------------------------------------|---|
| Generate Data Using: Individual Means | <input checked="" type="checkbox"/> Common Standard Deviation |
| Mean Control (μ_c): 0 | SD Control (σ_c): 7.5 |
| Mean Treatment (μ_t): 1.6 | SD Treatment (σ_t): 7.5 |

The **Sample Size Re-estimation** tab:

| | | |
|--|---------------------------|--|
| Use Adaptation Method | | Weights... |
| <input checked="" type="radio"/> CHW | <input type="radio"/> CDL | <input type="radio"/> Müller and Schäfer |
| Adapt at: | Look # | 1 |
| Max. Sample Size if Adapt (multiplier; total #): | 2 | 884 |
| Target CP for Re-estimating Sample Size: | 0.8 | |
| Promising Zone Scale: | Cond. Power | <input checked="" type="radio"/> CP |
| Promising Zone: | Min. CP: | 0.3 |
| | Max. CP: | 0.8 |
| CP Computation Based on: | Estimated δ/σ | |

Most of these simulation parameters are self-explanatory. Some of them need further explanation. This is provided below.

Adapt at: For a K -look group sequential design, one can decide the time at which

54 The Cui, Hung and Wang Method

conditions for adaptations are to be checked and actual adaptation is to be carried out. This can be done either at some intermediate look or after some specified information fraction. The possible values of this parameter depends upon the choice of the user. If it is **Look no.**, then this parameter can be any integer number from 1 to $K - 1$. If the adaptation is to be carried out after reaching specified information fraction then this parameter can be a fraction between 0 and 1. The default choice in East is **Look no.** to decide the time of adaptation.

Adapt at:

Target CP for Re-estimating Sample Size: The primary driver for increasing the sample size at the interim look is the desired (or target) conditional power or probability of obtaining a positive outcome at the end of the trial, given the data already observed. For this example we have set the conditional power at the end of the trial to be 80%. East then computes the sample size that would be required to achieve this desired conditional power. The computation assumes that the estimated $\hat{\delta}$ obtained at the interim look is the true δ . Refer to Section 54.1.3 for the relevant formula for this computation.

Maximum Sample Size if Adapt (multiplier; total): As just stated, a new sample size is computed at the interim analysis on the basis of the observed data so as to achieve some target conditional power. However the sample size so obtained will be overruled unless it falls between pre-specified minimum and maximum values. For this example, the range of allowable sample sizes is [442, 884]. If the newly computed sample size falls outside this range, it will be reset to the appropriate boundary of the range. For example, if the sample size needed to achieve the desired 80% conditional power is less than 442, the new sample size will be reset to 442. In other words we will not decrease the sample size from what was specified initially. On the other hand, the upper bound of 884 subjects demonstrates that the sponsor is prepared to increase the sample size up to double the initial investment in order to achieve the desired 80% conditional power. But if 80% conditional power requires more than 884 subjects, the sample size will be reset to 884, the maximum allowed.

Promising Zone Scale: One can define the promising zone as an interval based on conditional power or test statistic or δ/σ . The input fields change according to this choice. The decision of altering the sample size is taken based on whether the interim value of conditional power / test statistic / δ/σ lies in this interval or not.

Let us keep the default scale which is Conditional Power.

Promising Zone: Minimum/Maximum Conditional Power (CP): The sample size

will only be altered if the estimate of CP at the interim analysis lies in a pre-specified range, referred to as the "Promising Zone". Here the promising zone is stipulated to be $0.30 - 0.80$. The idea is to invest in the trial in stages. Prior to the interim analysis the sponsor is only committed to a sample size of 442 subjects. If, however, the results at the interim analysis appear reasonably promising, the sponsor would be willing to make a larger investment in the trial and thereby improve the chances of success. Here we have somewhat arbitrarily set the lower bound for a promising interim outcome to be $CP = 0.30$. An estimate $CP < 0.30$ at the interim analysis is not considered promising enough to warrant a sample size increase. It might sometimes be desirable to also specify an upper bound beyond which no sample size change will be made. Here we have set that upper bound of the promising zone at $CP = 0.80$. In effect we have partitioned the range of possible values for conditional power at the interim analysis into three zones; *unfavorable* ($CP \leq 0.3$), *promising* ($0.3 \leq CP < 0.8$), and *favorable* ($CP \geq 0.8$). Sample size adaptations are attempted only if CP (with no sample size adaptation) falls in the promising zone at the interim analysis.

The promising zone defined on the Test Statistic scale or δ/σ scale work on the similar lines.

The **Simulation Control Info** tab:

| | |
|--|-------------------------------------|
| Number of Simulations: | <input type="text" value="100000"/> |
| Refresh Frequency: | <input type="text" value="10000"/> |
| Random Number Seed | |
| <input checked="" type="radio"/> Clock | |
| <input type="radio"/> Fixed | <input type="text" value="100"/> |

| |
|---|
| Output Options |
| Output Type: <input type="text" value="Case Data"/> |
| <input type="checkbox"/> Save summary statistics for every simulation run |
| <input type="checkbox"/> Save subject-level data for <input type="text" value="1"/> simulation runs |
| Note: Max. 100,000 records will be saved. |

Operating Characteristics of Adaptive Implementation of Des 3

Having entered the above simulation parameters into the simulation tabs, we simulate the adaptive implementation of Des 3 100,000 times. An entry gets added in the **Output Preview** pane. Save this Simulation node in the workbook and either double click on the node or click the  icon to see the details for the complete simulation

54 The Cui, Hung and Wang Method

output.

Simulation: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means (CHW Simulation)

| Simulation Parameters | |
|--------------------------------------|---------------------------|
| Simulation ID | CHWSim1 |
| Design Type | Superiority |
| Number of Looks | 2 |
| Test Type | 1-Sided |
| Sample Size (n) | 442 |
| Variance | Equal |
| Test Statistic | t |
| Avg. Power at Termination | 0.671 |
| Response Generation Parameters | |
| Generate Data Using | Individual Means |
| Mean Control (μ_c) | 0 |
| Mean Treatment (μ_t) | 1.6 |
| SD Control (σ_c) | 7.5 |
| SD Treatment (σ_t) | 7.5 |
| Sample Size Re-estimation Parameters | |
| Method of Adaptation | Cui-Hung-Wang |
| Adapt At Look No. | 1 |
| Max. Sample Size if Adapt | |
| Multiplier | 2 |
| Total # | 884 |
| Target CP | 0.8 |
| Promising Zone Scale | Cond. Power |
| Min. CP | 0.3 |
| Max. CP | 0.8 |
| CP Computation Based on | Estimated δ/σ |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 100000 |

Zone-wise Averages

| | Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|---|-------------|--------------------------|----------|------------------------------|---------|-------------------|----------|---------------------|
| | | Count | Row % | Count | Row % | Count | Column % | |
| ⊖ | Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| | Unfavorable | 9155 | 28.155% | 23362 | 71.845% | 32517 | 32.517% | 442 |
| | Promising | 21986 | 83.004% | 4502 | 16.996% | 26488 | 26.488% | 697.409 |
| ⊖ | Favorable | 35889 | 87.592% | 5084 | 12.408% | 40973 | 40.973% | 442 |
| | Efficacy | 22 | 100.000% | 0 | 0.000% | 22 | 0.022% | 208 |
| | All Trials | 67052 | 67.052% | 32948 | 32.948% | 100000 | 100.000% | 509.601 |

Promising Zone defined as $0.3 \leq CP < 0.8$

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 208 |
| 2 | 509.668 |
| Average | 509.601 |

Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | Efficacy | Count | % |
| | | Upper | | | |
| 1 | 208 | 5.251 | 22 | 22 | 0.022% |
| 2 | 442 | 1.96 | 67030 | 99978 | 99.978% |
| Total | | | 67052 | 100000 | |
| % | | | 67.052% | | |

The results from the 100,000 simulated trials are displayed in three tables titled **Simulation Boundaries and Boundary Crossing Probabilities**, **Average Sample Size and Look Times** and **Simulation Results by Zone**. We observe from the table that the power of the adaptive implementation of Des 3 at $\delta = 1.6$ is 67.05%, an improvement of about 6% over the power of Des 1 at the same value of δ . This increase in power has come at an average cost of $510 - 442 = 68$ additional subjects. Next we observe from the **Simulation Results by Zone** that 26,488 of the 100,000 trials (26.49%) underwent a sample size adaptation and of these 26,488 trials, 21,986 (83%) were able to reject the null hypothesis. The average sample size, conditional on adaptation was 697.41. To examine these same results in more details, we see the table **Zone-wise Averages**.

This table contains the results from all the six zones - Futility, Unfavorable, Promising, Favorable, Efficacy and All Trials.

The simulations fall in the unfavorable zone, promising, favorable and efficacy zones

32.52%, 26.49%, 40.97% and 0.022% of the time respectively. Observe that while the overall probability of obtaining a significant result is only 67.05%, this probability jumps up to 83.0% conditional on falling in the promising zone.

We repeat these simulations with other values of δ between 1.6 and 2. The operating characteristics for the adaptive Des 3, are compared to those of the fixed sample Des 1 in Table 54.3. All results for Des 3 are based on 100,000 simulated trials and rounded to the nearest percentage point.

Table 54.3: Operating Characteristics of Des 1 (Fixed Sample) and Des 3 (Adaptive)

| Value of δ | Des 1(Fixed Sample) | | Des 3- Sim 1 to Sim 5(Adaptive) | |
|-------------------|---------------------|---------------------|---------------------------------|----------------------|
| | Power | Expected SampleSize | Power | Expected Sample Size |
| 1.6 | 61% | 442 | 67% | 509 |
| 1.7 | 66% | 442 | 72% | 508 |
| 1.8 | 71% | 442 | 77% | 506 |
| 1.9 | 76% | 442 | 81% | 502 |
| 2.0 | 80% | 442 | 84% | 499 |

The power of the adaptive Des 3 has increased by about 6% at $\delta = 1.6$ and by about 4% at $\delta = 2$ compared to Des 1. These power gains were obtained at the cost of corresponding average sample size increases of 67 subjects at $\delta = 1.6$ and 57 subjects at $\delta = 2$. Although these power gains appear fairly modest, Des 3 offers a significant benefit in terms of risk reduction, not reflected in Table 54.3. To see this, it is important to note that the sample size under Des 3 is only increased when the interim results are promising; i.e., when the conditional power at the interim analysis is greater than or equal to 30% but less than 80%. This is the very situation in which it is advantageous to increase the sample size and thereby avoid an underpowered trial. When the interim results are unfavorable (conditional power $< 30\%$) or favorable (conditional power $\geq 80\%$), a sample size increase is not warranted and hence it is unchanged at 442 subjects for both Des 1 and Des 3. But when the interim results are promising (conditional power between 30% and 80%) the sample size is increased under Des 3 in an attempt to boost the conditional power back to 80%. It is this feature of the adaptive design that makes it more attractive than the simpler fixed sample design.

In order to compare Des 1(the fixed sample design) with Des 3 (the group sequential design designed with adaptive simulations) conditional on zone, let us edit the simulations inputs associated with Des 3. Select Simulation node in the **Library** and click the  icon and make the changes as below:

54 The Cui, Hung and Wang Method

The **Response Generation Info** tab:

| | | |
|-----------------------------|------------------|---|
| Generate Data Using: | Individual Means | <input checked="" type="checkbox"/> Common Standard Deviation |
| Mean Control (μ_c): | 0 | SD Control (σ_c): 7.5 |
| Mean Treatment (μ_t): | 1.6:2.0:0.1 | SD Treatment (σ_t): 7.5 |

and the **Sample Size Re-estimation** tab:

| | | | |
|--|---------------------------|--|--|
| Use Adaptation Method | | Weights... | |
| <input checked="" type="radio"/> CHW | <input type="radio"/> CDL | <input type="radio"/> Müller and Schäfer | |
| Adapt at: | Look # | 1 | |
| Max. Sample Size if Adapt (multiplier; total #): | 1 | 442 | |
| Target CP for Re-estimating Sample Size: | 0.8 | | |
| Promising Zone Scale: | Cond. Power | CP | |
| Promising Zone: | Min. CP: | 0.3 | |
| | Max. CP: | 0.8 | |
| CP Computation Based on: | Estimated δ/σ | | |

Here we are simulating the same design except that we will not make any modification to the sample size at the interim look. Des 3 stipulates, however, that an interim look will be taken after 26 weeks of follow-up data are available on 208 of the planned 442 subjects. At that interim look the sample size may be increased. But in this modified setup, we will not increase the sample size at the interim look.

Note that we have kept the cap on max. sample size after adaptation as 442 under modified setup, compared to 884 under Sim 1. Now we can run the adaptive simulation under this modified setup and make a comparison of the results with the results obtained under Sim 1. Table 54.4 displays the probability of falling into the unfavorable, promising and favorable zones at the interim look, along with the power and expected sample size, conditional on falling into each zone, under various values of δ .

The table highlights the key advantage of the adaptive design (Sim 1 to Sim 5) compared to the traditional group sequential (Sim 6 to Sim 10) i.e., the ability to invest

Table 54.4: Operating Characteristics of Traditional Group Sequential Trial and an Adaptive Group Sequential Trial Conditional on Interim Outcome

| δ | Interim Outcome | Probability of Interim Outcome | Power Conditional on Interim Outcome | | Expected Sample Size | |
|---|-----------------|--------------------------------|--------------------------------------|-------|----------------------|-------|
| | | | Des 1 | Des 3 | Des 1 | Des 3 |
| 1.6 | Unfav + Fut | 32% | 28% | 28% | 442 | 442 |
| | Promising | 26% | 62% | 83% | 442 | 697 |
| | Fav + Eff | 41% | 87% | 87% | 442 | 442 |
| 1.7 | Unfav + Fut | 29% | 32% | 32% | 442 | 442 |
| | Promising | 26% | 65% | 86% | 442 | 694 |
| | Fav + Eff | 45% | 89% | 89% | 442 | 442 |
| 1.8 | Unfav + Fut | 26% | 36% | 37% | 442 | 442 |
| | Promising | 26% | 69% | 88% | 442 | 692 |
| | Fav + Eff | 48% | 91% | 91% | 442 | 442 |
| 1.9 | Unfav + Fut | 23% | 40% | 41% | 442 | 442 |
| | Promising | 25% | 73% | 91% | 442 | 687 |
| | Fav + Eff | 52% | 93% | 93% | 442 | 442 |
| 2.0 | Unfav + Fut | 20% | 45% | 45% | 442 | 442 |
| | Promising | 24% | 76% | 93% | 442 | 684 |
| | Fav + Eff | 56% | 95% | 94% | 442 | 442 |
| All results are based on 100,000 simulated trials | | | | | | |

in the trial in stages, with the second stage of the investment being required only if promising results are obtained at the first stage. This feature of adaptive design makes it far more attractive as an investment strategy than fixed sample or non-adaptive group sequential design which has no provision for increasing the sample size if a promising interim outcome is obtained. Suppose, for example that $\delta = 1.6$, the smallest clinically meaningful treatment effect. The trial sponsor only commits the resources needed for 442 subjects at the start of the trial, at which point the chance of success is 61%, as shown in Table 54.3. The additional sample size commitment is forthcoming only if promising results are obtained at the interim analysis, and in that case the sponsor's risk is substantially reduced because the chance of success jumps to 83%, as shown in Table 54.4. Similar results are observed for the other values of δ .

The probabilities of entering the unfavorable, promising and favorable zones at the interim analysis, displayed in Table 54.4, are instructive. Consider again the case

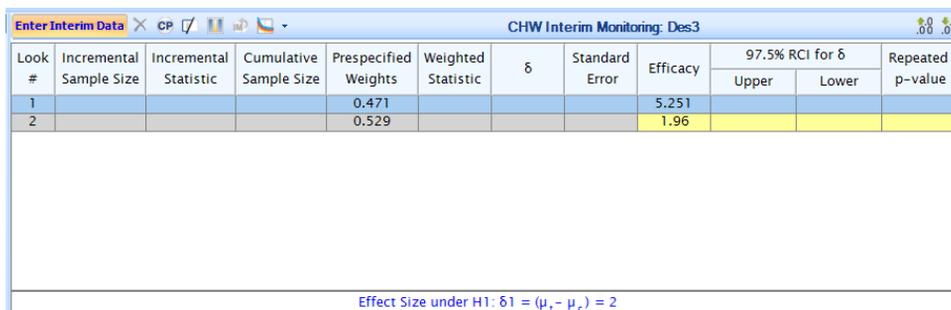
54 The Cui, Hung and Wang Method

$\delta = 1.6$. At this value of δ there is a 26% chance of landing in the promising zone and thereby obtaining a substantial power boost under adaptive setup as compared to non-adaptive. That is, 26% of the time the adaptive strategy can rescue a trial that is underpowered at the interim look. The chance of entering the favorable + efficacy zone is 41%. That is, 41% of the time the sponsor will be lucky and have a well powered trial at the interim look without the need to increase the sample size. The remaining 32% of the time the sponsor will be unlucky and will enter the unfavorable zone from which also there is no sample size increase, and the chance of success is only 28%. These odds improve with larger values of δ . The adaptive implementation satisfies the objective of powering the study primarily for $\delta = 2$ while providing a hedge against substantial power loss if $1.6 \leq \delta < 2$. It is thus a good compromise between Des 1 which is powered to detect $\delta = 2$ without any means of improving power if $\delta = 1.6$, and Des 2 which is powered to detect $\delta = 1.6$ but utilizes excessive sample size resources if $\delta = 2$.

54.3.3 Interim Monitoring

Now we will discuss the interim monitoring procedure taking the example of Des 3. Accordingly we invoke the CHW IM dashboard associated with Des 3 by clicking on the  icon from the toolbar.

The following dashboard appears.



| Look # | Incremental Sample Size | Incremental Statistic | Cumulative Sample Size | Prespecified Weights | Weighted Statistic | δ | Standard Error | Efficacy | 97.5% RCI for δ | | Repeated p-value |
|--------|-------------------------|-----------------------|------------------------|----------------------|--------------------|----------|----------------|----------|------------------------|-------|------------------|
| | | | | | | | | | Upper | Lower | |
| 1 | | | | 0.471 | | | | 5.251 | | | |
| 2 | | | | 0.529 | | | | 1.96 | | | |

Effect Size under H1: $\delta_1 = (\mu_1 - \mu_c) = 2$

This dashboard differs from the usual interim monitoring dashboard for a classical group sequential trial in the following major ways: The Pre-specified Nominal Critical Points (stopping boundaries) are written into the dashboard as soon as it is invoked, and are non-editable. Patient accruals and corresponding test statistics are entered incrementally for each look, rather than cumulatively for all looks taken thus far. The weighted statistic is obtained by combining these incremental test statistics using Pre-specified Weights that are written into the dashboard as soon as it is invoked. One

is free to change the incremental sample size at each look from what was originally specified at the design stage. But if the sample sizes that correspond to the original study design are entered, then the weighted statistic is the same as the usual Wald statistic used for conventional (non-adaptive) interim monitoring

Suppose the first look is taken as planned after enrolling 208 subjects. Suppose we observe $\hat{\delta} = 1.7$ and $\hat{\sigma} = 7.6$ thus leading to a standard error of $\sqrt{(4 * 7.6^2 / 208)} = 1.0539$. The incremental statistic at the first look is thus $(1.7 / 1.0539) = 1.613$. Invoke the Test Statistic Calculator by clicking on the **Enter Interim Data** button. We enter these quantities into the Test Statistic Calculator as shown below.

| Editing Look # 1 For Incremental Sample Size | |
|--|--------|
| Sample Size: | 208 |
| Input for Normal end point | |
| Estimate of δ : $\delta = (\mu_t - \mu_c)$ | 1.7 |
| Standard Error of Estimate of δ : | 1.0539 |
| Output | |
| Weighted Statistic: | 1.613 |
| Incremental Statistic: | 1.613 |

Buttons: Recalc, OK, Cancel

Since the nominal critical value for early stopping is 5.251, the trial continues. We now need to decide on the sample size to use for the second and final look. We invoke the

54 The Cui, Hung and Wang Method

conditional power calculator to assist with this decision.

The screenshot shows a dialog box titled "Conditional Power Calculator" with a close button (X) in the top right corner. The dialog is divided into two main sections: "Input" and "Input/Output".

Input Section:

- Look #:
- Cumulative Sample Size:
- Weighted Test Statistic:

Input/Output Section:

- Value of δ :
- Value of σ :
- Value of δ/σ :
- Computed Conditional Power:
- Sample Size (Overall):

*Use radio button to select the quantity to be computed.

At the bottom of the dialog, there are four buttons: "Recalc", "Plot", "Details...", and "Close".

Suppose we specify to the calculator that we wish to obtain 80% conditional power to detect $\delta=1.6$ with a hypothesized value of 7.5 for sigma. Upon entering these terms

into the calculator we obtain a final (overall) sample size of 564.7 subjects.

Conditional Power Calculator

Input

Look #:

Cumulative Sample Size:

Weighted Test Statistic:

Input/Output

Value of δ :

Value of σ :

Value of δ/σ :

Desired Conditional Power:

Computed Sample Size (Overall):

*Use radio button to select the quantity to be computed.

Based on the guidance provided by the calculator, suppose we decide to enroll a total of 565 subjects. This implies that the incremental number to be entered into the interim monitoring dashboard is $565 - 208 = 357$ subjects.

Suppose that, based only on these 357 incremental subjects, the estimate of delta is 1.5 and the estimate of sigma is 7.7. The standard error of $\hat{\delta}$ is thus

54 The Cui, Hung and Wang Method

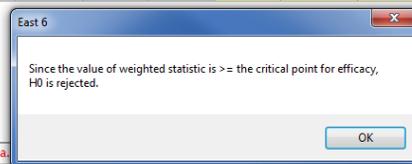
$\sqrt{(4 * 7.7^2 / 357)} = 0.8151$, leading to an incremental test statistic of 1.8404.

| Field | Value |
|--|--------|
| Sample Size: | 357 |
| Estimate of δ : $\delta = (\mu_t - \mu_c)$ | 1.5 |
| Standard Error of Estimate of δ : | 0.8151 |
| Weighted Statistic: | 2.446 |
| Incremental Statistic: | 1.84 |

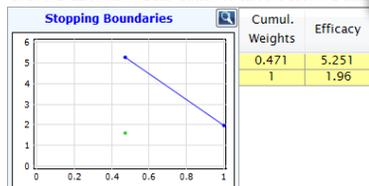
Upon pressing the OK button the incremental test statistic is entered into the interim monitoring dashboard and the weighted statistic that combines the two incremental statistics by the square roots of the pre-specified weights (as described in Section 54.1, equation 54.4) is computed as 2.446. Since the weighted statistic exceeds the nominal critical value, the null hypothesis is rejected. The confidence interval for delta is (0.3146, infty) and the p-value is 0.0072. These estimates are based on the methods described in Section 54.1 and are appropriately adjusted to preserve their validity in

the face of adaptive sample size changes.

| Look # | Incremental Sample Size | Incremental Statistic | Cumulative Sample Size | Prespecified Weights | Weighted Statistic | δ | Standard Error | Efficacy | 97.5% RCI for δ | | Repeat... p-value |
|--------|-------------------------|-----------------------|------------------------|----------------------|--------------------|----------|----------------|----------|------------------------|--------|-------------------|
| | | | | | | | | | Upper | Lower | |
| 1 | 208 | 1.613 | 208 | 0.471 | 1.613 | 1.7 | 1.054 | 5.251 | Infinity | -3.834 | 1 |
| 2 | 357 | 1.84 | 565 | 0.529 | 2.446 | 1.5 | 0.815 | 1.96 | Infinity | 0.315 | 0.007 |



Click the "Edit Interim Data" button to edit the Look # 2 data.



54.4 Binomial Endpoint: Acute Coronary Syndromes

54.4.1 Fixed Sample Design

54.4.2 Group Sequential Design

54.4.3 Adaptive Group Sequential Design

54.4.4 Operating Characteristics

54.4.5 Adding a Futility Boundary

Consider a two-arm, placebo controlled randomized clinical trial for subjects with acute cardiovascular disease undergoing percutaneous coronary intervention (PCI). The primary endpoint is a composite of death, myocardial infarction or ischemia-driven revascularization during the first 48 hours after randomization. We assume on the basis of prior knowledge that the event rate for the placebo arm is 8.7%. The investigational drug is expected to reduce the event rate by at least 20%. The investigators are planning to randomize a total of 8000 subjects in equal proportions to the two arms of the study.

54.4.1 Fixed Sample Design

We show with the help of East that a conventional fixed sample design enrolling a total of 8000 subjects will have 83% power to detect a 20% risk reduction with a one-sided level-0.025 test of significance (with 0.087 on the control arm and

54 The Cui, Hung and Wang Method

$0.8 \times 0.087 = 0696$ on the treatment arm).

| | Des 1 |
|--|-------------------|
| Mnemonic | PN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Variance | Unpooled Estimate |
| Specified α | 0.025 |
| Power | 0.8259 |
| Model Parameters | |
| Proportion under Control (π_c) | 0.087 |
| Proportion under Treatment (π_t) | 0.07 |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.017 |
| Allocation Ratio (n_t/n_c) | 1 |
| Sample Size | |
| Maximum | 8000 |

The actual risk reduction is expected to be larger, but could also be as low as 15%, a treatment effect that would still be of clinical interest given the severity and importance of the outcomes. In addition, there is some uncertainty about the magnitude of the placebo event rate. For these reasons the investigators wish to build into the trial design some flexibility for adjusting the sample size. Two options under consideration are, a group sequential design with the possibility of early stopping in case the risk reduction is large, and an adaptive design with the possibility of increasing the sample size in case the risk reduction is small. In the remainder of this section we shall discuss these two options and show how they may be combined into a single design that captures the benefits of both.

54.4.2 Group Sequential Design

We first transform the fixed sample design into an 8000 person group sequential design with two interim looks, one after 4000 subjects are enrolled (50% of total information) and the second after 5600 subjects are enrolled (70% of total information). Early stopping efficacy boundaries are derived from the Lan and DeMets (1983) O'Brien-Fleming type error spending function. This group sequential design is shown as Des 2 in the following screen shot. Along with this plan, its operating characteristics

are also shown by the side.

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Test Parameters | |
|-----------------------------------|-------------------|
| Design ID | Des5 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.82 |
| Model Parameters | |
| Prop. under Control (π_c) | 0.087 |
| Prop. under Treatment (π_t) | 0.07 |
| $\delta = \pi_t - \pi_c$ | |
| Under H0 | 0 |
| Under H1 | -0.017 |
| Allocation Ratio (n_t/n_c) | 1 |
| Variance | Unpooled Estimate |
| Boundary Parameters | |
| Spacing of Looks | Unequal |
| Efficacy Boundary | LD (OF) |

⊖ Sample Size Information

| | Control Arm | Treatment Arm | Total |
|------------------------------------|-------------|---------------|----------|
| Sample Size (n) | | | |
| Maximum | 4000 | 4000 | 8000 |
| Expected H1 | 3267.517 | 3267.517 | 6535.034 |
| Expected H0 | 3989.918 | 3989.918 | 7979.837 |
| Maximum Information (I): 27741.783 | | | |

⊖ Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|------------|---|----------|
| | | | | Efficacy Z | Under H0 | Under H1 |
| | | | | | Efficacy | Efficacy |
| 1 | 0.5 | 4000 | 0.002 | -2.963 | 0.002 | 0.181 |
| 2 | 0.7 | 5600 | 0.007 | -2.462 | 0.006 | 0.31 |
| 3 | 1 | 8000 | 0.025 | -2.002 | 0.018 | 0.33 |

The output tells us that for this design, where the risk reduction is 20%; the probabilities of crossing boundary at Look1 (N=4000) is 0.181, at Look2 (N=5600) 0.31, and at Final Look 0.33; the overall power is 82%.

We can also create different designs by changing the value of risk reduction in Des 2 and obtain their corresponding results. A summary of such results is displayed in Table 54.5. The first column of Table 54.5 is a list of potential risk reductions, defined as $100 \times (1 - \rho)\%$ where $\rho = \pi_t/\pi_c$, π_t is the event rate for the treatment arm, and π_c is the event rate for the control arm. The remaining columns display early stopping probabilities, power and expected sample size. Since the endpoint is observed with 48 hours, the problem of overruns that we encountered in the schizophrenia trial is negligible and may be ignored.

54 The Cui, Hung and Wang Method

Table 54.5: Operating Characteristics of Des 2, a Three-Look 8000-Person Group Sequential Design

| Risk Reduction $100 \times (1 - \rho)$ | Probability of Crossing Efficacy Boundary | | | Overall Power | Expected Sample Size |
|---|---|-----------------------------|---------------------------------|---------------|----------------------|
| | At Look 1 ($N = 4000$) | At Look 2 ($N = 5600$) | At Final Look ($N = 8000$) | | |
| 15% | 0.074 | 0.183 | 0.309 | 57% | 7264 |
| 17% | 0.109 | 0.235 | 0.335 | 68% | 7002 |
| 20% | 0.181 | 0.310 | 0.330 | 82% | 6535 |
| 23% | 0.279 | 0.362 | 0.275 | 92% | 6017 |
| 25% | 0.357 | 0.376 | 0.222 | 95% | 5671 |

Table 54.5 shows that Des 2 is well powered, with large savings of expected sample size for risk reductions of 20% or more. It is thus a satisfactory design if, as is initially believed, the magnitude of the risk reduction is in the range 20% to 25%. This design does not, however, offer as good protection against a false negative conclusion for smaller risk reductions. In particular, even though 15% is still a clinically meaningful risk reduction, Des 2 offers only 57% power to detect this treatment effect. One possibility then is to increase the up-front sample size commitment of the group sequential design so that it has 80% power if the risk reduction is 15%. This leads to Des 3, a three-look group sequential design with a maximum sample size commitment of 13,853 subjects, one interim look after 6927 subjects (50% of total information) and a second interim look after 9697 subjects (70% of total information). Des 3 has 80% power to detect a risk reduction of 15% with a one-sided level-0.025 test.

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

| Test Parameters | |
|--------------------------------------|-------------------|
| Design ID | Des3 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.8 |
| Model Parameters | |
| Prop. under Control (π_c) | 0.087 |
| Prop. under Treatment (π_{t1}) | 0.074 |
| $\delta = \pi_c - \pi_{t1}$ | |
| Under H0 | 0 |
| Under H1 | -0.013 |
| Allocation Ratio (n_c/n_t) | 1 |
| Variance | Unpooled Estimate |
| Boundary Parameters | |
| Spacing of Looks | Unequal |
| Efficacy Boundary | LD (OF) |

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|------------------------------------|-------------|---------------|-----------|
| Sample Size (n) | | | |
| Maximum | 6927 | 6926 | 13853 |
| Expected H1 | 5728.406 | 5727.406 | 11455.811 |
| Expected H0 | 6909.542 | 6908.542 | 13818.085 |
| Maximum Information (I): 46828.394 | | | |

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries Efficacy Z | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|--------------------------|---|----------|
| | | | | | Under H0 | Under H1 |
| | | | | | Efficacy | Efficacy |
| 1 | 0.5 | 6927 | 0.002 | -2.962 | 0.002 | 0.167 |
| 2 | 0.7 | 9697 | 0.007 | -2.462 | 0.006 | 0.298 |
| 3 | 1 | 13853 | 0.025 | -2.002 | 0.018 | 0.335 |

Table 54.6 displays operating characteristics of Des 3 for risk reductions between 15%, and 25%, while keeping the maximum sample size as 13,853 . Notice that by attempting to provide adequate power at 15% risk reduction, the low end of clinically meaningful treatment effects, we have significantly over-powered the trial for values of risk reduction in the expected range of risk reductions, 20% to 25% . If, as expected, the risk reduction exceeds 20%, the large up-front sample size commitment of 13,853 subjects under Des 3 is unnecessary. Des 2 with an up-front commitment of only 8000 subjects will provide sufficient power in this setting. From this point of view, Des 3 is

Table 54.6: Operating Characteristics of Des 3, a Three-Look 13,853-Person Grp Sequential Design

| Risk Reduction $100 \times (1 - \rho)$ | Probability of Crossing Efficacy Boundary | | | Overall Power | Expected Sample Size |
|---|---|-----------------------------|------------------------------------|---------------|----------------------|
| | At Look 1 ($N = 6926$) | At Look 2 ($N = 9697$) | At Final Look ($N = 13, 853$) | | |
| 15% | 0.167 | 0.298 | 0.335 | 80% | 11,456 |
| 17% | 0.246 | 0.349 | 0.296 | 89% | 10,699 |
| 20% | 0.395 | 0.375 | 0.196 | 97% | 9558 |
| 23% | 0.565 | 0.329 | 0.099 | 99.3% | 8574 |
| 25% | 0.675 | 0.269 | 0.054 | 99.8% | 8061 |

not a very satisfactory design. It commits the investigators to a very large and expensive trial in order to provide adequate power in the pessimistic range of risk reductions, without any evidence that the true risk reduction does indeed lie in the pessimistic range. Evidently a single group sequential design cannot provide adequate power for the "worst-case" scenario, and at the same time avoid overpowering the more optimistic range of scenarios. This leads us to consider building an adaptive sample size re-estimation option into the group sequential design Des 2, such that the adaptive component will provide the necessary insurance for the worst-case scenario, and thereby free the group sequential component to provide adequate power for the expected scenario, without a large and unnecessary up-front sample size commitment.

54.4.3 Adaptive Group Sequential Design

We convert the three-look group sequential design Des 2 into an adaptive group sequential design by inserting into it the option to increase the sample size at look 2, when 5600 subjects have been enrolled. Recreate the Des 2 by clicking on the  icon and just clicking **Compute** button. This will create Des 4 in the **Output Preview** pane. Save it in the workbook. The sample size re-estimation or adaptation can be done through simulations. The rules governing the sample size increase for Des 4 are

54 The Cui, Hung and Wang Method

similar to the rules specified in Section 53.2.4 for the schizophrenia trial, but tailored to the needs of the current trial. The idea is to identify unfavorable, promising and favorable zones for the interim results at look 2, based on the attained conditional power. Sample size should only be increased if the interim results fall in the promising zone. Subject to an upper limit, the sample size should be increased by just the right amount to boost the current conditional power to some desired level (say 80%). The following are the design specifications for Des 4:

1. The starting design is Des 2 with a sample size of 8000 subjects, one interim look after enrolling 4000 subjects and a second interim look after enrolling 5600 subjects. The efficacy stopping boundaries at these two interim looks are derived from the Lan and DeMets (1983) error spending function of the O'Brien-Fleming type.
2. At the second interim analysis, with data available on 5600 subjects, the conditional power is computed using the estimated value $\hat{\rho}$ as though it were the true relative risk ρ . If the conditional power is no greater than 30%, the outcome is deemed to be unfavorable. If the conditional power is between 30% and 80%, the outcome is deemed to be promising. If the conditional power is at least 80%, the outcome is deemed to be favorable.
3. If the interim outcome is promising, the sample size is re-computed so as to achieve 80% conditional power at the estimated value $\hat{\rho}$. The original sample size is then updated to the re-computed sample size, subject to the constraint in item 4 shown below.
4. If the re-computed sample size is less than 8000, the original sample size of 8000 subjects is used. If the re-computed sample size exceeds 16,000, the sample size is curtailed at 16,000 subjects.

Some features of this adaptive strategy are worth pointing out. First, the sample size is re-computed on the basis of data from 5600 subjects from the trial itself. Therefore the estimate of ρ available at the interim analysis is substantially more reliable than the estimate that was used at the start of the trial to compute an initial sample size of 8000 subjects. The latter estimate is typically derived from smaller pilot studies or from other phase 3 studies in which the patient population might not be exactly the same as that of the current trial. Second, a sample size increase is only requested if the interim results are promising, in which case the trial sponsor should be willing to invest the additional resources needed to power the trial adequately. In contrast Des 3 increases the sample size substantially at the very beginning of the trial, before any data are available to determine if the large sample size is justified.

54.4.4 Operating Characteristics of Adaptive Group Sequential Design

The East software provides a simulation tool for studying the consequences of increasing the sample size of Des 4 at the interim look. To implement this tool we must add the sample size re-estimation tab for Des 4. Select Des 4 in the **Library** and click the **S** icon. Click the **Include Options** button and select Sample Size Re-Estimation from the list. This will add a tab named as **Sample Size Re-estimation** as shown below:



The first two tabs **Simulation Parameters** and **Response Generation Info** contains all the information about the design Des 4 in the absence of any adaptive change. It is a three-look design with a sample size of 8000 and first interim look after 4000 subjects, second interim look after 5600 subjects and the last look after 8000 subjects. The early stopping boundaries generated by the $LD(OF)$ spending function equals -2.963 and -2.462 at the first look and the second look respectively.

The third tab **Sample Size Re-estimation** is used to specify the rules for modifying the initial sample size of Des 4, based on the data at the interim analysis. The description of these parameters is similar to what is described for the normal endpoint example in section 54.3.2. We will run simulations for different risk reduction values (15% to 25%) by changing the proportion response (treatment) values correspondingly from 0.85×0.087 to 0.75×0.087 . Before running the simulations we must input suitable values into the cells of this tab. enter the following values of proportion under treatment as 0.07395, 0.07221, 0.0696, 0.06699, 0.06525. The **Response Generation Info** Tab:

Specify Proportion

Prop. under Control (π_c):

Prop. under Treatment (π_t):

Table 54.7: Operating Characteristics of Des 2 (Group Sequential) and Des 4 (Adaptive Group Sequential) Designs

| Risk Reduction $100 \times (1 - \rho)$ | Des 2 (Group Sequential) | | Des 4 (Adaptive Group Sequential) | |
|---|--------------------------|----------------------|-----------------------------------|----------------------|
| | Power | Expected Sample Size | Power | Expected Sample Size |
| 15% | 57% | 7264 | 63% | 8265 |
| 17% | 68% | 7002 | 73% | 7919 |
| 20% | 82% | 6535 | 86% | 7289 |
| 23% | 92% | 6017 | 94% | 6543 |
| 25% | 95% | 5671 | 97% | 6027 |
| All results for Des 4 are based on 100,000 simulated trials | | | | |

design Des 2. As discussed previously in the schizophrenia example, the real benefit of an adaptive design is the opportunity it provides to invest in the trial in stages with the second stage investment forthcoming only if promising results are obtained at the first stage. To explain this better it is necessary to display power and expected sample size results conditional on the zone (unfavorable, promising or favorable) into which the results of the trial fall at the second interim analysis. To this end we run through the entire set of 100,000 simulations for Des 4 twice. In the first run we do not allow the sample size to change even when the conditional power lies in the promising zone. In effect we are simulating Des 2. The choice of simulation parameters for adaptation is as shown below:

Use Adaptation Method
 CHW CDL Müller and Schäfer Weights...

Adapt at: Look # 2

Max. Sample Size if Adapt (multiplier; total #): 1 8000

Target CP for Re-estimating Sample Size: 0.8

Promising Zone Scale: Cond. Power CP

| | | |
|-----------------|----------|-----|
| Promising Zone: | Min. CP: | 0.3 |
| | Max. CP: | 0.8 |

CP Computation Based on: Estimated (n_c, n_t)

These simulations produced 56% overall power and 15%, 57% and 83% power

This time the simulations produced 62% overall power and 15%, 80% and 83% power conditional on being in the unfavorable, promising and favorable zones, respectively.

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions (CHW Simulation)

| Simulation Parameters | |
|--------------------------------------|------------------------------|
| Simulation ID | CHWSim6 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 8000 |
| Variance | Unpooled Estimate |
| Avg. Power at Termination | 0.622 |
| Response Generation Parameters | |
| Prop. under Control (π_c) | 0.087 |
| Prop. under Treatment (π_t) | 0.074 |
| Sample Size Re-estimation Parameters | |
| Method of Adaptation | Cui-Hung-Wang |
| Adapt At Look No. | 2 |
| Max. Sample Size if Adapt | |
| Multiplier | 2 |
| Total # | 16000 |
| Target CP | 0.8 |
| Promising Zone Scale | Cond. Power |
| Min. CP | 0.3 |
| Max. CP | 0.8 |
| CP Computation Based on | Estimated (π_c, π_t) |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 100000 |

Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|-------------|--------------------------|----------|------------------------------|---------|-------------------|----------|---------------------|
| | Count | Row % | Count | Row % | Count | Column % | |
| Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| Unfavorable | 5404 | 14.989% | 30650 | 85.011% | 36054 | 36.054% | 8000 |
| Promising | 19611 | 80.245% | 4828 | 19.755% | 24439 | 24.439% | 12006.355 |
| Favorable | 11613 | 83.128% | 2357 | 16.872% | 13970 | 13.970% | 8000 |
| Efficacy | 25537 | 100.000% | 0 | 0.000% | 25537 | 25.537% | 5147.699 |
| All Trials | 62165 | 62.165% | 37835 | 37.835% | 100000 | 100.000% | 8250.721 |

Promising Zone defined as 0.3 <= CP < 0.8

Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 4000 |
| 2 | 5600 |
| 3 | 9314.899 |
| Average | 8250.721 |

Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | Total Simulations | |
|--------|-----------------|------------|----------|--------------|-------------------|---|
| | | Lower | Efficacy | | Count | % |
| 1 | 4000 | -2.963 | 7219 | 7219 | 7.219% | |
| 2 | 5600 | -2.462 | 18318 | 18318 | 18.318% | |
| 3 | 8000 | -2.002 | 36628 | 74463 | 74.463% | |
| Total | | | 62165 | 100000 | | |
| % | | | 62.165% | | | |

Similar simulation operations were carried out for other values of risk reduction under both the designs. Finally all these results representing the operating characteristics of both Des 2 and Des 4 conditional on the zone into which the conditional power falls at the second interim analysis, are displayed in Table 54.8.

(or) The table reveals substantial gains in power for Des 4 compared to Des 2 at all values of risk reduction if the second interim outcome falls in the promising zone, thereby leading to an increase in the sample size. Outside this zone the two designs have the same operating characteristics since the sample size does not change. If the second interim outcome falls in the unfavorable zone, the trial appears to be headed for failure and an additional sample size investment would be risky. If the second interim outcome falls in the favorable zone, the trial is headed for success without the need to increase the sample size. Thus the adaptive design provides the opportunity to increase the sample size only when the results of the second interim analysis fall in the promising zone. This is precisely when the trial can most benefit from a sample size increase.

54 The Cui, Hung and Wang Method

Table 54.8: Operating Characteristics of Des 2 (Group Sequential) and Des 4 (Adaptive Group Sequential) Designs Conditional on Second Interim Outcome

| Risk Reduction $100 \times (1 - \rho)$ | Second Interim Outcome | Probability of Interim Outcome | Power Conditional on Second Interim Outcome | | Expected Sample Size | |
|---|------------------------|--------------------------------|---|-------|----------------------|-------|
| | | | Des 2 | Des 4 | Des 2 | Des 4 |
| 15% | Unfavorable | 36% | 15% | 15% | 8000 | 8000 |
| | Promising | 24% | 57% | 81% | 8000 | 12098 |
| | Favorable | 40% | 94% | 94% | 6148 | 6147 |
| 17% | Unfavorable | 27% | 20% | 20% | 8000 | 8000 |
| | Promising | 24 % | 64% | 87% | 8000 | 11925 |
| | Favorable | 49 % | 96% | 96% | 5989 | 5989 |
| 20% | Unfavorable | 16% | 30% | 30% | 8000 | 8000 |
| | Promising | 20% | 73% | 93% | 8000 | 11781 |
| | Favorable | 64% | 98% | 98 % | 5726 | 5738 |
| 23% | Unfavorable | 8% | 40% | 40% | 8000 | 8000 |
| | Promising | 14% | 81% | 96% | 8000 | 11599 |
| | Favorable | 78% | 99% | 99% | 5440 | 5447 |
| 25% | Unfavorable | 5% | 48% | 48% | 8000 | 8000 |
| | Promising | 10% | 86% | 97% | 8000 | 11443 |
| | Favorable | 85% | 99.6% | 99.6% | 5253 | 5251 |

All results are based on 100,000 simulated trials

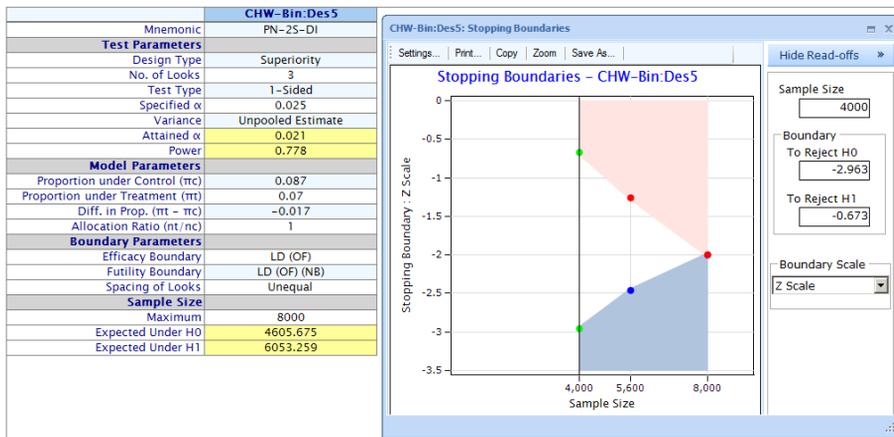
54.4.5 Adding a Futility Boundary

One concern with design Des 4 is that it lacks a futility boundary. There is thus the risk of proceeding to the end, possibly with a sample size increase, when the magnitude of the risk reduction is small and unlikely to result in a successful trial. In particular, suppose that the null hypothesis is true. In that case we can show that the power (i.e., the type-1 error) is 2.5% and the expected sample size under Des 4 is 8293 subjects. It might thus be desirable to include some type of futility stopping rule for the trial. In this trial the investigators proposed the following futility stopping rules at the two interim analysis time points:

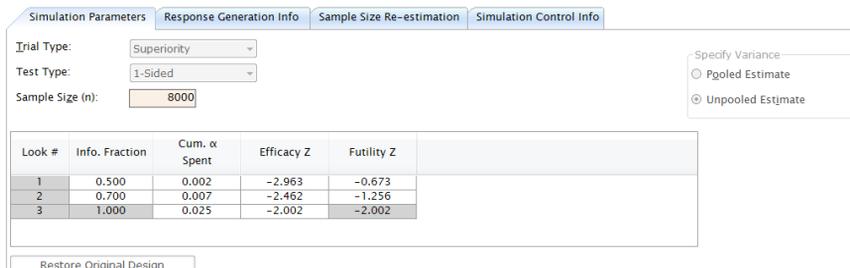
1. Stop for futility at the first interim analysis ($N = 4000$) if the estimated event rate for the experimental arm is at least 1% higher than the estimated event rate for the control arm
2. Stop for futility at the second interim analysis ($N = 5600$) if the conditional

power, based on the estimated risk ratio $\hat{\rho}$, is no greater than 20%

We will implement these futility rules by simulation. To this end create Des 5 with the same **LD (OF)** efficacy boundaries as Des 4, but also include non-binding **LD (OF)** futility boundaries after selecting Des 4 in the **Library** and clicking the  icon.



The futility boundary of Des 5 is not the one we intend to use. This is not a problem, however, since East permits us to edit all the boundaries in any of the simulation tabs. Accordingly we invoke the Simulations for Des 5 and add the **Sample Size Re-estimation** by selecting from the **Include Options** button. The following screen appears.



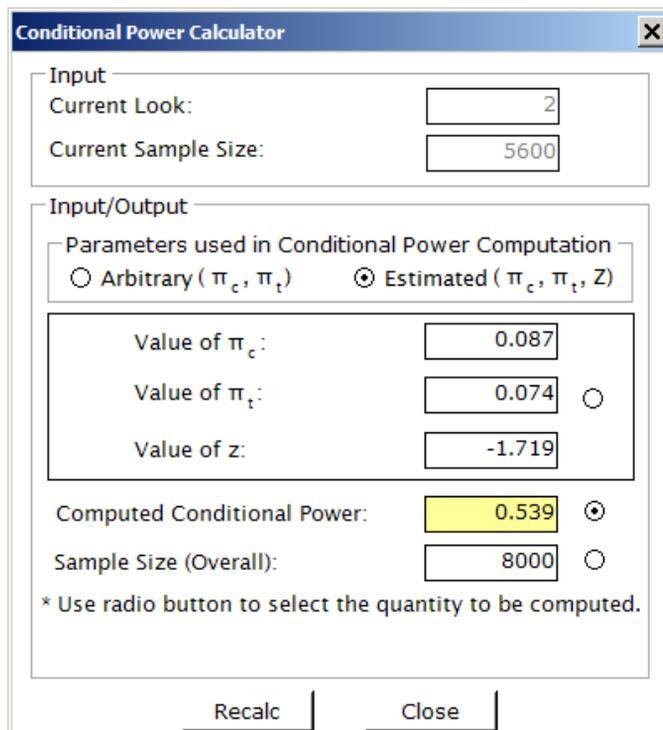
The first step is to edit the futility boundaries. The futility boundary for the first look, using the rule 1 mentioned at the beginning of this section, can be calculated manually as shown below:

54 The Cui, Hung and Wang Method

$$\begin{aligned} N &= 4000 \\ \pi_c &= 0.087 \\ \pi_t = 1.01\pi_c &= 0.08787 \\ \delta = \pi_t - \pi_c &= 0.00087 \\ se = \sqrt{\pi_c(1 - \pi_c)/2000 + \pi_t(1 - \pi_t)/2000} &= 0.008932521 \\ z = \delta/se &= 0.097396916 \end{aligned}$$

We will thus use 0.0974 as the futility boundary for the first interim look. Before we make this change and run CHW Simulations, however, we must determine the futility boundary for the second interim look, under rule 2. This is achieved by using the conditional power calculator available on the **Sample Size Re-estimation** tab. Click

on the  button for invoking CP Calculator. The following calculator appears.



| Field | Value |
|----------------------------|--------|
| Current Look | 2 |
| Current Sample Size | 5600 |
| Value of π_c | 0.087 |
| Value of π_t | 0.074 |
| Value of z | -1.719 |
| Computed Conditional Power | 0.539 |
| Sample Size (Overall) | 8000 |

We make the following changes to this dialog box. At the top of the **Input /Output** section we select the radio button that indicates that conditional power will be based on the values estimated at the interim look and not based on user-defined values. We choose the radio button from the three available at the right hand side of the dialog box that specifies what it is that we wish to compute. In the present case we wish to compute the Z-statistic that corresponds to a conditional power of 0.2, and so we select the top radio button from the three that are available. Finally, we edit box for Conditional Power and enter 0.2, since this is the conditional power for which we wish

54 The Cui, Hung and Wang Method

to determine the corresponding futility boundary.

The image shows a software dialog box titled "Conditional Power Calculator". It is divided into several sections:

- Input:** Contains two text input fields: "Current Look:" with the value "2" and "Current Sample Size:" with the value "5600".
- Input/Output:** Contains a sub-section titled "Parameters used in Conditional Power Computation" with two radio buttons: "Arbitrary (π_c, π_t)" (unselected) and "Estimated (π_c, π_t, Z)" (selected). Below this are three rows of input/output fields:
 - "Value of π_c :" with the value "0.087".
 - "Computed value of π_t :" with the value "Computed" and a radio button.
 - "Computed value of z:" with the value "Computed".
- Below the Input/Output section are two more text input fields: "Conditional Power:" with the value "0.2" and "Sample Size (Overall):" with the value "8000", each followed by an unselected radio button.
- At the bottom, there are two buttons: "Recalc" and "Close".

Below the input fields, there is a red asterisk followed by the text: "* Use radio button to select the quantity to be computed." and "* Please click the Recalc button."

Upon pressing the **Recalc** button, the calculator is updated.

We see that the Z-statistic corresponding to the futility boundary at look 2 is equal to -1.289. We may now edit the futility boundaries at look 1 and look 2 as shown below. Click on **Simulation Parameters** tab and edit the boundaries as given below:

Trial Type: Superiority
 Test Type: 1-Sided
 Sample Size (n): 8000

| Look # | Info. Fraction | Cum. α Spent | Efficacy Z | Futility Z |
|--------|----------------|--------------|------------|------------|
| 1 | 0.500 | | -2.963 | 0.0974 |
| 2 | 0.700 | | -2.462 | -1.2891 |
| 3 | 1.000 | | -2.002 | -2.0018 |

54 The Cui, Hung and Wang Method

We now proceed to simulate this trial as before. The other parameters on **Sample Size Re-estimation** tab are set as below:

| | | | |
|--|---------------------------|--|-------|
| Use Adaptation Method | | Weights... | |
| <input checked="" type="radio"/> CHW | <input type="radio"/> CDL | <input type="radio"/> Müller and Schäfer | |
| Adapt at: | Look # | 2 | |
| Max. Sample Size if Adapt (multiplier; total #): | | 2 | 16000 |
| Target CP for Re-estimating Sample Size: | | 0.8 | |
| Promising Zone Scale: | | Cond. Power | CP |
| Promising Zone: | Min. CP: | 0.3 | |
| | Max. CP: | 0.8 | |
| CP Computation Based on: | | Estimated (n_c, n_t) | |

The impact of the futility boundary on the unconditional operating characteristics of the Des 4 design are displayed in Table 54.9.

Table 54.9: Operating Characteristics of the Des 4 Design with and without a Futility Boundary

| Risk Reduction $100 \times (1 - \rho)$ | Des 4 with No Futility Boundary | | Des 4 with Futility Boundary | |
|---|---------------------------------|----------------------|------------------------------|----------------------|
| | Power | Expected Sample Size | Power | Expected Sample Size |
| 0% | 2.5% | 8259 | 2.81% | 5339 |
| 15% | 63% | 8265 | 59% | 7440 |
| 20% | 86% | 7289 | 83% | 6939 |
| 25% | 97% | 6027 | 95% | 5928 |
| All results are based on 100,000 simulated trials | | | | |

The inclusion of the futility boundary has resulted in a dramatic saving of more than 3000 subjects, on average, at the null hypothesis of no risk reduction. Furthermore, notwithstanding a small power loss of 2-5%, the trial continues to have well over 80% power for risk reductions of 20% or more. The trial suffers a power loss of 7% if the magnitude of the risk reduction is 15%, the low end of the range of clinical interest. In this situation, however, the unconditional power is inadequate (only 63%) even without a futility boundary.

To fully appreciate the impact of the futility boundary on power and expected sample size, it is necessary to study the operating characteristics of the trial conditional on the results of the second interim analysis. These results are displayed in Table 54.10.

54 The Cui, Hung and Wang Method

Table 54.10: Operating Characteristics of Des 4 Design with and without a Futility Boundary, Conditional on the Second Interim Outcome

| Risk Reduction $100 \times (1 - \rho)$ | Second Interim Outcome | Prob. of Interim Outcome | Power Conditional on Second Interim Outcome | | Expected Sample Size | |
|---|------------------------|--------------------------|---|----------|----------------------|----------|
| | | | No Fut | With Fut | No Fut | With Fut |
| 0% | Unfav + Fut | 92% | 0.44% | 0.14% | 8000 | 4851 |
| | Promising | 6% | 15% | 16% | 12985 | 12946 |
| | Fav + Eff | 2% | 64% | 64% | 6918 | 6923 |
| 15% | Unfav + Fut | 36% | 15% | 5 % | 8000 | 5705 |
| | Promising | 24 % | 81% | 81% | 12098 | 12098 |
| | Fav + Eff | 40 % | 94% | 94% | 6147 | 6139 |
| 20% | Unfav + Fut | 16% | 30% | 10.2% | 8000 | 5930 |
| | Promising | 20% | 93% | 93% | 11781 | 11746 |
| | Fav + Eff | 64% | 98% | 98% | 5738 | 5729 |
| 25% | Unfav + Fut | 5% | 47% | 18% | 8000 | 6106 |
| | Promising | 10% | 98% | 97% | 11443 | 11443 |
| | Fav + Eff | 85% | 99.5% | 99.5% | 5251 | 5245 |
| All results are based on 100,000 simulated trials | | | | | | |

It is seen that the presence of the futility boundary does not cause any loss of power for trials that enter the promising or favorable zones at the second interim analysis. Additionally the presence of the futility boundary causes the average sample size to be reduced substantially in the unfavorable zone, moderately in the promising zone while remaining the same in the favorable zone. In effect, the futility boundary terminates a proportion of trials that enter the unfavorable zone thereby preventing them from proceeding to conclusion. It has no impact on trials that enter the favorable zone.

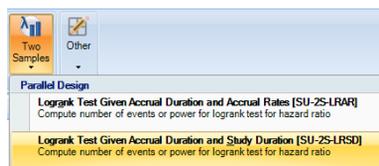
54.5 Survival Endpoint: Lung Cancer Trial

A two-arm multi-center randomized clinical trial is planned for subjects with advanced metastatic non-small cell lung cancer with the goal of comparing the current standard second line therapy (docetaxel+cisplatin) to a new docetaxel containing combination regimen. The primary endpoint is Overall Survival (OS). The study is required to have one-sided $\alpha = 0.025$, and 90% power to detect an improvement in median survival, from 8 months on the control arm to 11.4 months on the experimental arm, which corresponds to a hazard ratio of 0.7. We shall first create a group sequential design for this study in East, and shall then show how the design may be improved by permitting an increase in the number of events and sample size at the time of the interim analysis.

54.5.1 Group Sequential Design

We begin by constructing a two-look group sequential design with an efficacy boundary derived from the Lan and DeMets (1983) O’Brien-Fleming type spending function, a futility boundary derived from the γ -spending function of Hwang, Shih and DeCani (1990) with parameter $\gamma = -5$, and an interim analysis at 50% of the total information. It is required to enroll subjects over 24 months and extend the follow-up for six additional months, thereby completing the study in 30 months.

We begin by using East to design a trial under these basic assumptions. First, click **Survival: Two Samples** on the **Design** tab and then click **Parallel Design: Logrank Test Given Accrual Duration and Study Duration** as shown below.



This will launch a new input window. Enter the appropriate design parameters into the dialog box as shown below. Enter median survival times of 8 months for the Control arm and a hazard ratio of 0.7

Design Type: Superiority Number of Looks: 2

Design Parameters Boundary Info Accrual/Dropout Info

Test Type: 1-Sided # of Hazard Pieces: 1 Input Method: Median Survival Times

Type I Error (α): 0.025 Hazard Ratio (Optional) Alternative

Power: 0.9 Hazard Ratio (λ_t/λ_c) 0.7

Sample Size (n): Computed Ratio of Medians (m_t/m_c) 1.429

No. of Events: Computed

| Period # | Med. Surv. Time (Control) | Med. Surv. Time (Treatment: Alt.) |
|----------|---------------------------|-----------------------------------|
| 1 | 8.000 | 11.429 |

Allocation Ratio: 1 (n_t/n_c)

Variance of Log Hazard Ratio Null Alternative

Next, click on the **Boundary Info** tab. Be sure to select the nonbinding futility

54 The Cui, Hung and Wang Method

boundary as below:

| | | | | | | | |
|--|----------------|-------------------------------------|-------------------------------------|---|-------------------|--------------------|-------------------|
| Efficacy Boundary Family: Spending Functions Spending Function: Lan-DeMets Parameter: OF Type I Error (α): 0.025 | | | | Futility Boundary Family: Spending Functions Spending Function: Gamma Family Parameter (γ): -5 Type II Error (β): 0.1 | | | |
| Spacing of Looks: <input checked="" type="radio"/> Equal <input type="radio"/> Unequal | | | | Boundary Scale: Z Scale | | | |
| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Cum. β Spent | Futility Boundary |
| 1 | 0.500 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.002 | -2.963 | 0.008 | 0.127 |
| 2 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.025 | -1.969 | 0.100 | -1.969 |

Next click on the **Accrual/Dropout Info** tab. The **Accrual Duration** is 24 and the **Study Duration** is 30. In this trial everyone will be followed for survival until the end of the study, thus the **Until End of Study** entry is selected.

| | | |
|--|---------|----------------|
| Subjects are followed: Until End of Study | | |
| Accrual Info Accrual Duration: 24 Study Duration: 30 # of Accrual Periods: 1 | | |
| Period # | By Time | Cum. % Accrued |
| 1 | 24.000 | 100.000 |

| | | | |
|--|------------------|-----------------------|-------------------------|
| Piecewise Dropout Information # of Pieces: 0 Input Method: Hazard Rates | | | |
| Period # | Starting at Time | Hazard Rate (Control) | Hazard Rate (Treatment) |
| | | | |

Click **Compute** to complete the design. Here is the Output Summary of this design.

| CHW-Surv:Des 1 | |
|---|--------------------|
| Mnemonic | SU-25-LRSD |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 2 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Attained α | 0.025 |
| Power | 0.901 |
| Model Parameters | |
| Hazard Ratio (Alt.) | 0.7 |
| Var (Log HR) | Null |
| Allocation Ratio (nt/nc) | 1 |
| Boundary Parameters | |
| Futility Boundary | Gm (-5) (NB) |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Accrual & Dropout Parameters | |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 483 |
| Expected Under H0 | 421.701 |
| Expected Under H1 | 453.671 |
| Events | |
| Maximum | 334 |
| Expected Under H0 | 258.635 |
| Expected Under H1 | 290.11 |
| Study Duration | |
| Maximum | 30 |
| Expected Under H0 | 22.92 |
| Expected Under H1 | 26.954 |
| Accrual Duration | |
| Maximum | 24 |
| Expected Under H0 | 20.954 |
| Expected Under H1 | 22.543 |

Des 1 requires an up-front commitment of 334 events to achieve 90% power. With an enrollment of 483 subjects over 24 months, the required 334 events are expected to arrive within 30 months. An interim analysis will be performed after 167 events are obtained (50% of the total information). Under the alternative hypothesis that the hazard ratio is 0.7, the chance of crossing the efficacy boundary at the interim look is about 26% leading to an expected sample size of 454 subjects and an expected study duration of 27 months. Keeping the cursor on Des1 node, if you click on the 

54 The Cui, Hung and Wang Method

icon, you will see the following output.

⊖ Sample Size Information

| | Control Arm | Treatment Arm | Total |
|------------------------------|-------------|---------------|---------|
| Sample Size (n) | | | |
| Maximum | 241 | 242 | 483 |
| Expected H1 | 226.836 | 226.836 | 453.671 |
| Expected H0 | 210.851 | 210.851 | 421.701 |
| Events (s) | | | |
| Maximum | 180 | 154 | 334 |
| Expected H1 | 162.827 | 135.155 | 290.11 |
| Expected H0 | 130.437 | 130.437 | 258.635 |
| Maximum Information (I):83.5 | | | |

⊖ Accrual and Study Duration

| | Accrual Duration | Study Duration |
|-------------|------------------|----------------|
| Maximum | 24 | 29.984 |
| Expected H1 | 22.543 | 26.954 |
| Expected H0 | 20.954 | 22.92 |

54.5.2 Adaptive Design: Motivation

Des1 is adequately powered to detect a hazard ratio of 0.7. It is possible however, either because the new treatment is somewhat less effective than anticipated or because of improved standard of care for patients on the control arm, that the underlying hazard ratio could be larger. If this were the case, the study would be underpowered. For example, if the true hazard ratio was 0.77, an effect that is still considered clinically meaningful, the power of a 483-subject study would drop from 90% to 67.2% as

shown below under Des2.

Test Type: 1-Sided # of Hazard Pieces: 1 Input Method: Median Survival Times

Type I Error (α): 0.025

Power: Computed

Sample Size (n): 483

No. of Events: Computed

Allocation Ratio: 1 (n_1/n_2)

Hazard Ratio (Optional) Alternative 0.77

Hazard Ratio (λ_1/λ_2)

Ratio of Medians (m_1/m_2) 1.299

| Period # | At | Med. Surv. Time (Control) | Med. Surv. Time (Treatment: Alt.) |
|----------|----|---------------------------|-----------------------------------|
| 1 | | 8.000 | 10.390 |

Variance of Log Hazard Ratio

Null Alternative

| | Wbk1:Des1 | Wbk1:Des2 |
|------------------------|-------------|-------------|
| Mnemonic | SU-2S-LRSD | SU-2S-LRSD |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 2 | 2 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Attained α | 0.025 | 0.025 |
| Power | 0.901 | 0.672 |

Thus one possibility would be to design the trial from the very beginning to have 90% power to detect a hazard ratio of 0.77.

Test Type: 1-Sided # of Hazard Pieces: 1 Input Method: Median Survival Times

Type I Error (α): 0.025

Power: 0.9

Sample Size (n): Computed

No. of Events: Computed

Allocation Ratio: 1 (n_1/n_2)

Hazard Ratio (Optional) Alternative 0.77

Hazard Ratio (λ_1/λ_2)

Ratio of Medians (m_1/m_2) 1.299

| Period # | At | Med. Surv. Time (Control) | Med. Surv. Time (Treatment: Alt.) |
|----------|----|---------------------------|-----------------------------------|
| 1 | | 8.000 | 10.390 |

Variance of Log Hazard Ratio

Null Alternative

54 The Cui, Hung and Wang Method

Such a design is displayed below as Des 3 and requires 621 events. In order to complete the trial in 30 months it would be necessary to enroll 878 subjects over 24 months with an additional 6 months of follow-up.

| | CHW-Surv:Des 1 | CHW-Surv:Des 2 | CHW-Surv:Des 3 |
|---|--------------------|--------------------|--------------------|
| Mnemonic | SU-25-LRSD | SU-25-LRSD | SU-25-LRSD |
| Test Parameters | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 2 | 2 | 2 |
| Test Type | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 | 0.025 |
| Attained α | 0.025 | 0.025 | 0.025 |
| Power | 0.901 | 0.672 | 0.9 |
| Model Parameters | | | |
| Hazard Ratio (Alt.) | 0.7 | 0.77 | 0.77 |
| Var (Log HR) | Null | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Boundary Parameters | | | |
| Futility Boundary | Gm (-5) (NB) | Gm (-5) (NB) | Gm (-5) (NB) |
| Spacing of Looks | Equal | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) |
| Accrual & Dropout Parameters | | | |
| Subjects are Followed | Until End of Study | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 | 1 |
| No. of Dropout Pieces | 0 | 0 | 0 |
| Sample Size | | | |
| Maximum | 483 | 483 | 878 |
| Expected Under H0 | 421.701 | 430.32 | 770.565 |
| Expected Under H1 | 453.671 | 468.22 | 823.951 |
| Events | | | |
| Maximum | 334 | 341 | 621 |
| Expected Under H0 | 258.635 | 272.475 | 480.693 |
| Expected Under H1 | 290.11 | 318.735 | 539.239 |
| Study Duration | | | |
| Maximum | 30 | 30 | 30 |
| Expected Under H0 | 22.92 | 23.853 | 23.356 |
| Expected Under H1 | 26.954 | 28.407 | 26.922 |
| Accrual Duration | | | |
| Maximum | 24 | 24 | 24 |
| Expected Under H0 | 20.954 | 21.382 | 21.063 |
| Expected Under H1 | 22.543 | 23.266 | 22.523 |

The sponsor is either unable or unwilling to make such a large sample size commitment up-front purely on the basis of the limited prior data available on the new compound. However, since an independent data monitoring committee (DMC) will be reviewing the interim efficacy data in an unblinded fashion at 50% of the total information, the sponsor might be prepared to authorize the investment of additional resources on the recommendation this committee. In a manner analogous to the pre-specification of group sequential boundaries for early stopping, the sponsor must pre-specify to the DMC the precise data dependent rules for increasing the number of events and sample size at the time of the interim analysis. (Note, however, that these rules may be modified at the time of the interim analysis if the DMC believes it is in the best interests of the patients to modify them. The statistical methodology described in this volume permits such modifications without type-1 error inflation.) These rules are best constructed with the help of the simulation tools available in East as we now show.

54.5.3 Adaptive Design: Construction

The starting point for constructing the adaptive design is the group sequential design, Des1. This design is entirely satisfactory if the true hazard ratio is 0.7 but is unsatisfactory if the hazard ratio is 0.77, a hazard ratio that is still clinically meaningful. Designing a group sequential trial to detect a hazard ratio of 0.77, as in Des 3 above, is unfortunately not an option, for it requires too large a commitment of resources up front. It is possible, however, for the sponsor to start out with Des1, requiring only 334 events and 483 subjects, but build in the option for an increase in the number of events and subjects if the results obtained at the interim analysis are promising.

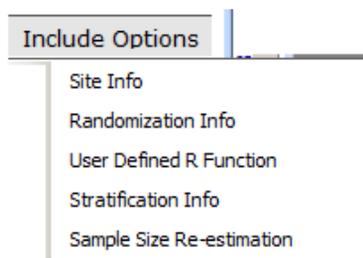
The adaptive design is constructed by means of simulation. Select Des1 in the **Library** and click the  icon. You will be taken to the following simulation input window.

Trial Type: Superiority
 Test Type: 1-Sided
 Maximum Events: 334
 Fix at Each Look: Total No. of Events
 Test Statistic: LogRank

| Look # | Info. Fraction | Cum. α Spent | Efficacy Z | Futility Z |
|--------|----------------|---------------------|------------|------------|
| 1 | 0.500 | 0.002 | -2.963 | 0.126 |
| 2 | 1.000 | 0.025 | -1.969 | -1.969 |

Restore Original Design

In addition to the four tabs appearing by default on inserting Simulations, one can add more tabs to enter information available on randomization, stratification and sample size re-estimation. This can be done by clicking the **Include Options** button on this right hand top corner of the screen.



The **Sample Size Re-estimation** tab is added by clicking the appropriate option as shown above. Let us focus on five such tabs shown below. Several parameters on these

54 The Cui, Hung and Wang Method

tabs can play vital role in simulation and adaptation of a design.



The default values on the **Simulation Parameters** tab are those that were specified at the design stage. However, all the entries in the white cells are editable and can be used to alter the simulation parameters. Thus we could alter the **Info Fraction**, **Cum. α** spent and the **Simulation Boundaries** as well.

| Look # | Info. Fraction | Cum. α Spent | Efficacy Z | Futility Z |
|--------|----------------|---------------------|------------|------------|
| 1 | 0.500 | 0.002 | -2.963 | 0.126 |
| 2 | 1.000 | 0.025 | -1.969 | -1.969 |

or we could alter the **Survival Information** on the **Response Generation Info** tab

Survival Information

Using Hazard Rates
 Using Cum. % Survival

of Hazard Pieces

| Piece | Starting At | Hazard Rates | | Hazard Ratio |
|-------|-------------|--------------|-----------|--------------|
| | | Control | Treatment | |
| 1 | 0.000 | 0.0866 | 0.0607 | 0.700 |

or we could alter the Accrual and Dropout Information on the **Accrual/Dropout Info** tab.

Sample Size:

Subjects are followed:

Accrual Info

Accrual Duration:

of Accrual Periods: Input Method:

| Period # | At | Cum. % Accrued |
|----------|--------|----------------|
| 1 | 24.000 | 100.000 |

Distribution of Accrual Time: Uniform

Piecewise Constant Dropout Rates

of Pieces: Input Method:

| Period # | Starting At | Hazard Rate (Control) | Hazard Rate (Treatment) |
|----------|-------------|-----------------------|-------------------------|
| | | | |

and so on. Suppose, for example that we wish to edit the input parameters in the **Survival Information** panel. The current panel displays hazard rates of 0.0866

and 0.0607 for the Control and Treatment arms, respectively, implying a hazard ratio of 0.7

We know from the design Des 1 that a hazard ratio of 0.7 will yield 90% power. But what if the true hazard ratio was 0.77? The resultant deterioration in power can be evaluated by simulation. Accordingly we shall alter the Treatment cell, containing the hazard 0.0607, by replacing it with $0.77 * 0.0866 = 0.0667$.

Survival Information

of Hazard Pieces: Input Method:

Hazard Ratio

| Piece | Starting at Time | Hazard Rates | | Hazard Ratio |
|-------|------------------|--------------|-----------|--------------|
| | | Control | Treatment | |
| 1 | 0.000 | 0.087 | 0.067 | 0.770 |

The total number of simulations shall be 10000 and the screen will be refreshed after every 1000 trials.

Number of Simulations:

Refresh Frequency:

Random Number Seed

Clock

Fixed

Output Options

Save summary statistics for every simulation run

Save subject-level data for simulation runs

Note: Max. 100,000 records will be saved.

Simulation without Adaptation:

Note that we have not changed any of the adaptation parameters on the Sample Size Re-estimation tab. This means we are not carrying out any adaptation at this point of time. To run 10,000 simulations with a hazard ratio of 0.77, click on the **Simulate**

54 The Cui, Hung and Wang Method

button. The following simulation output is displayed.

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

| Simulation Parameters | |
|-------------------------------|---------------------|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 2 |
| Test Type | 1-Sided |
| Sample Size (n) | 483 |
| Fix at Each Look | Total No. of Events |
| Test Statistic | Logrank |
| Average Events | 312.223 |
| Total Accrual Duration | 24 |
| Avg. Power at Termination | 0.654 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

Average Sample Size and Look Times

| Look # | Average Sample Size | Average Events | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|-------------------|-------------------|
| | | Control | Treatment | | |
| 1 | 364.604 | 90.608 | 76.392 | 18.093 | 5.998 |
| 2 | 483 | 177.341 | 156.659 | 29.233 | 9.091 |
| Average | 467.564 | 166.39 | 145.833 | 27.779 | 8.687 |

Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Events | Boundaries | | Stopping For | | Total Simulations | |
|--------|--------|------------|----------|--------------|----------|-------------------|---------|
| | | Efficacy | Futility | Efficacy | Futility | Count | % |
| | | Lower | Upper | | | | |
| 1 | 167 | -2.963 | 0.126 | 962 | 342 | 1304 | 13.040% |
| 2 | 334 | -1.969 | -1.969 | 5579 | 3117 | 8696 | 86.960% |
| Total | | | | 6541 | 3459 | 10000 | |
| % | | | | 65.410% | 34.590% | | |

Response Generation Parameters

No. of Hazard Pieces: 1
Input Method: Hazard Rates

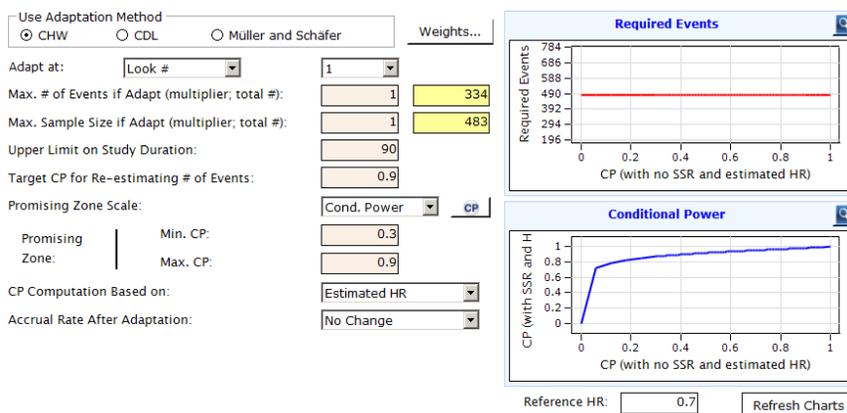
| Piece # | Starting at Time | Control | Treatment | Hazard Ratio |
|---------|------------------|---------|-----------|--------------|
| 1 | 0 | 0.087 | 0.067 | 0.77 |

The overall power is only 65.4% suggesting that it might be useful to consider an adaptive increase in the number of events and sample size at the interim look.

The “Sample Size Re-Estimation” Tab

Select **CHWSim1** in the **Library** and click the  icon. You will be taken to the following simulation input window.

The impact of an adaptive increase in the number of events and sample size on power and study duration can be evaluated by simulation. Click the **Sample Size Re-estimation** tab. This tab contains the input parameters for performing the adaptive simulations and sample size re-estimation in the on-going trial.



Use Adaptation Method: CHW CDL Müller and Schäfer

Adapt at: Look # [1] [1] [334]

Max. # of Events if Adapt (multiplier; total #): [1] [483]

Max. Sample Size if Adapt (multiplier; total #): [1] [483]

Upper Limit on Study Duration: [90]

Target CP for Re-estimating # of Events: [0.9]

Promising Zone Scale: Cond. Power [0.3] [CP]

Promising Zone: Min. CP: [0.3] Max. CP: [0.9]

CP Computation Based on: [Estimated HR]

Accrual Rate After Adaptation: [No Change]

Required Events: [784, 686, 588, 490, 392, 294, 196] vs CP (with no SSR and estimated HR) [0, 0.2, 0.4, 0.6, 0.8, 1]

Conditional Power: [1, 0.8, 0.6, 0.4, 0.2, 0] vs CP (with no SSR and estimated HR) [0, 0.2, 0.4, 0.6, 0.8, 1]

Reference HR: [0.7] Refresh Charts

The **Sample Size Re-estimation** tab is the main location from which you will be using East to design adaptive time-to-event trials. The left hand side of this tab contains the Input Parameters for adaptive simulations and the right hand side contains two charts.

54 The Cui, Hung and Wang Method

Input Parameters for Sample Size Re-estimation

This window consists of 10 input fields into which one may enter various design parameters.

| | | | |
|--|---------------------------|--|--|
| Use Adaptation Method | | Weights... | |
| <input checked="" type="radio"/> CHW | <input type="radio"/> CDL | <input type="radio"/> Müller and Schäfer | |
| Adapt at: | Look # | 1 | |
| Max. # of Events if Adapt (multiplier; total #): | 1.5 | 501 | |
| Max. Sample Size if Adapt (multiplier; total #): | 1.5 | 724 | |
| Upper Limit on Study Duration: | 90 | | |
| Target CP for Re-estimating # of Events: | 0.9 | | |
| Promising Zone Scale: | Cond. Power | CP | |
| Promising Zone: | Min. CP: | 0.3 | |
| | Max. CP: | 0.9 | |
| CP Computation Based on: | Estimated HR | | |
| Accrual Rate After Adaptation: | No Change | | |

For a given set of design parameters, East will run a number of simulated trials as specified in the **Simulation Control Info** tab:

| Number of Simulations: 100000 Refresh Frequency: 1000 Random Number Seed <input checked="" type="radio"/> Clock <input type="radio"/> Fixed 100 <input type="checkbox"/> Suppress All Intermediate Output <input type="checkbox"/> Pause after Refresh <input checked="" type="checkbox"/> Stop At End | Output Options Output Type: Case Data <input type="checkbox"/> Save summary statistics for every simulation run <input type="checkbox"/> Save subject-level data for 1 simulation runs Note: Max. 100,000 records will be saved. | | | | | | |
|---|--|----------------|---|----|----|----|----|
| | Output for All Trials <table border="1"> <thead> <tr> <th>Percentile (%)</th> </tr> </thead> <tbody> <tr><td>5</td></tr> <tr><td>25</td></tr> <tr><td>50</td></tr> <tr><td>75</td></tr> <tr><td>95</td></tr> </tbody> </table> | Percentile (%) | 5 | 25 | 50 | 75 | 95 |
| Percentile (%) | | | | | | | |
| 5 | | | | | | | |
| 25 | | | | | | | |
| 50 | | | | | | | |
| 75 | | | | | | | |
| 95 | | | | | | | |

On running the simulations, an entry for Simulation output gets added in the **Output**

Preview pane and the detailed output can be seen in the Output Summary of Simulations.

The input quantities in the **Sample Size Re-estimation** tab are described below in detail.

1. **Adaptation at:** For a K -look group sequential design, one can decide the time at which conditions for adaptations are to be checked and actual adaptation is to be carried out. This can be done either at some intermediate look or after accumulating data on specified number of events or after some specified information fraction. The value of this parameter depends upon the choice of the user. If it is **Look no.** then this parameter can be any integer number from 1 to $K - 1$. If the adaptation is to be carried out after observing specified events then this parameter can be some integer between [4, No. of events at design stage] and so on. The default choice in East is **look number** to decide the time of adaptation.

Adapt at:

2. **Max Number of Events if Adapt :** This quantity is a multiplier with value ≥ 1 for specifying the upper limit (or cap) on the increase in the number of events, should an adaptive increase be called for based on the target conditional power. Notice that, in keeping with the FDA Guidance on Adaptive Clinical Trials (2010), East does not permit an adaptive decrease in the number of events. Therefore multipliers less than 1 are not accepted in this cell. For example, if you use the multiplier 1.5 and if adaptation takes place, the modified number of events is capped at 501. The 501-event cap becomes effective only if the increased number of events (as calculated by the criteria of cells 4, 5 and 6) exceed 501.

Max. # of Events if Adapt (multiplier; total #):

| | |
|-----|-----|
| 1.5 | 501 |
|-----|-----|

3. **Max Subjects if Adapt :** This quantity is a multiplier with value ≥ 1 for specifying the upper limit (or cap) on the number of subjects to be enrolled in the study. Although the power of the trial is determined by the number of events

54 The Cui, Hung and Wang Method

and not the number of subjects, the number of subjects play a role in determining how long it will take to observe the required number of events, and hence for determining the study duration. The number of subjects may only be increased, never decreased. Therefore multipliers less than 1 are not accepted in this cell. For example, if you use the multiplier 1.5 and if adaptation takes place, the modified number of subjects is capped at 724 subjects. The trial will continue to enroll subjects until either the required number of events is reached or the cap on the number of subjects is reached.

Max. Sample Size if Adapt (multiplier; total #):

- Upper Limit on Study Duration :** An event driven trial ordinarily continues until the required number of events arrive. This input parameter is provided merely as a safety factor in order to prevent the trial from being prolonged excessively should the required number of events be very large or their rate of arrival be very slow. Its default value is set at three times the expected study duration obtained from the initial design of the trial. Consequently, if the scenarios being simulated are realistic, the required number of events will almost always be attained much before this upper limit parameter becomes operational. It is recommended to leave this parameter unchanged at least for the initial set of simulation experiments since it would interfere with the operating characteristics of the study if it were to become operational.

Upper Limit on Study Duration:

- Target Conditional Power for Re-estimating Events :** This parameter ranges between 0 and 1 and is the *target conditional power* desired at the end of the study. Suppose, for example that the Target CP is set at 0.9.

Target CP for Re-estimating # of Events:

Let the value of the test statistic obtained in the current simulation be z_L at

look L , where an adaptive increase in the number of events is being considered.

Then, by setting the left hand side of equation (54.21) to 0.9 we have:

$$0.9 = 1 - \Phi \left\{ b_K \sqrt{1 + \frac{D_L}{D_K - D_L}} - z_L \sqrt{\frac{D_L}{D_K - D_L}} - \delta \sqrt{r(1-r)} \sqrt{D_K^* - D_L} \right\}. \quad (54.23)$$

Upon solving equation (56.11) for D_K^* we obtain the increased number of events that are needed to achieve the target conditional power of 0.9 in this simulation. Let us illustrate with Des 1. In Des 1 $K = 2, L = 1, r = 0.5$ and the critical value for declaring statistical significance at the end of the trial is $b_2 = -1.9687$, as can be seen by examining the stopping boundaries displayed in the **Simulation Parameters** tab. The interim analysis is performed when $D_1 = 167$ events are obtained. In the absence of any adaptive change, the trial will terminate when $D_2 = 334$ events are obtained. Suppose the current simulation generates a value $z_1 = 1.5$ for the logrank statistic at look 1. Since the target conditional power is 0.9, equation (56.11) takes the form

$$0.9 = 1 - \Phi \left\{ -1.9687 \sqrt{1 + \frac{167}{334 - 167}} - 1.5 \sqrt{\frac{167}{334 - 167}} - 0.5\delta \sqrt{D_2^* - 167} \right\}. \quad (54.24)$$

In order to evaluate D_2^* , however, it is necessary to specify a value for the log hazard ratio δ in equation (56.12). This parameter is of course unknown. East gives you the option to perform simulations with either the current estimate $\hat{\delta}_1$ or to use the value of δ specified under the alternative hypothesis at the design stage. The choice can be made by selecting **Estimated HR** or **Design HR** from a drop-down list of the quantity **CP Computation Based on** of the **Sample Size Re-estimation** tab.

The default value is **Estimated HR**, (or equivalently $\hat{\delta}_1 = \ln \hat{HR}_1$) and we recommend using this default until you have gained some experience with the simulation output and can judge for yourselves which option provides better operating characteristics for your studies. East uses the formula

$$\hat{\delta}_1 = \frac{z_1}{\sqrt{r(1-r)}D_1}$$

to obtain the current estimate of δ . Upon substituting $z_1 = 1.5, D_1 = 167$ and $r = 0.5$ in the above expression we obtain $\hat{\delta}_1 = 0.232$, or equivalently a hazard ratio estimate of $\exp(0.232) = 1.2611$. Substituting the estimate of $\hat{\delta}_1$ into equation (56.12) and solving for D_2^* yields $D_2^* = 656$. Since the maximum number of events has been capped at 501, this simulation will terminate the trial when the number of events reaches 501 instead of going all the way to 656 events. In this case the desired target conditional power of 0.9 will not be met.

54 The Cui, Hung and Wang Method

Indeed in this case the conditional power (with $\hat{\delta}_1$ being used in place of the unknown true δ) is only

$$1 - \Phi \left\{ 1.9687 \sqrt{1 + \frac{167}{333 - 167}} - 1.5 \sqrt{\frac{333}{333 - 167}} - 0.5\delta \sqrt{500 - 167} \right\} = 0.798$$

For a more detailed discussion of conditional power, including the use of a special conditional power calculator that computes conditional power accurately without relying on the approximate assumption that the next look will be the last one, see Chapter 57.

6. **Promising Zone Scale :** Promising Zone is such that the number of events will only be increased if the conditional power at the interim look falls in this zone. East asks you to select the scale on which the promising zone is to be defined. It can be defined based on the conditional power or the test statistic or the estimated effect size and should be specified by entering the minimum and maximum of these quantities.
Let us go ahead with the default option which is Conditional Power.
7. **Promising Zone – Min CP :** In this cell you specify the minimum conditional power (in the absence of any adaptive change) at which you will entertain an increase in the number of events. That is, you specify the lower limit of the promising zone.
8. **Promising Zone – Max CP :** In this cell you specify the maximum conditional power (in the absence of any adaptive change) at which you will entertain an increase in the number of events. That is, you specify the upper limit of the promising zone.
Suppose, for example, that the number of events is only increased in a promising zone specified by the range $0.45 \leq CP < 0.8$, and suppose that in that case, the number of events is re-estimated so as to achieve a target conditional power of 0.99. Then the Input Parameters Table will contain the entries shown below.

| | |
|--|---------------|
| <u>T</u> arget CP for Re-estimating # of Events: | 0.99 |
| <u>P</u> romising Zone Scale: | Cond. Power ▼ |
| Promising Zone: Min. CP: | 0.45 |
| Max. CP: | 0.8 |

The zone to the left of the *promising zone* ($CP < 0.45$) is known as the

unfavorable zone. The zone to the right of the *promising zone* ($CP \geq 0.8$) is known as the *favorable zone*. In a group sequential design that includes early stopping boundaries for futility and efficacy, the unfavorable zone contains within it an even more extreme region for early futility stopping and the favorable zone contains within it an even more extreme region for early efficacy stopping.

9. **HR Used in CP Computations:** In this cell you specify whether the simulations should utilize conditional power based on $\hat{\delta}_L$ estimated at the time of the interim analysis or should utilize the value of δ specified under the alternative hypothesis, in equations (54.21) and (56.11). The adaptive design will have rather different operating characteristics in each case. The default is to use the estimated value $\hat{\delta}_L$.

CP Computation Based on:

Estimated HR ▼

10. **Accrual Rate After Adaptation :** East gives you the option to alter the rate of enrollment after an adaptive increase in the number of events. This feature would be useful, for example, to evaluate the extent to which the follow-up time and hence the total study duration can be shortened if the rate of enrollment is increased after the adaptive change is implemented.

Accrual Rate After Adaptation:

No Change ▼

54 The Cui, Hung and Wang Method

Required Events Chart

The upper chart at the extreme right of the **Sample Size Re-estimation** tab is called the Required Events Chart. The X-axis of this chart, labeled **CP(Dsgn.Events, Est.HR)**, tracks the conditional power obtained at the interim look based on the total number of events D_K specified at the design stage (334 events under Des 1) and the interim estimate $\hat{\delta}_L$ of the log hazard ratio. To be specific,

$$= 1 - \Phi \left\{ b_K \sqrt{1 + \frac{D_L}{D_K - D_L}} - z_L \sqrt{\frac{D_L}{D_K - D_L}} - \hat{\delta}_L \sqrt{r(1-r)} \sqrt{D_K - D_L} \right\} \quad \text{CP(Dsgn.Events, Est.HR)} \quad (54.25)$$

Since $\hat{\delta}_L$ and z_L are related through the relationship

$$\hat{\delta}_L = \frac{z_L}{\sqrt{r(1-r)} D_L},$$

equation (54.25) shows that there is a one-to-one correspondence between **CP(Dsgn.Events, Est.HR)**, $\hat{\delta}_L$ and z_L . It is thus reasonable to use any one of these three variables on the X-axis of the Required Events Chart. We have chosen **CP(Dsgn.Events, Est.HR)** because it has a natural interpretation that is easily understood by non-statisticians. The Y-axis, labeled **Required Events** displays the number of events that are required to complete the trial. This number is computed as the minimum of the re-estimated number of events and the cap on the maximum number of events. To be specific, let D_{\max} be the maximum number of events permitted if an adaptation occurs. (This is the entry to the right of the multiplier in cell 1 of the Input Parameters Table.)

Max. # of Events if Adapt (multiplier; total #): Let D_K^* be the solution to the equation

$$\text{Target CP} = 1 - \Phi \left\{ b_K \sqrt{1 + \frac{D_L}{D_K - D_L}} - z_L \sqrt{\frac{D_L}{D_K - D_L}} - \hat{\delta}_L \sqrt{r(1-r)} \sqrt{D_K^* - D_L} \right\}, \quad (54.26)$$

where **Target CP** is the entry in cell 4.

Target CP for Re-estimating # of Events:

Then

$$\text{RequiredEvents} = \min(D_{\max}, D_K^*)$$

We will illustrate with a couple of examples.

Example 1: Suppose the input parameters are as displayed below:

| | | | |
|--|-------------------------------------|---|---------------------------------|
| Adapt at: | <input type="text" value="Look #"/> | <input type="text" value="1"/> | |
| Max. # of Events if Adapt (multiplier; total #): | <input type="text" value="1.5"/> | <input type="text" value="501"/> | |
| Max. Sample Size if Adapt (multiplier; total #): | <input type="text" value="1.5"/> | <input type="text" value="724"/> | |
| Upper Limit on Study Duration: | <input type="text" value="90"/> | | |
| Target CP for Re-estimating # of Events: | <input type="text" value="0.8"/> | | |
| Promising Zone Scale: | | <input type="text" value="Cond. Power"/> | <input type="text" value="CP"/> |
| Promising Zone: | Min. CP: | <input type="text" value="0.45"/> | |
| | Max. CP: | <input type="text" value="0.8"/> | |
| CP Computation Based on: | | <input type="text" value="Estimated HR"/> | |
| Accrual Rate After Adaptation: | | <input type="text" value="No Change"/> | |

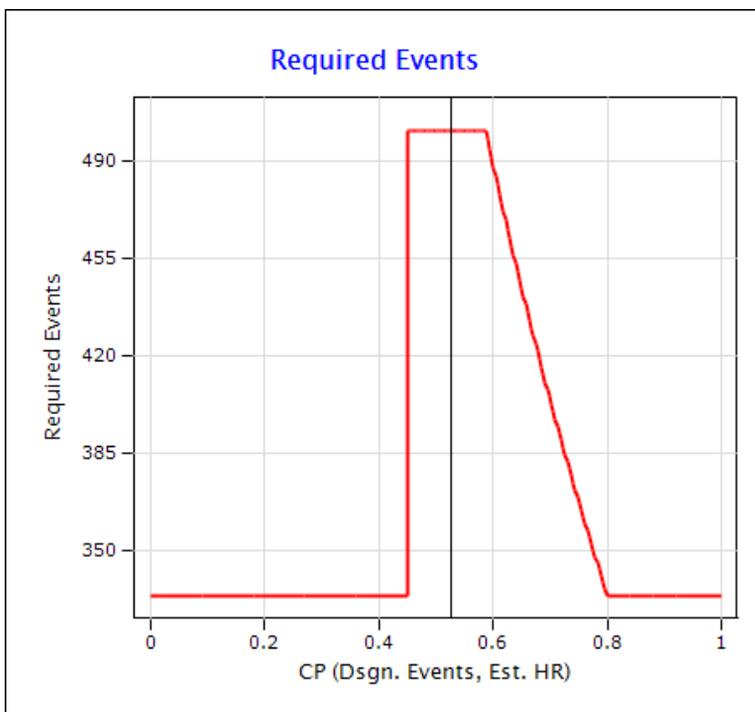
With these inputs the **Required Events** will be re-computed for values of **CP(Dsgn.Events, Est.HR)** that fall in the promising zone, specified by $0.45 \leq \text{CP}(\text{Dsgn.Events}, \text{Est.HR}) < 0.8$. For all values of **CP(Dsgn.Events, Est.HR)** outside this zone, the **Required Events** will remain the unchanged at 334, the number specified at the design stage. Inside the promising zone, however, East will re-estimate D_2^* , the number events that are needed to achieve the target conditional power of 0.8 displayed in cell 4, using equation (54.26). It can be shown that for values of **CP(Dsgn.Events, Est.HR)** on the X-axis between 0.45 and 0.58, the value of D_2^* needed to boost the conditional power to the 0.8 target exceeds 501. Since the cap on the number of events is set at $D_{\max} = 501$, East will set

$$\text{Required Events} = \min(501, D_2^*) = 501$$

in the chart for all $0.45 \leq \text{CP}(\text{Dsgn.Events}, \text{Est.HR}) \leq 0.58$. However, at values of **CP(Dsgn.Events, est.HR)** on the X-axis that exceed 0.59, the re-estimated number

54 The Cui, Hung and Wang Method

of events D_2^* is less than 500, and hence the **Required Events** gradually drops down until it reaches 334 at $CP(Dsgn.Events, Est.HR) = 0.8$. Thereafter the **Required Events** remains constant at 334. Thus the shape of the Required Events Chart is as shown below.



The shape of the Required Events Chart depends on the value of the target conditional power that is one of the inputs. To see this, consider the next example.

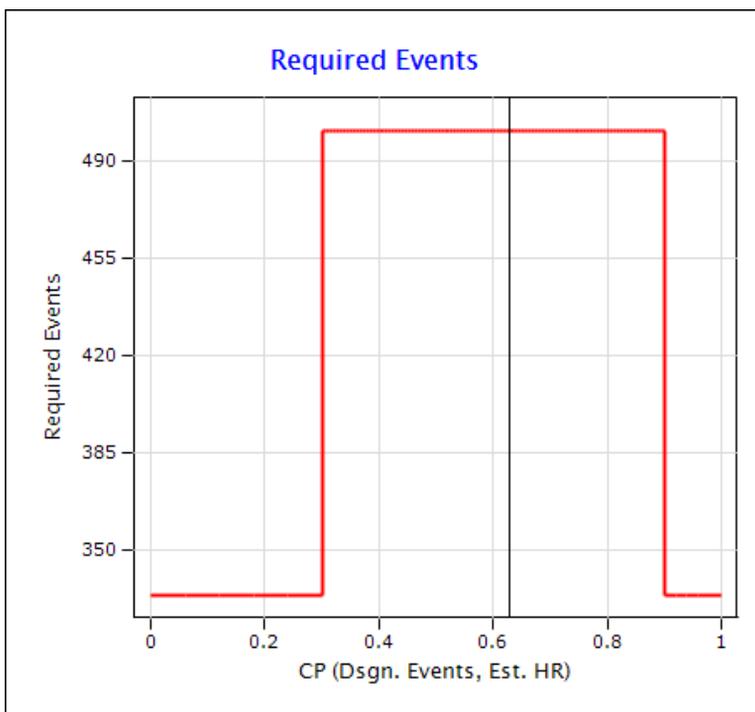
Example 2 : Suppose the input parameters are as displayed below:

| | | | |
|--|-------------------------------------|---|-----------------------------------|
| Adapt at: | <input type="text" value="Look #"/> | <input type="text" value="1"/> | |
| Max. # of Events if Adapt (multiplier; total #): | <input type="text" value="1.5"/> | <input type="text" value="501"/> | |
| Max. Sample Size if Adapt (multiplier; total #): | <input type="text" value="1.5"/> | <input type="text" value="724"/> | |
| Upper Limit on Study Duration: | <input type="text" value="90"/> | | |
| Target CP for Re-estimating # of Events: | <input type="text" value="0.99"/> | | |
| Promising Zone Scale: | | <input type="text" value="Cond. Power"/> | <input type="button" value="CP"/> |
| Promising Zone: | Min. CP: | <input type="text" value="0.3"/> | |
| | Max. CP: | <input type="text" value="0.9"/> | |
| CP Computation Based on: | | <input type="text" value="Estimated HR"/> | |
| Accrual Rate After Adaptation: | | <input type="text" value="No Change"/> | |

This time the promising zone ranges from 0.3 to 0.9. The target conditional power (Shape Parameter) is 0.99. It can be shown that more than 501 events (the cap in cell 1) will be needed to reach this target, for all values of **CP(Dsgn.Events, Est.HR)** in the promising zone. Therefore the Required Events Chart will be a step function taking on values 334 outside the promising zone and taking on values 501 inside the promising

54 The Cui, Hung and Wang Method

zone.



Thus by entering different target conditional power values as input in cell 4 and pressing **Refresh Charts** button, you can experiment with different shapes on the Required Events Chart. The step function shape is favored in many trials both for its simplicity and because it prevents "reverse engineering" the precise value of **CP(Dsgn.Events, est.HR)** by anyone who, for regulatory reasons, has to remain blind to the interim results. For example, suppose it is known that the number of events has increased from 334 to 501. Even then all one can conclude is that **CP(Dsgn.Events, est.HR)** falls between 0.3 and 0.9.

The Conditional Power Chart

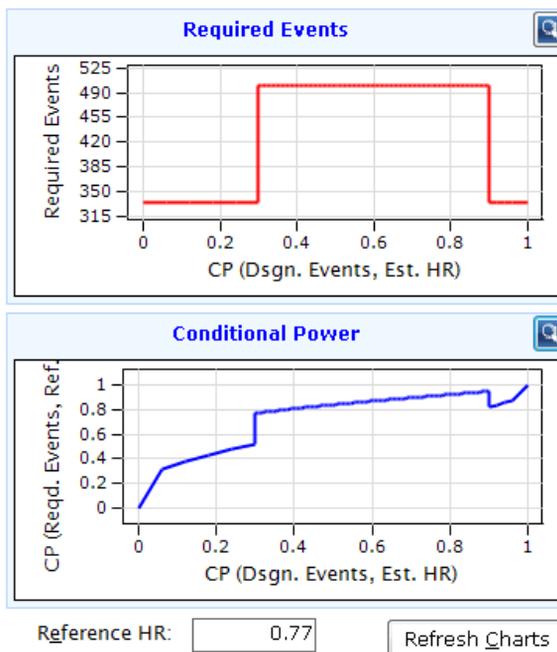
The lower chart on the right side panel of the **Sample Size Re-estimation** tab is called the Conditional Power Chart. As the name suggests, this chart plots the actual conditional power of the study given the observed data at the interim analysis. As was the case for the Required Events Chart, the data at interim analysis results are summarized in terms of **CP(Desgn.Events, Est.HR)** and displayed on the X-axis. The Y-axis, titled **CP(Req.Events, Ref.HR)** then plots the actual conditional power for the reference hazard ratio contained in the edit box below this chart, where **Req.Events** refers to the **Required Events** displayed in the chart above the conditional power chart. Consider again the inputs that were entered into the input parameter table in Example 2.

| | | | |
|--|-------------------------------------|---|-----------------------------------|
| Adapt at: | <input type="text" value="Look #"/> | <input type="text" value="1"/> | |
| Max. # of Events if Adapt (multiplier; total #): | <input type="text" value="1.5"/> | <input type="text" value="501"/> | |
| Max. Sample Size if Adapt (multiplier; total #): | <input type="text" value="1.5"/> | <input type="text" value="724"/> | |
| Upper Limit on Study Duration: | <input type="text" value="90"/> | | |
| Target CP for Re-estimating # of Events: | <input type="text" value="0.99"/> | | |
| Promising Zone Scale: | | <input type="text" value="Cond. Power"/> | <input type="button" value="CP"/> |
| Promising Zone: | Min. CP: | <input type="text" value="0.3"/> | |
| | Max. CP: | <input type="text" value="0.9"/> | |
| CP Computation Based on: | | <input type="text" value="Estimated HR"/> | |
| Accrual Rate After Adaptation: | | <input type="text" value="No Change"/> | |

For these inputs the conditional power chart looks as shown below if the Reference HR

54 The Cui, Hung and Wang Method

is equal to 0.77.



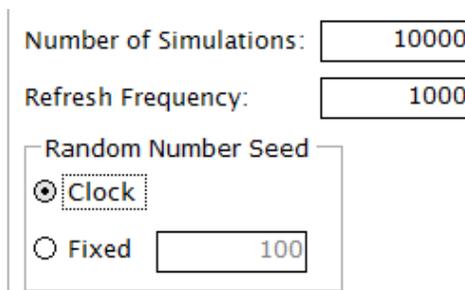
The chart shows that true conditional power gradually climbs from below 20% to about 50% in the *unfavorable zone* ($CP(\text{Dsgn.Events, Est.HR}) < 0.3$). The true conditional power receives a substantial boost in the *promising zone* ($0.3 \leq CP(\text{Dsgn.Events, Est.HR}) < 0.9$), because the **Required Events** jump from 334 to 501 in this zone. Now the conditional power climbs from slightly below 80% to slightly above 90%. There is a slight decline in the true conditional power upon entering the *favorable zone* $CP(\text{Dsgn.Events, Est.HR}) \geq 0.9$, for now the **Required Events** drop back to 334. However in this zone the true conditional power starts out at 82% and rapidly climbs up to well over 90%.

The conditional power chart is useful because it provides a good idea of the type of power one can expect, conditional on falling in the unfavorable, promising and favorable zones, even before any simulations are performed. The simulation results, to be discussed next, provide additional insights.

Table of Simulation Results by Zone

We have already seen in the **Simulation Outputs Without Adaptation** that if the underlying hazard ratio is 0.77 and there is no adaptive change to the number of events then the study only has about 66% power. The power can be improved by increasing the number of events. The traditional approach is to commit up-front to an increase in the number of events. This was the approach used for creating Des 3. We saw in Section 54.5.2 that while Des 3 does indeed have 90% power to detect a hazard ratio of 0.77, it requires a considerably larger up-front commitment of resources; 539 events to be obtained from 823 subjects enrolled over 24 months with 6 additional months of follow-up. A commitment of this magnitude based solely on limited phase 2 data from other trials was not feasible for the sponsor of the current study.

We now consider an alternative approach that has lower overall power than Des 3 under a hazard ratio of 0.77, but might be more acceptable to the sponsor. This is the adaptive approach in which the commitment of resources occurs in two stages with the second stage commitment forthcoming only if the first stage results are in the promising zone. We will evaluate the operating characteristics of this approach by generating 10,000 simulated trials.



Number of Simulations: 10000
Refresh Frequency: 1000
Random Number Seed
 Clock
 Fixed 100

Enter the following Accrual / Dropout Information and Survival Information in the

54 The Cui, Hung and Wang Method

respective tabs of the Simulation Input tabs.

Accrual Info

of Accrual Periods: Input Method:

| Period # | Starting at Time | Accr. Rates |
|----------|------------------|-------------|
| 1 | 0.000 | 20.08 |

Survival Information

of Hazard Pieces Input Method:

Hazard Ratio

| Piece | Starting at Time | Hazard Rates | | Hazard Ratio |
|-------|------------------|--------------|-----------|--------------|
| | | Control | Treatment | |
| 1 | 0.000 | 0.087 | 0.067 | 0.770 |

and the following values in the **Sample Size Re-estimation** tab.

Adapt at:

Max. # of Events if Adapt (multiplier; total #):

Max. Sample Size if Adapt (multiplier; total #):

Upper Limit on Study Duration:

Target CP for Re-estimating # of Events:

Promising Zone Scale:

| | | |
|-----------------|----------|----------------------------------|
| Promising Zone: | Min. CP: | <input type="text" value="0.3"/> |
| | Max. CP: | <input type="text" value="0.9"/> |

CP Computation Based on:

Accrual Rate After Adaptation:

These inputs imply that the data for each of the 10,000 simulated trials will be generated from exponential distributions with a hazard ratio of 0.77, with patients arriving at the rate of 20.08/month, and no drop-outs. At the interim analysis, when 167 events have been observed, there will be an increase of resources **only if** the stage 1 conditional power lies in the promising zone (between 0.3 and 0.9). In that case, the maximum number of events will increase by 50%, from 334 to 501, and the maximum number of subjects will also increase by 50%, from 483 to 724.

To run the simulations, click the **Simulate** button. Save the overall Simulation output in the **Library**. This will get saved as **CHWSim1**. The Table of Simulation Results by Zone gets filled in and is displayed below as Figure 54.1

Figure 54.1: Simulation Results for 10,000 Trials of Des 1 with Adaptation at Look 1

Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Rejecting H1 | | Total Simulations | | Average Sample Size | Average Number of Events | Average Accrual Duration | Average Study Duration |
|------------------------|--------------------------|---------|--------------------------|---------|-------------------|----------|---------------------|--------------------------|--------------------------|------------------------|
| | Count | Row % | Count | Row % | Count | Column % | | | | |
| Futility + Unfavorable | 908 | 31.181% | 2004 | 68.819% | 2912 | 29.120% | 468.259 | 313.24 | 23.271 | 27.851 |
| Promising | 2958 | 86.365% | 467 | 13.635% | 3425 | 34.250% | 724 | 501 | 30.067 | 33.911 |
| Efficacy + Favorable | 3376 | 92.165% | 287 | 7.835% | 3663 | 36.630% | 451.012 | 288.591 | 22.405 | 26.27 |
| All Trials | 7242 | 72.420% | 2758 | 27.580% | 10000 | 100.000% | 549.533 | 368.519 | 25.282 | 29.347 |

Promising Zone defined as 0.3 <= CP < 0.9

This table displays five rows for tracking the outcomes of the 10,000 simulated clinical trials zone by zone, plus a sixth row that combines the results across all five zones. The entries in the table are self-explanatory. For comparison purposes run the simulations again, this time without adaptation. One simple way to do this is to set the two multipliers equal to 1. Edit the Simulation node **CHWSim1** by selecting it and clicking the  icon.

Make the two multipliers equal to 1 as shown below:

Max. # of Events if Adapt (multiplier; total #):

Max. Sample Size if Adapt (multiplier; total #):

The results from 10,000 simulations are re-computed, this time without any adaptation of events or sample size and are displayed below as Figure 54.2.

54 The Cui, Hung and Wang Method

Figure 54.2: Simulation Results for 10,000 Trials of Des 1 without Adaptation at Look 1

☰ Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Rejecting H1 | | Total Simulations | | Average Sample Size | Average Number of Events | Average Accrual Duration | Average Study Duration |
|--------------------------|--------------------------|---------|--------------------------|---------|-------------------|----------|---------------------|--------------------------|--------------------------|------------------------|
| | Count | Row % | Count | Row % | Count | Column % | | | | |
| ⊕ Futility + Unfavorable | 884 | 30.336% | 2030 | 69.664% | 2914 | 29.140% | 467.807 | 312.738 | 23.245 | 27.802 |
| Promising | 2370 | 68.716% | 1079 | 31.284% | 3449 | 34.490% | 483 | 334 | 24.003 | 29.257 |
| ⊕ Efficacy + Favorable | 3388 | 93.154% | 249 | 6.846% | 3637 | 36.370% | 451.547 | 289.323 | 22.428 | 26.314 |
| All Trials | 6642 | 66.420% | 3358 | 33.580% | 10000 | 100.000% | 467.133 | 311.555 | 23.209 | 27.763 |

Promising Zone defined as $0.3 \leq CP < 0.9$

Figure 54.1 displays the simulation based operating characteristics of the Des 1 with the adaptive option enabled while Figure 54.2 displays corresponding operating characteristics of Des 1 with the adaptive option disabled. Although Des 1 was designed under the optimistic assumption that the true hazard ratio is 0.7, both sets of simulations are performed under the pessimistic assumption that the true hazard ratio is 0.77. In order to conveniently compare the operating characteristics of the non-adaptive and adaptive designs, we have combined the relevant data from Figures 54.1 and 54.2 into a single table, Table 54.11.

Table 54.11: Operating Characteristics of Optimistic Design (Powered to Detect HR=0.7) under the Pessimistic Scenario (true HR=0.77)

| 10,000 Simulations Under the Pessimistic Scenario that HR = 0.77 | | | | | | | |
|--|---------|---------|-------|-------------------|-------|---------------|-------|
| Zone | P(Zone) | Power | | Duration (months) | | # of Subjects | |
| | | NonAdpt | Adapt | NonAdpt | Adapt | NonAdpt | Adapt |
| Unf+Fut | 29% | 30% | 31% | 27.8 | 27.8 | 468 | 468 |
| Prom | 34% | 68% | 86% | 29.3 | 33.9 | 483 | 724 |
| Fav+Eff | 37% | 93% | 92% | 26.3 | 26.3 | 452 | 451 |
| Total | — | 66% | 72% | 27.7 | 29.3 | 467 | 550 |

The fourth row of Table 54.11 displays the overall simulation results combined across all zones. The non-adaptive design has 66% power, average study duration of 27.7 months and an average sample size of 467 subjects. In contrast the adaptive design boosts the power by 7 percentage points to 73%, but requires average study duration of 29.34 months and an average sample size of 550 subjects. This is to be expected. If additional study duration and sample size resources are allocated to a trial, its power must increase.

It is more instructive to compare the results Table 54.11 by zone rather than overall. In

this type of comparison it is seen that the adaptive and non-adaptive designs behave identically (up to Monte Carlo accuracy) in the **Unfavorable+Futility** zone as well as in the **Favorable+Efficacy** zone. Both designs end up in the **Unfavorable+Futility** zone about 29% of the time and in that case both designs have similar power of about 30% with identical average study duration, and average number of subjects. Again, both designs end up in the **Favorable+Efficacy** zone about 37% of the time, and in that case they have about 93% power with practically identical average study duration and average number of subjects. In other words the adaptive design produces the same power and consumes the same resources as the conventional design if the interim result falls in either of these two zones. However, 34% of the time both designs end up in the **Promising: $0.3 \leq CP < 0.9$** zone, and where the adaptive design produces about 86% power whereas the non-adaptive design produces only 68% power. To be sure the adaptive design consumes more resources in the promising zone (study duration = 34 months versus 29.3 months; average events = 501 versus 334; average number of subjects = 724 versus 483), but these additional resources are worth spending since they can boost the power by about 20% and might make all the difference between a successful trial and a failure. In summary the adaptive design calls up the additional event and sample size resources only when they are needed and not otherwise.

Although the tables in Figure 54.1 and Figure 54.2 have partitioned the simulation results into three zones there are in fact five zones in the East output. The simulations in the **Unfavorable+Futility** zone are further separated into those simulations that were terminated for futility at the interim analysis and those that were unfavorable but did not cross the futility boundary. Similarly the simulations in the **Favorable+Efficacy** zone are further separated into those that crossed the efficacy boundary at the interim look and those that were favorable but did not cross the efficacy boundary.

The Table of Simulation Results by Zone as seen above for adaptive design:

☰ Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Rejecting H1 | | Total Simulations | | Average Sample Size | Average Number of Events | Average Accrual Duration | Average Study Duration |
|---------------|--------------------------|----------|--------------------------|----------|-------------------|----------|---------------------|--------------------------|--------------------------|------------------------|
| | Count | Row % | Count | Row % | Count | Column % | | | | |
| ☐ Futility | 0 | 0.000% | 362 | 100.000% | 362 | 3.620% | 364.417 | 167 | 18.101 | 18.15 |
| ☐ Unfavorable | 909 | 35.608% | 1642 | 64.392% | 2550 | 25.500% | 483 | 334 | 24.005 | 29.228 |
| ☐ Promising | 2958 | 86.365% | 467 | 13.635% | 3425 | 34.250% | 724 | 501 | 30.067 | 33.911 |
| ☐ Favorable | 2380 | 89.239% | 287 | 10.761% | 2667 | 26.670% | 483 | 334 | 24.004 | 29.233 |
| ☐ Efficacy | 996 | 100.000% | 0 | 0.000% | 996 | 9.960% | 365.356 | 167 | 18.124 | 18.175 |
| ☐ All Trials | 7242 | 72.420% | 2758 | 27.580% | 10000 | 100.000% | 549.533 | 368.519 | 25.282 | 29.347 |

Promising Zone defined as $0.3 \leq CP < 0.9$

Examined in this way, it is seen that of the 2912 simulations entering the

54 *The Cui, Hung and Wang Method*

Unfavorable+Futility zone, only 362 (3.62%) stop early for futility. Of the 3663 simulations entering the **Favorable+Efficiency** zone, only 996 (9.96%) cross the efficacy boundary and stop early.

Table of Zone wise Percentiles

The Table of Simulation Results by Zone reports only the average number of events, sample size, accrual duration and study duration. One can examine the percentiles of the distributions of these statistics from the Table of Zone-Wise Percentiles. Double click the node named **CHWSim2** to see the detailed simulation output for adaptive

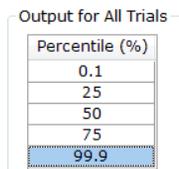
design.

| | Zone | Percentile | Number of Events | Sample Size | Accrual Duration | Study Duration |
|---|----------------|------------|------------------|-------------|------------------|----------------|
| - | Futility | 5.000% | 167 | 343 | 16.93 | 16.947 |
| | | 25.000% | 167 | 355 | 17.629 | 17.669 |
| | | 50.000% | 167 | 364 | 18.088 | 18.156 |
| | | 75.000% | 167 | 372 | 18.546 | 18.594 |
| | | 95.000% | 167 | 388 | 19.227 | 19.256 |
| | Average | 167 | 364.417 | 18.101 | 18.15 | |
| - | Unfavorable | 5.000% | 334 | 483 | 23.91 | 27.856 |
| | | 25.000% | 334 | 483 | 23.985 | 28.609 |
| | | 50.000% | 334 | 483 | 24.018 | 29.226 |
| | | 75.000% | 334 | 483 | 24.039 | 29.804 |
| | | 95.000% | 334 | 483 | 24.051 | 30.661 |
| | Average | 334 | 483 | 24.005 | 29.228 | |
| - | Promising | 5.000% | 501 | 724 | 28.949 | 32.544 |
| | | 25.000% | 501 | 724 | 29.6 | 33.314 |
| | | 50.000% | 501 | 724 | 30.059 | 33.89 |
| | | 75.000% | 501 | 724 | 30.529 | 34.487 |
| | | 95.000% | 501 | 724 | 31.194 | 35.38 |
| | Average | 501 | 724 | 30.067 | 33.911 | |
| - | Favorable | 5.000% | 334 | 483 | 23.901 | 27.849 |
| | | 25.000% | 334 | 483 | 23.985 | 28.652 |
| | | 50.000% | 334 | 483 | 24.02 | 29.287 |
| | | 75.000% | 334 | 483 | 24.04 | 29.899 |
| | | 95.000% | 334 | 483 | 24.051 | 30.853 |
| | Average | 334 | 483 | 24.004 | 29.293 | |
| - | Efficacy | 5.000% | 167 | 343 | 16.958 | 17.032 |
| | | 25.000% | 167 | 356 | 17.634 | 17.689 |
| | | 50.000% | 167 | 366 | 18.122 | 18.19 |
| | | 75.000% | 167 | 374 | 18.589 | 18.657 |
| | | 95.000% | 167 | 386 | 19.273 | 19.292 |
| | Average | 167 | 365.356 | 18.124 | 18.175 | |
| - | All Trials | 5.000% | 167 | 360 | 17.892 | 17.941 |
| | | 25.000% | 334 | 483 | 23.978 | 28.549 |
| | | 50.000% | 334 | 483 | 24.036 | 29.707 |
| | | 75.000% | 501 | 724 | 29.635 | 33.372 |
| | | 95.000% | 501 | 724 | 30.795 | 34.82 |
| | Average | 368.519 | 549.533 | 25.282 | 29.347 | |

By default this table displays the 5th, 25th, 50th, 75th and 95th percentiles of the relevant distributions for all 10,000 trials. For example, the 95th percentile of the **Study Duration** for **all Trials** is displayed as 38.14 months.

54 The Cui, Hung and Wang Method

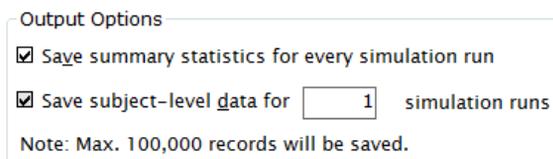
The 95% percentile of the study duration for trials that enter the promising zone, and therefore adapt, is 35.4 months. It might be of interest to know that how short and how long the study duration could be among the simulations that have entered the promising zone. To see this one may edit the **Percentile** column of small table named **Output for all Trials** on **Simulation Control Info** tab



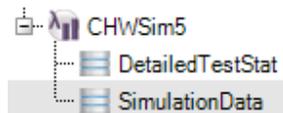
and run the simulations again. Observe the Promising zone table.

| | Zone | Percentile | Number of Events | Sample Size | Accrual Duration | Study Duration |
|---|-----------|------------|------------------|-------------|------------------|----------------|
| ⊖ | Promising | 0.100% | 501 | 724 | 27.855 | 31.172 |
| | | 25.000% | 501 | 724 | 29.583 | 33.343 |
| | | 50.000% | 501 | 724 | 30.06 | 33.925 |
| | | 75.000% | 501 | 724 | 30.537 | 34.561 |
| | | 99.900% | 501 | 724 | 32.243 | 36.942 |
| | | Average | 501 | 724 | 30.065 | 33.953 |

The 0.1 percentile of the study duration is 32 months while the 99.9 percentile is 37 months. East also provides the capability to store the summary statistics for every simulation run and the subject level data. This is achieved by checking off the following checkboxes



on the **Simulation Control Info** tab. When we keep this simulation output in the **Library**, two more nodes get saved under the simulation node as shown below:



Simulating Multiple Scenarios

The simulations that were performed in the earlier section were based on a multiplier of 1.5 for the maximum number of subjects, if an adaptation were to occur (cell 3 in the Table of Input Parameters). The choice of 1.5 was arbitrary. It is possible that a smaller multiplier might result in almost the same average study duration, and hence produce a more efficient design from the sponsor’s perspective. It would therefore be desirable to conduct several simulation experiments with different multipliers for the number of subjects. It is possible to conduct such multiple experiments from the **Sample Size Re-estimation** tab.

Edit the Simulations and click on **Sample Size Re-estimation** tab.

| | | | |
|--|-------------------------------------|---|-----------------------------------|
| Adapt at: | <input type="text" value="Look #"/> | <input type="text" value="1"/> | |
| Max. # of Events if Adapt (multiplier; total #): | <input type="text" value="1.5"/> | <input type="text" value="501"/> | |
| Max. Sample Size if Adapt (multiplier; total #): | <input type="text" value="1.5"/> | <input type="text" value="724"/> | |
| Upper Limit on Study Duration: | <input type="text" value="90"/> | | |
| Target CP for Re-estimating # of Events: | <input type="text" value="0.99"/> | | |
| Promising Zone Scale: | | <input type="text" value="Cond. Power"/> | <input type="button" value="CP"/> |
| Promising Zone: | Min. CP: | <input type="text" value="0.3"/> | |
| | Max. CP: | <input type="text" value="0.9"/> | |
| CP Computation Based on: | | <input type="text" value="Estimated HR"/> | |
| Accrual Rate After Adaptation: | | <input type="text" value="No Change"/> | |

The inputs on this tab can be used to conduct simulation experiments over a range of multipliers for the maximum number of subjects, while keeping the multiplier for an adaptive increase in the number of events constant. The magnitude of the multiplier applied to the maximum number of subjects does not affect the power of the study but it does have a direct impact on the study duration. It is thus preferable to experiment with a range of multipliers so as to gain a better understanding of the relationship between maximum number of subjects and study duration in an adaptive design.

Suppose we wish to conduct simulation experiments over a range of sample sizes, with the **Max. Events if Adapt** multiplier fixed at 1.5.

54 The Cui, Hung and Wang Method

We may enter a range of multipliers for sample size into the field for **Max. Sample size if adapt** using the convention $x : y : z$ to denote entries ranging from x to y in steps of size z . Let us enter multiplier values for sample size ranging from 1.25 to 1.9 in steps of size 0.05. The complete input table will look like:

Adapt at:

Max. # of Events if Adapt (multiplier; total #):

Max. Sample Size if Adapt (multiplier; total #):

Upper Limit on Study Duration:

Target CP for Re-estimating # of Events:

Promising Zone Scale:

Promising Zone: Min. CP:

Max. CP:

CP Computation Based on:

Accrual Rate After Adaptation:

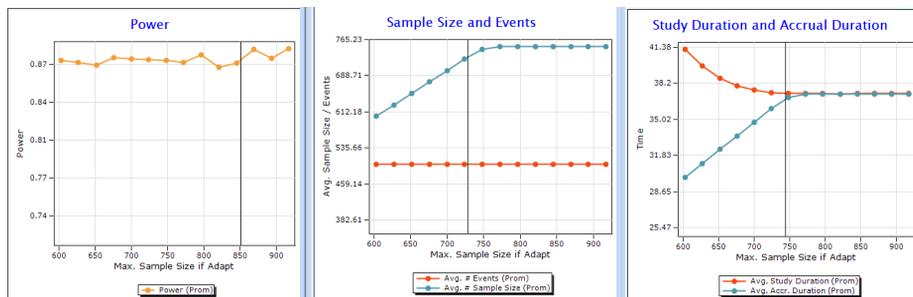
Upon pressing the **Simulate** button, all the scenarios are simulated and can be seen in the **Output Preview** pane.

| ID | Power | Power (Promising) | Method of Adaptation | Adaptation Stage | Max. # of Events if Adapt | Max. Sample Size if Adapt | Target CP | CP Computation Based on | Promising Zone | Average Study Duration | Average Sample Size | Average Events |
|----------|-------|-------------------|----------------------|------------------|---------------------------|---------------------------|-----------|-------------------------|-----------------|------------------------|---------------------|----------------|
| CHWSim5 | 0.722 | 0.873 | Cui-Hung-Wang | Look #: 1 | 501 | 603 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 41,203 | 603 | 501 |
| CHWSim6 | 0.723 | 0.871 | Cui-Hung-Wang | Look #: 1 | 501 | 627 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 39,75 | 627 | 501 |
| CHWSim7 | 0.725 | 0.869 | Cui-Hung-Wang | Look #: 1 | 501 | 652 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 38,674 | 652 | 501 |
| CHWSim8 | 0.731 | 0.875 | Cui-Hung-Wang | Look #: 1 | 501 | 676 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 38 | 676 | 501 |
| CHWSim9 | 0.726 | 0.874 | Cui-Hung-Wang | Look #: 1 | 501 | 700 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 37,615 | 700 | 501 |
| CHWSim10 | 0.728 | 0.874 | Cui-Hung-Wang | Look #: 1 | 501 | 724 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 37,406 | 723,942 | 501 |
| CHWSim11 | 0.728 | 0.873 | Cui-Hung-Wang | Look #: 1 | 501 | 748 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 37,295 | 744,188 | 501 |
| CHWSim12 | 0.722 | 0.871 | Cui-Hung-Wang | Look #: 1 | 501 | 772 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 37,307 | 750,31 | 501 |
| CHWSim13 | 0.726 | 0.878 | Cui-Hung-Wang | Look #: 1 | 501 | 796 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 37,296 | 750,136 | 501 |
| CHWSim14 | 0.722 | 0.867 | Cui-Hung-Wang | Look #: 1 | 501 | 821 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 37,28 | 750,159 | 501 |
| CHWSim15 | 0.728 | 0.87 | Cui-Hung-Wang | Look #: 1 | 501 | 845 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 37,298 | 750,515 | 501 |
| CHWSim16 | 0.731 | 0.883 | Cui-Hung-Wang | Look #: 1 | 501 | 869 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 37,299 | 750,366 | 501 |
| CHWSim17 | 0.722 | 0.875 | Cui-Hung-Wang | Look #: 1 | 501 | 893 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 37,315 | 750,651 | 501 |
| CHWSim18 | 0.731 | 0.883 | Cui-Hung-Wang | Look #: 1 | 501 | 917 | 0.99 | Estimated HR | 0.3 <= CP < 0.9 | 37,315 | 750,683 | 501 |

The above simulation output displays results for all 10,000 simulated trials. The column **Power** contains the overall power based on 10,000 simulated trials. It might be of greater interest to examine the results only for those trials that entered the promising zone and hence were adapted. The above simulation output also has some columns which correspond to the promising zone. These columns are **Power**

(Promising), Average Study Duration, Average Sample Size and Average Events.

These same simulation results for promising zone are also displayed graphically on the three charts shown below.



We have performed ten simulation runs with the **Maximum Number of Subjects if Adapt** input parameter ranging from 603 to 821. Let us analyze these outputs.

Power The **Power (Promising)** column show a relatively constant power of about 87% for the entire range of proposed values for **Maximum Number of Subjects if Adapt**. This is what one would expect in an event driven trial. The mild fluctuation in power that are observed are due to Monte Carlo sampling error.

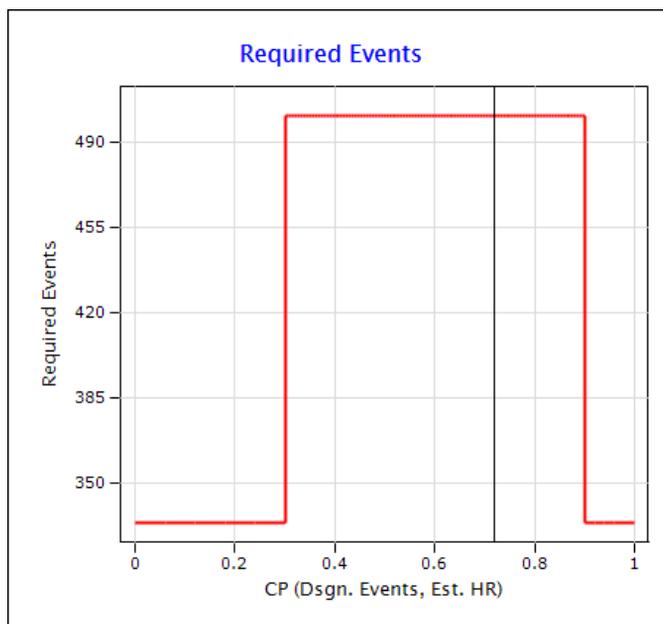
Average Number of Events The required number of events for trials that enter the promising zone is determined by the **Target CP**.

Target CP for Re-estimating # of Events:

Since the value of this parameter has been set to 0.99, the Required Events Chart displayed in the **Sample Size Re-estimation** tab is a step function with the

54 The Cui, Hung and Wang Method

constant value 501 for all value of conditional power in the promising zone.



Consequently the **Average Number of Events** for all trials in the promising zone is 501 regardless of the value of the **Maximum Number of Subjects if Adapt** parameter.

Average Sample Size Observe from the table of simulation results that the numbers in the **Maximum Sample Size if Adapt** column and the **Average Sample Size** are close to one another between 604 and 748. Thereafter the **Average Sample Size** level off to a constant value of 750 even though the **Maximum Sample Size if Adapt** continue to grow. The same behavior is evident in the Number of Subjects Chart which displays a 45 degree line for values between 603 and 748 on the X-axis and a horizontal line thereafter. This is so because the time that it takes for the 748 subjects to be enrolled is about 37 months and that is about the same as the average time that it takes for the required 501 events to arrive. Once 501 events have arrived, additional enrollment stops. Thus values on the Y-axis of the Number of Subjects Chart do not change after an average enrollment of 748 subjects.

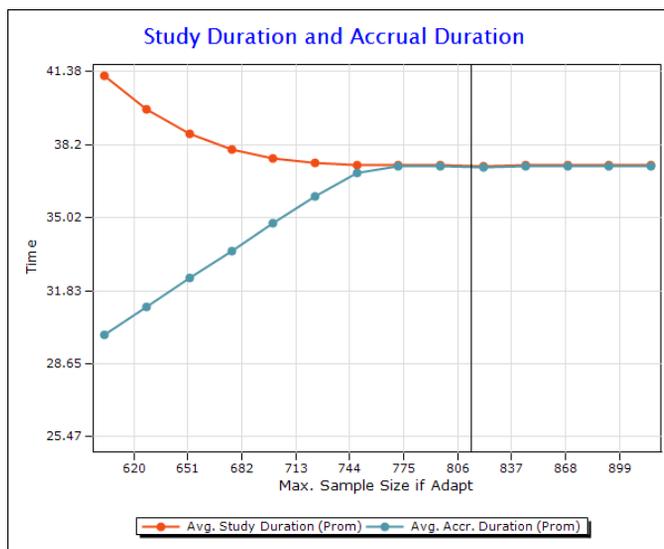
Average Study Duration As the magnitude of the **Maximum Number of**

Subjects if Adapt parameter increases, the **Average Study Duration** decreases. This is so because with increased accrual the required 501 events arrive earlier. Notice, however, from the Study Duration and Accrual Duration Chart as well as from the tabulated values in the **Average Study Duration** column that the rate of decrease in the average study duration continues to decline until it gradually comes to a halt at a value between 724 and 748 for the **Maximum Number of Subjects if Adapt** value on the X-axis. Thereafter the **Average Study Duration** value remains constant at 37 months even though the **Maximum Number of Subjects if Adapt** value continues to increase. This is so because on average by the time about 748 subjects have enrolled, the required 501 events will have arrived and the trial will be terminated.

Average Accrual Duration As the magnitude of the **Maximum Number of Subjects if Adapt** parameter increases, the **Average Accrual Duration** increase as well since more subject are being enrolled while the rate of accrual is constant. However, as seen from the Study Duration and Accrual Duration Chart, the rate of increase in **Average Accrual Duration** continues to decline until it comes to a halt at a value close to 748 for the **Maximum Number of Subjects if Adapt** value on the X-axis. Thereafter the **Average Accrual Duration** value remains constant at about 37 months even though the **Maximum Number of Subjects if Adapt** value continues to increase. This is so because on average by the time about 748 subjects have enrolled, the required 501 events will have arrived and further enrollment will be halted. Indeed, as can be seen on the Study Duration and Accrual Duration Chart, the graphs of **Average Study Duration** and **Average Accrual Duration** begin to converge and meet at a value of

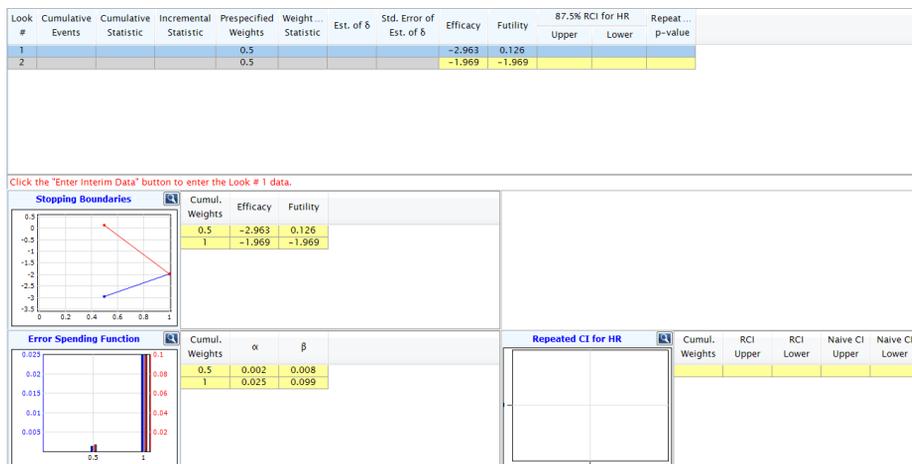
54 The Cui, Hung and Wang Method

about 772 subjects on the X-axis and 37 months on the Y-axis approximately.



54.5.4 Interim Monitoring

Now we will discuss the CHW interim monitoring procedure taking the example of Des 1. Select Des 1 in the **Library** and click on the icon  to create a CHW Interim Monitoring Dashboard for this design as shown below.



This dashboard differs from the usual interim monitoring dashboard for a classical group sequential trial in the following major ways: The Pre-specified Nominal Critical Points (stopping boundaries) are written into dashboard as soon as it is created, and are non-editable. Incremental Statistic value is derived at each look from Cumulative Events and Cumulative Statistic values of that look and the previous look, except at the first look, the Incremental Statistic value remains same as the Cumulative Statistic value. The weighted statistic is obtained by combining the incremental test statistics using Pre-specified Weights. In actual trial, the cumulative events at each look need not correspond to what was originally specified at the design stage. But if the cumulative events that correspond to the original study design are entered, then the weighted statistic is the same as the usual Wald statistic employed in conventional (non-adaptive) interim monitoring.

The values of cumulative test statistics $Z_{j,cum}^*$ at the interim look j are calculated by clicking on the **Enter Interim Data** button. This calculator uses as an input the estimates of the treatment effect $\hat{\delta}_j$ and estimated value of the standard error of $\hat{\delta}_j$. These values may be obtained by fitting a Cox proportional hazard model to the dataset available at look j or by calculating the Z-score based on the log-rank test statistics

54 The Cui, Hung and Wang Method

$Z_{j,LR}^*$ and using the following (approximating) expressions

$$\hat{\delta}_j^* = \frac{Z_{j,LR}^*}{\sqrt{r(1-r)D_j^*}} \quad (54.27)$$

$$Var(\hat{\delta}_j^*) = \frac{1}{r(1-r)D_j^*} \quad (54.28)$$

Here r is the proportion of subjects randomized to the active treatment group and D_j^* is the number of events observed at the look j .

Example: IM Inputs taken from the results of Cox proportional hazards model

Suppose the first look is taken as planned after an accrual of 167 events. Suppose we observe $\hat{\delta} = -0.288$ and a standard error of 0.236. The cumulative statistic at the first look is thus $(-0.288/0.236) = -1.220$. We enter these quantities into the Test statistic calculator as shown below. On pressing OK, the IM dashboard is updated with the first look computation.

Test Statistic Calculator

Editing Look # 1 For Cumulative Events

Events:

Input for Survival end point

Estimate of δ :
 $\delta = \ln(\lambda_t / \lambda_c)$

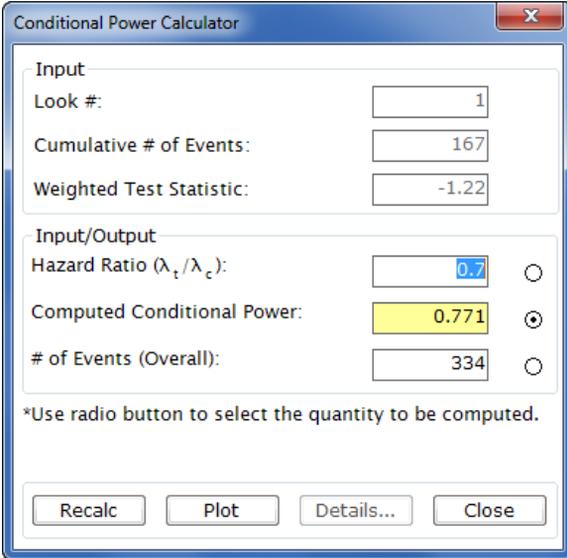
Standard Error of Estimate of δ :

Output

Weighted Statistic:
 Cumulative Statistic:

Recalc OK Cancel

Since the nominal critical value for early stopping is -2.963, the trial continues. We now need to decide on the sample size to use for the second and final look. We invoke the conditional power calculator to assist with this decision.



The screenshot shows a dialog box titled "Conditional Power Calculator". It has a blue title bar with a close button (X) in the top right corner. The dialog is divided into two main sections: "Input" and "Input/Output".

Input Section:

- Look #:
- Cumulative # of Events:
- Weighted Test Statistic:

Input/Output Section:

- Hazard Ratio (λ_t/λ_c):
- Computed Conditional Power:
- # of Events (Overall):

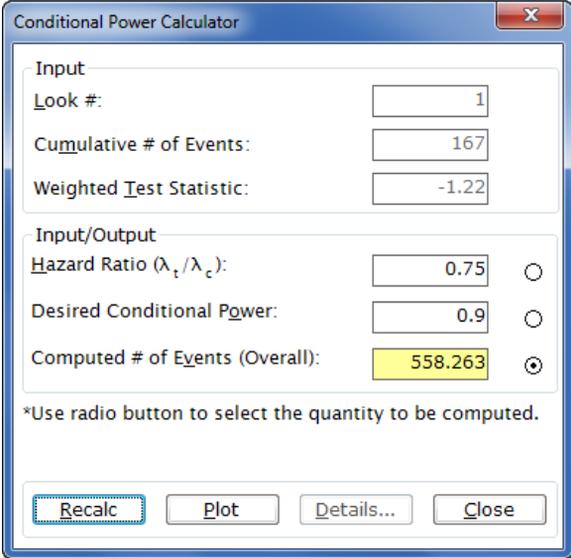
*Use radio button to select the quantity to be computed.

At the bottom of the dialog, there are four buttons: "Recalc", "Plot", "Details...", and "Close".

Suppose we specify to the calculator that we wish to obtain 90% conditional power to detect HR=0.75. Upon entering these terms into the calculator we obtain a final

54 The Cui, Hung and Wang Method

(overall) tally of 558.263 events.



The image shows a software dialog box titled "Conditional Power Calculator". It contains several input fields and radio buttons. The "Input" section has three fields: "Look #:" with value 1, "Cumulative # of Events:" with value 167, and "Weighted Test Statistic:" with value -1.22. The "Input/Output" section has three fields: "Hazard Ratio (λ_t/λ_c):" with value 0.75, "Desired Conditional Power:" with value 0.9, and "Computed # of Events (Overall):" with value 558.263. There are three radio buttons to the right of the Hazard Ratio, Desired Conditional Power, and Computed # of Events fields. A note below the radio buttons says "*Use radio button to select the quantity to be computed." At the bottom, there are four buttons: "Recalc", "Plot", "Details...", and "Close".

| Field | Value |
|---|---------|
| Look #: | 1 |
| Cumulative # of Events: | 167 |
| Weighted Test Statistic: | -1.22 |
| Hazard Ratio (λ_t/λ_c): | 0.75 |
| Desired Conditional Power: | 0.9 |
| Computed # of Events (Overall): | 558.263 |

Based on the guidance provided by the calculator, suppose we decide to continue the trial to observe 560 events by suitably increasing the sample size and the study duration.

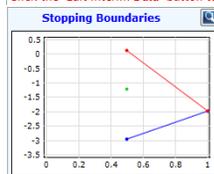
Suppose that, based on these 560 events, the estimate of delta is -0.272 corresponding to a HR value of 0.762 and the estimate of standard error of $\hat{\delta}$ as 0.135, leading to

cumulative test statistic of -2.015.

Upon pressing the OK button the cumulative statistic is entered into the interim monitoring dashboard, the incremental statistic and the weighted statistic are computed as -1.605 and -1.998 respectively. Since the weighted statistic exceeds the nominal critical value, the null hypothesis is rejected. The repeated confidence interval for HR is (0.697,0.996) and the repeated p-value is 0.023. These estimates are based on the methods described in Section 54.1 and are appropriately adjusted to preserve their validity in the face of adaptive sample size changes.

| Look # | Cumulative Events | Cumulative Statistic | Incremental Statistic | Prespecified Weights | Weight... Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | Futility | 87.5% RCI for HR | | Repeat... p-value |
|--------|-------------------|----------------------|-----------------------|----------------------|---------------------|-----------|-------------------------|----------|----------|------------------|-------|-------------------|
| | | | | | | | | | | Upper | Lower | |
| 1 | 167 | -1.22 | -1.22 | 0.5 | -1.22 | -0.288 | 0.236 | -2.963 | 0.126 | 1.509 | 0.509 | 0.26 |
| 2 | 560 | -2.015 | -1.605 | 0.5 | -1.998 | -0.272 | 0.135 | -1.969 | -1.969 | 0.996 | 0.697 | 0.023 |

Click the "Edit Interim Data" button to edit the Look # 2 data.



| Cumul. Weights | Efficacy | Futility |
|----------------|----------|----------|
| 0.5 | -2.963 | 0.126 |
| 1 | -1.969 | -1.969 |

54 The Cui, Hung and Wang Method

The above computations in the IM sheet were carried out using the formulas specified in section 54.2 as detailed below.

At the first look,

$$\hat{\delta}_1^* = -0.288, SE(\hat{\delta}_1^*) = 0.236, I_1^* = \frac{1}{[SE(\hat{\delta}_1^*)]^2} = 17.955, Z_{1,cum}^* = \frac{\hat{\delta}_1^*}{SE(\hat{\delta}_1^*)} = -1.22$$

By definition, for the first look, the incremental statistic and the weighted statistic are $Z^{*(1)} = Z_{1,CHW}^* = Z_{1,cum}^* = -1.22$

At the second look,

$$\hat{\delta}_2^* = -0.272, SE(\hat{\delta}_2^*) = 0.135, I_2^* = \frac{1}{[SE(\hat{\delta}_2^*)]^2} = 54.870, Z_{2,cum}^* = \frac{\hat{\delta}_2^*}{SE(\hat{\delta}_2^*)} = -2.015$$

The incremental statistic at the second look is

$$Z^{*(2)} = \frac{\sqrt{I_2^*} Z_{2,cum}^* - \sqrt{I_1^*} Z_{1,cum}^*}{\sqrt{I_2^* - I_1^*}} = -1.605$$

The weighted statistic

$$Z_{2,CHW}^* = \frac{\sqrt{w^{(1)}} Z^{*(1)} + \sqrt{w^{(2)}} Z^{*(2)}}{\sqrt{w^{(1)} + w^{(2)}}} = \frac{\sqrt{0.5}(-1.22) + \sqrt{0.5}(-1.605)}{\sqrt{0.5 + 0.5}} = -1.998$$

Example: IM Inputs taken from the results of Logrank test

Suppose the first look is taken after an accrual of 160 events. Further we apply Logrank test to the data, and obtain the value of χ_{1df}^2 to be 1.456 or equivalently $Z_1^* = \sqrt{1.456} = 1.2066$. The cumulative statistic at the first look is thus 1.2066. We will first estimate $\hat{\delta}$ and $SE(\hat{\delta})$ using the approximation formulas 54.27 and 54.28 and then use the test statistic calculator to post these values. Thus using the formulas,

$$\delta_1^* = \frac{Z_1^*}{\sqrt{r(1-r)D_1^*}} = \frac{1.2066}{\sqrt{0.5(1-0.5)160}} = \frac{1.2066}{6.3246} = 0.1908;$$

$$Var(\delta_1^*) = \frac{1}{r(1-r)D_1^*} = \frac{1}{0.5(1-0.5)160} = 0.025, SE(\delta_1^*) = \sqrt{0.025} = 0.1581.$$

Another way to estimate δ_1^* is

$$\delta_1^* = (Z_1^*)(SE(\delta_1^*)) = (1.2066)(0.1581) = 0.1908.$$

Now bring up CHW-IM dashboard, select the first look row and click on the **Enter Interim Data** button to input the look-wise information. Enter Cumulative Events as 160. Enter the value of $\hat{\delta}$ as -0.1908, the value of $SE(\hat{\delta})$ as 0.1581 and click on **Recalc** and then on **OK**. The values in the IM sheet for the first look will appear as shown below.

| Look # | Cumulative Events | Cumulative Statistic | Incremental Statistic | Prespecified Weights | Weight... Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | Futility | 87.5% RCI for HR | | Repeat ... p-value |
|--------|-------------------|----------------------|-----------------------|----------------------|---------------------|------------------|--------------------------------|----------|----------|------------------|-------|--------------------|
| | | | | | | | | | | Upper | Lower | |
| 1 | 160 | -1.207 | -1.207 | 0.5 | -1.207 | -0.191 | 0.158 | -2.963 | 0.126 | 1.32 | 0.567 | 0.263 |
| 2 | | | | 0.5 | | | | -1.969 | -1.969 | | | |

The values of cumulative, incremental and weighted statistics are all same as -1.207. Since the nominal critical value for early stopping is -2.963, the trial continues. We

now need to decide on the sample size to use for the second and final look. We invoke the conditional power calculator to assist with this decision.

The screenshot shows a dialog box titled "Conditional Power Calculator" with a close button (X) in the top right corner. The dialog is divided into two main sections: "Input" and "Input/Output".

Input Section:

- Look #: 1
- Cumulative # of Events: 160
- Weighted Test Statistic: -1.207

Input/Output Section:

- Hazard Ratio (λ_t/λ_c): 0.7 (radio button is unselected)
- Computed Conditional Power: 0.781 (radio button is selected, and the value is highlighted in yellow)
- # of Events (Overall): 334 (radio button is unselected)

Below the input/output section, there is a note: "*Use radio button to select the quantity to be computed."

At the bottom of the dialog, there are four buttons: "Recalc", "Plot", "Details...", and "Close".

Suppose we specify to the calculator that we wish to obtain 90% conditional power to detect HR=0.75. Upon entering these terms into the calculator we obtain a final

54 The Cui, Hung and Wang Method

(overall) tally of 555.0 events.

Based on the guidance provided by the calculator, suppose we decide to continue the trial to accrue 560 events by suitably increasing the number of subjects and the study duration. Suppose that, based on these 560 events, the estimate of Z_2^* from Logrank test is -2.135. Now as in the first look, we can estimate $SE(\hat{\delta})$ and $\hat{\delta}$ using the formulas 54.27 and 54.28.

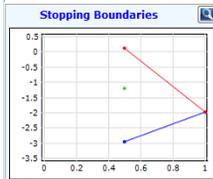
These estimates work out as

$SE(\hat{\delta}_2^*) = 0.0845$, the default value that appears in the test statistic calculator and $\hat{\delta}_2^* = (SE(\hat{\delta}_2^*))(Z_2^*) = (0.0845)(-2.135) = -0.1804$. Enter these values in the CHW IM dashboard. Now the cumulative statistic is entered into the interim monitoring dashboard, the incremental statistic and the weighted statistic are computed as -1.7624 and -2.0995 respectively. Since the weighted statistic exceeds the nominal critical value, the null hypothesis is rejected. The repeated confidence interval for HR is (0.6921, 0.9887) and the repeated p-value is 0.0182. These estimates are based on the methods described in Section 54.1 and are appropriately adjusted to preserve their

validity in the face of adaptive sample size changes.

| Look # | Cumulative Events | Cumulative Statistic | Incremental Statistic | Prespecified Weights | Weight... Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | Futility | 87.5% RCI for HR | | Repeat... p-value |
|--------|-------------------|----------------------|-----------------------|----------------------|---------------------|------------------|--------------------------------|----------|----------|------------------|-------|-------------------|
| | | | | | | | | | | Upper | Lower | |
| 1 | 160 | -1.207 | -1.207 | 0.5 | -1.207 | -0.191 | 0.158 | -2.963 | 0.126 | 1.32 | 0.567 | 0.263 |
| 2 | 560 | -2.13 | -1.757 | 0.5 | -2.096 | -0.18 | 0.085 | -1.969 | -1.969 | 0.989 | 0.692 | 0.018 |

Click the "Edit Interim Data" button to edit the Look # 2 data.



| Cumul. Weights | Efficacy | Futility |
|----------------|----------|----------|
| 0.5 | -2.963 | 0.126 |
| 1 | -1.969 | -1.969 |

East 6

Since the value of weighted statistic is \leq the critical point for efficacy, H_0 is rejected.

OK

55

The Chen, DeMets and Lan Method

Two objections are sometimes leveled at the CHW method discussed in Chapter 54. They both relate to the use of the CHW statistic (54.2) instead of the classical Wald statistic (54.3) or its variant (54.4) for performing the hypothesis tests. Specifically, it is felt by some statisticians that the incremental Wald statistics $(Z^{*(1)}, Z^{*(2)}, \dots, Z^{*(K)})$ generated at the K stages should be combined by utilizing weights derived from the actual sample sizes $(n_1^*, n_2^*, \dots, n_K^*)$ at each stage rather than by weights that depend on the pre-specified sample sizes (n_1, n_2, \dots, n_K) . There is a concern that if the actual number of subjects entering the trial differs from the number pre-specified at the start of the trial, then the use of pre-specified weights will distort the scientific contribution of each cohort entering the trial. This is a philosophical rather than statistical objection, since the use of pre-specified weights controls the type-1 error in the presence of sample size changes, whereas the use of actual weights, in general, does not. It has, however, led to some interesting theoretical research on the loss of efficiency resulting from use of the CHW statistic. (See, for example, Tsiatis and Mehta, 2003; Jennison and Turnbull, 2006). In practice, the magnitude of the adaptive sample size increase is seldom greater than two-fold and within this limit, the loss of efficiency is rather small. Indeed some of the EastAdapt tools described in the present chapter will show that in most practical settings, the loss of efficiency is negligible.

This chapter discusses a method proposed by Chen, DeMets and Lan (2004) (the CDL method) for making sample size modifications to an ongoing trial and then performing the interim monitoring and final analysis with the classical Wald statistic rather than the weighted CHW statistic. The method is further extended to a more general setting by Gao, Ware and Mehta (2008) (the extended CDL method). The main limitation of these two methods is that they are only applicable if the sample size is altered at the penultimate stage of a K -stage group sequential trial. Thus, for simplicity, we will illustrate the methods for two-stage trials only. Furthermore, in the current implementation of East, they are only applicable if the sample size is increased adaptively, but not if it is decreased.

This chapter pre-supposes familiarity with the CHW method and examples presented in Chapter 54. The same three designs, normal (schizophrenia example), binomial (acute coronary syndromes example) and survival (lung cancer example), that were used to illustrate the CHW method in Chapter 54 will be re-visited in the present chapter. Thus some of the steps used to construct these designs in East may be skipped since they will have already been presented in Chapter 54.

55.1 The CDL Method

55.1.1 Normal Endpoint

55.1.2 Binomial Endpoint

55.1.3 Survival Endpoint

Consider a two-sided level- α test of the null hypothesis

$$H_0: \delta = 0$$

versus the two-sided alternative hypothesis

$$H_1: \delta \neq 0$$

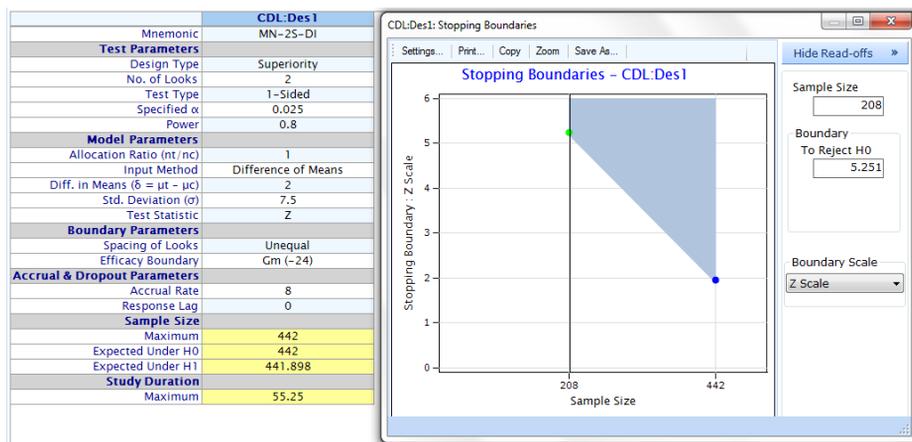
for a two-arm randomized clinical trial. We assume that the null hypothesis will be tested by a two-look group sequential trial with cumulative sample sizes (n_1, n_2) and stopping boundaries (b_1, b_2) derived from some level- α spending function. The data will be examined at the end of look 1 and the sample size for the remainder of the trial may then be changed. Ordinarily if the sample size is changed in a data dependent manner in the middle of a trial, we would be obliged to use the CHW weighted statistic (54.2) described in Chapter 54 instead of the conventional Wald statistic (54.3) for the final analysis, in order to preserve the type-1 error. Intuitively, however, it could be argued that if under the null hypothesis the interim value of the test statistic is large, then it would stand a better chance of regressing to the mean if the sample size of the second stage was increased. Therefore a sample size increase would make it more difficult to achieve statistical significance at the final analysis. Chen, DeMets and Lan (2006) have formalized this intuition by demonstrating mathematically that if the conditional power at the interim look, evaluated at the **estimated** value $\hat{\delta}$ obtained at the interim analysis, is at **least 50%**, one can increase the sample size for the remainder of the trial and still use the conventional Wald statistic for the final analysis, **and the type-1 error won't be inflated thereby**. This important result makes it possible to design two-stage adaptive trials in which the sample size may be increased in a data dependent manner at the interim look, but all the conventional methods of obtaining p-values, confidence intervals and point estimates, available in standard software packages, are applicable at the time of the final analysis.

The above CDL result applies only to a sample size increase and not to a sample size decrease. In order to use the conventional statistic under a sample size decrease the reverse condition must hold. That is, if the conditional power is **no greater than 50%** at the interim look, the sample size can be decreased and the conventional Wald statistic can be used for the final analysis without inflating the type-1 error. However, the discussion in this chapter focuses on sample size increases only. This is entirely in keeping with the recommendations in the FDA Guidance on Adaptive Design (2010) where the use of adaptive methods to decrease sample size is discouraged.

55 The Chen, DeMets and Lan Method

55.1.1 Normal Endpoint: Schizophrenia Trial

We will apply the CDL method to the Schizophrenia trial discussed in detail in Chapter 54. The starting point is a two-look design enrolling 442 subjects, with an interim look planned after obtaining data on 208 completers. The trial is designed to test the null hypothesis $\delta = 0$ versus the one-sided alternative that $\delta > 0$. The standard deviation is assumed to be $\sigma = 7.5$. As the only purpose of the interim analysis is to re-estimate the sample size, but not to stop early, we use the conservative $\gamma(-24)$ spending function (Hwang, Shih and DeCani, 1990) to obtain the efficacy stopping boundary for the interim look. Thereby the amount of type-1 error spent at the interim look is negligible and practically the entire $\alpha = 0.025$ is available for the final analysis. With these specification the trial has just over 80% power to detect $\delta = 2$.



As pointed out in Chapter 54, the true value of δ which might actually be less than 2. It is thus possible that this trial is underpowered at a sample size of 442. We can, however, examine the data at the interim look and estimate the conditional power, and increase the sample size if the conditional power falls in a promising zone. The approach is identical to that discussed in Chapter 54 for the CHW design. We partition the sample space into following zones - futility, unfavorable, promising, favorable and efficacy, based on the conditional power attained at the interim look. The sample size may then be increased if the interim results fall inside the promising zone, thereby recovering the lost power. The additional feature of the CDL design is, however, that if conditional power at the interim look is at least 50% it is not necessary to use the CHW statistic at the final analysis. The conventional Wald statistic may then be used without inflating the type-1 error. We shall study the operating characteristics of the above CDL design through simulation. The option for choosing CDL method for adaptation

is on the **Sample Size Re-estimation** tab. We call the CDL simulations by clicking on the radio-button for CDL, the resulting simulation input window will appear as shown below.

| | |
|--|--|
| Use Adaptation Method | |
| <input type="radio"/> CHW | <input checked="" type="radio"/> CDL |
| <input type="radio"/> Müller and Schäfer | |
| Adapt at: | Look # <input type="text" value="1"/> |
| Max. Sample Size if Adapt (multiplier; total #): | <input type="text" value="1"/> <input type="text" value="442"/> |
| Upper Limit on Study Duration: | <input type="text" value="165.75"/> |
| Target CP for Re-estimating Sample Size: | <input type="text" value="0.8"/> |
| Use Wald Stat. if CP >= | <input type="text" value="0.5"/> |
| Promising Zone Scale: | Cond. Power <input type="text" value="CP"/> |
| Promising Zone: | Min. CP: <input type="text" value="0.3"/> Max. CP: <input type="text" value="0.8"/> |
| CP Computation Based on: | Estimated δ/σ <input type="text"/> |
| Accrual Rate After Adaptation: | No Change <input type="text"/> |

The inputs on this window are almost the same as those for CHW simulations which was described in detail in Chapter 54.

Most of the entries are self-explanatory. Those that need special explanation are listed below. **All the conditional power calculations mentioned below will be performed at the estimated value, $\delta/\hat{\sigma}$, obtained at the time of the interim analysis.**

Min and Max CP: This range partitions the interim result into unfavorable, promising and favorable zones based on conditional power (CP). If the conditional power at the interim look, under the original sample size, falls in this range then the interim result is deemed to be promising and the sample size is re-estimated according to criteria specified in the remaining cells.

Max Sample Size if Adapt, multiplier : Use this cell to specify the cap for the re-estimated sample size. Since, we don't allow decrease in sample size after adaptation, the minimum sample size is the one coming from the study design. This interval [Min. Sample Size and Max. Sample Size] defines the range of re-estimated sample size after adaptation.

55 The Chen, DeMets and Lan Method

Use Wald Statistic if $CP(n_1) \geq$: The entry in this cell determines when to use the conventional Wald statistic and when to use the CHW statistic for the final analysis. The default entry is 0.5. Thus if the conditional power at the interim analysis is at least 50%, the simulations will use the conventional Wald statistic for the final analysis. Otherwise the CHW statistic will be used. The CDL method will preserve the type-1 error as long as this entry is at least 0.5. We shall show subsequently that by applying the Gao, Ware and Mehta (2008) extension, the probability in this cell can be lowered without inflating the type-1 error.

Target Conditional Power for Re-estimating Sample Size: This entry is the primary driver for the new sample size. It specifies what conditional power is desired at the end of the study. The sample size for the remainder of the trial is changed accordingly, subject to the constraints placed upon it by the **Max Sample Size if Adapt** cell.

Suppose, for example, that we wish to run 100,000 simulations at $\delta = 1.6$ and $\sigma = 7.5$, and to increase the sample size only if the conditional power at the interim analysis under the original sample size is between 0.5 and 0.9. And in that case suppose that we wish to increase the sample size by just the right amount so that the conditional power is boosted to 0.95. Furthermore suppose that the re-estimated sample size is constrained to remain between 442 and 884 subjects. To run the simulations with these specifications we would change the entries in the **Response Generation Info** tab, the **Sample Size Re-estimation** tab and the **Simulation Control Info** tab as shown below

The **Response Generation Info** tab:

| | | |
|-----------------------------|------------------|---|
| Generate Data Using: | Individual Means | <input checked="" type="checkbox"/> Common Standard Deviation |
| Mean Control (μ_c): | 0 | SD Control (σ_c): 7.5 |
| Mean Treatment (μ_t): | 1.6 | SD Treatment (σ_t): 7.5 |

The **Sample Size Re-estimation** tab:

Use Adaptation Method
 CHW CDL Müller and Schäfer

Adapt at:

Max. Sample Size if Adapt (multiplier; total #):

Upper Limit on Study Duration:

Target CP for Re-estimating Sample Size:

Use Wald Stat. if CP >=

Promising Zone Scale:

| | | |
|-----------------|----------|----------------------------------|
| Promising Zone: | Min. CP: | <input type="text" value="0.5"/> |
| | Max. CP: | <input type="text" value="0.9"/> |

CP Computation Based on:

Accrual Rate After Adaptation:

The **Simulation Control Info** tab:

| | |
|--|---|
| Number of Simulations: <input type="text" value="100000"/> | Output Options |
| Refresh Frequency: <input type="text" value="10000"/> | Output Type: <input type="text" value="Case Data"/> |
| Random Number Seed | <input type="checkbox"/> Save summary statistics for every simulation run |
| <input checked="" type="radio"/> Clock | <input type="checkbox"/> Save subject-level data for <input type="text" value="1"/> simulation runs |
| <input type="radio"/> Fixed <input type="text" value="100"/> | Note: Max. 100,000 records will be saved. |

We run the simulations by pressing the **Simulate** button. An entry for CDL simulation gets added in the **Output Preview** pane. Save this in the **Library** and

55 The Chen, DeMets and Lan Method

observe the detailed output.

☉ Average Sample Size and Look Times

| Look # | Average Sample Size (n) | Average Look Time |
|---------|-------------------------|-------------------|
| 1 | 208.001 | 25.943 |
| 2 | 530.321 | 66.142 |
| Average | 530.263 | 66.135 |

☉ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | | Efficacy | Count |
| 1 | 208 | 5.251 | 18 | 18 | 0.018% |
| 2 | 442 | 1.96 | 66445 | 99982 | 99.982% |
| Total | | | 66463 | 100000 | |
| % | | | 66.463% | | |

The null hypothesis was rejected 66,463 times in 100,000 trials for an overall power of 66.46%. The average sample size was 530.26. In contrast, if there is no sample size increase, the power would be 61% and the average sample size would be 442. This can be verified by setting the multiplier for Max. Sample Size if Adapt to 1 on **Sample Size Re-estimation** tab. This is not the full story, however. As discussed in Chapter 54, one of the major appeals of an adaptive design is the ability to invest in stages, with the additional sample size investment being required only if the interim result falls in the promising zone. From this point of view it is of interest to examine the power and expected sample size conditional on being in the unfavorable, promising and favorable zones. The top part of the simulation output shows that the trial falls into the promising zone, and thereby undergoes an adaptive sample size increase, in 25,030 of the of 100,000 simulations (25.03%). Moreover 90% of these simulated trials go on to reject the null hypothesis. This is a significant boost to the power of the study, conditional on having a favorable interim outcome. The simulation results are

displayed zone by zone as shown below.

☰ Zone-wise Averages

| | Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|---|-------------|--------------------------|----------|------------------------------|---------|-------------------|----------|---------------------|
| | | Count | Row % | Count | Row % | Count | Column % | |
| ☰ | Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| | Unfavorable | 14399 | 34.069% | 27865 | 65.931% | 42264 | 42.264% | 442 |
| | Promising | 22543 | 90.064% | 2487 | 9.936% | 25030 | 25.030% | 794.799 |
| ☰ | Favorable | 29503 | 90.256% | 3185 | 9.744% | 32688 | 32.688% | 442 |
| | Efficacy | 18 | 100.000% | 0 | 0.000% | 18 | 0.018% | 208 |
| | All Trials | 66463 | 66.463% | 33537 | 33.537% | 100000 | 100.000% | 530.263 |

Promising Zone defined as 0.5 <= CP < 0.9

The expected sample size of all the trials that undergo a sample size increase is 794.799. Although this is considerably greater than the overall average of 530.263, it is important to recognize that a sample size increase is only requested if a trial enters the promising zone at the interim look. In that case, the prospects of success become extremely promising (90 % power) and hence, the sponsor or investor might be willing to make the additional investment. The alternative approach, to commit a large sample size at the very beginning, before any interim results have been observed, might not be as attractive.

In the above figure, observe that trials fall into the favorable zone (conditional power at least 90%) 32.688% of the time. For such trials the success rate is 90.256%, and no sample size increase is called for. Trials fall into the unfavorable zone 42.264% of the time and only 34.069% of such trials go on to succeed. In this design, the adaptive option is invoked only 25.03% of the time, but once invoked, it greatly improves the chances of success. This example has highlighted the importance of evaluating any proposed adaptive strategy by simulation before adopting it. One should look at the operating characteristics of the proposed adaptive design over the entire range of plausible parameter values in order to determine if the rules for sample size increase are acceptable. If the operating characteristics are not satisfactory, it would be necessary to perform similar simulation experiments with a different adaptive strategy for sample size change. In this manner it is possible to converge to an acceptable design.

It is interesting to simulate the trial under the null hypothesis and verify that the type-1 error is indeed preserved. Accordingly set the **Mean Treatment** μ_t cell in

55 The Chen, DeMets and Lan Method

Response Generation Info tab to 0. The other simulation parameters are unchanged.

| | | |
|-----------------------------|--------------------------------|---|
| Generate Data Using: | Individual Means | <input checked="" type="checkbox"/> Common Standard Deviation |
| Mean Control (μ_c): | <input type="text" value="0"/> | SD Control (σ_c): <input type="text" value="7.5"/> |
| Mean Treatment (μ_t): | <input type="text" value="0"/> | SD Treatment (σ_t): <input type="text" value="7.5"/> |

The results based on 100,000 simulated trials are displayed below.

⊖ **Average Sample Size and Look Times**

| Look # | Average Sample Size (n) | Average Look Time |
|---------|-------------------------|-------------------|
| 1 | 208.001 | 25.941 |
| 2 | 466.652 | 58.199 |
| Average | 466.652 | 58.199 |

⊖ **Simulation Boundaries and Boundary Crossing Probabilities**

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|----------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 208 | 5.251 | 0 | 0 | 0.000% |
| 2 | 442 | 1.96 | 2426 | 100000 | 100.000% |
| Total | | | 2426 | 100000 | |
| % | | | 2.426% | | |

It is seen that only 2426 of the 100,000 trials rejected the null hypothesis, for an overall type-1 error of 2.426%. The type-1 error was thus preserved.

Suppose, in order to provide the maximum opportunity to increase the sample size we set the **Promising Zone: Min.CP** to 0, in addition to setting **Mean**

55 The Chen, DeMets and Lan Method

CP (442) >= cell from 0.5 to 0.0 as shown below and run simulations.

Use Adaptation Method
 CHW CDL Müller and Schäfer

Adapt at:

Max. Sample Size if Adapt (multiplier; total #): 884

Upper Limit on Study Duration:

Target CP for Re-estimating Sample Size:

Use Wald Stat. if CP >=

Promising Zone Scale:

| | | |
|-----------------|----------|----------------------------------|
| Promising Zone: | Min. CP: | <input type="text" value="0"/> |
| | Max. CP: | <input type="text" value="0.9"/> |

CP Computation Based on:

Accrual Rate After Adaptation:

This time the type-1 error is not preserved.

⊖ Average Sample Size and Look Times

| Look # | Average Sample Size (n) | Average Look Time |
|---------|-------------------------|-------------------|
| 1 | 208.001 | 25.942 |
| 2 | 646.826 | 80.671 |
| Average | 646.826 | 80.671 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|----------|
| | | Efficacy | | Efficacy | Count |
| 1 | 208 | 5.251 | 0 | 0 | 0.000% |
| 2 | 442 | 1.96 | 2590 | 100000 | 100.000% |
| Total | | | 2590 | 100000 | |
| % | | | 2.590% | | |

Of 100,000 simulated trials a total of 2590 rejected the null hypothesis, for a type-1 error of 2.59%. This shows that the CDL constraint is indeed necessary.

55.1.2 Binomial Endpoint: Acute Coronary Syndromes Trial

Consider a two-arm, placebo controlled randomized clinical trial for subjects with acute cardiovascular disease undergoing percutaneous coronary intervention (PCI), which we discussed in Section 54.4. The primary endpoint in this study is a composite of death, myocardial infarction or ischemia-driven revascularization during the first 48 hours after randomization. We assume on the basis of prior knowledge that the event rate for the placebo arm is 8.7%. The investigational drug is expected to reduce the event rate by at least 20%. The investigators are planning to randomize a total of 8000 subjects in equal proportions to the two arms of the study.

As explained in the beginning of this chapter, for applying CDL method, a 2 look group sequential design will suffice, without loss of generality.

It is easy to show that a group sequential design enrolling a total of 8000 subjects with an interim look after 4000 subjects are enrolled (50% of total information), will have 82% power to detect a 20% risk reduction with a one-sided level-0.025 test of significance, and early stopping efficacy boundary derived from the Lan and DeMets (1983) O'Brien-Fleming type error spending function.

| | CDL2:Des 1 |
|--|-------------|
| Mnemonic | PN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 2 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.825 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.087 |
| Proportion under Treatment (π_t) | 0.07 |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.017 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Sample Size | |
| Maximum | 8000 |
| Expected Under H0 | 7993.899 |
| Expected Under H1 | 7277.838 |

55 The Chen, DeMets and Lan Method

The actual risk reduction is expected to be larger, but could also be as low as 15%, a treatment effect that would still be of clinical interest given the severity and importance of the outcomes. In addition, there is some uncertainty about the magnitude of the placebo event rate. For these reasons the investigators wish to build into the trial design some flexibility for adjusting the sample size. Two options under consideration are, a group sequential design with the possibility of early stopping in case the risk reduction is large, and an adaptive design with the possibility of increasing the sample size in case the risk reduction is small. In the remainder of this section we shall discuss these two options and show how they may be combined into a single design that captures the benefits of both.

For this design, where the risk reduction is 20%; the probabilities of crossing boundary at Look1 (N=4000) is 0.181, and at Final Look 0.644; the overall power is 82%.

As we did in chapter 54, we partition the sample space into three important zones, unfavorable, promising and favorable, based on the conditional power attained at the interim look. The sample size may then be increased if the interim results fall inside the promising zone, thereby recovering the lost power. The additional feature of the CDL design is, however, that if conditional power at the interim look is at least 50% it is not necessary to use the CHW statistic at the final analysis. The conventional Wald statistic may then be used without inflating the type-1 error.

Adaptive Group Sequential Design We convert the two-look group sequential design Des 1 into an adaptive group sequential design to increase the sample size at look 1, when 4000 subjects have been enrolled. The rules governing the sample size increase similar to the rules specified in Section 55.1.1 for the schizophrenia trial. We shall study the operating characteristics of the above CDL design through simulation. The option for choosing CDL method for adaptation is on the **Sample Size Re-estimation** tab. We invoke the CDL simulation by clicking on the radio-button for

CDL, The resulting simulation input window will appear as shown below.

| | | | |
|--|--------------------------------------|--|----|
| Use Adaptation Method | | | |
| <input type="radio"/> CHW | <input checked="" type="radio"/> CDL | <input type="radio"/> Müller and Schäfer | |
| Adapt at: | Look # | 1 | |
| Max. Sample Size if Adapt (multiplier; total #): | 1 | 8000 | |
| Target CP for Re-estimating Sample Size: | 0.825 | | |
| Use Wald Stat. if CP >= | 0.5 | | |
| Promising Zone Scale: | Cond. Power | | CP |
| Promising Zone: | Min. CP: | 0.3 | |
| | Max. CP: | 0.825 | |
| CP Computation Based on: | Estimated (n_c, n_t) | | |

The inputs on this window are almost the same as those for CHW simulations which was described in detail in Chapter 54.

Most of the entries are self-explanatory. Those that need special explanation are similar to what have been described in section 55.1.1 for schizophrenia example. **All the conditional power calculations mentioned below will be performed at the estimated value, π_c, π_t , obtained at the time of the interim analysis.**

Suppose, for example, that we wish to run 100,000 simulations at risk reduction $\rho = 0.15$ and to increase the sample size only if the conditional power at the interim analysis under the original sample size is between 0.5 and 0.9. And in that case suppose that we wish to increase the sample size by just the right amount so that the conditional power is boosted to 0.95. Furthermore suppose that the re-estimated sample size is constrained to remain between 8000 and 16000 subjects. To run the simulations with these specifications we would change the entries in the three tabs as shown below.

observe the detailed output.

⊖ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 4000 |
| 2 | 9731.685 |
| Average | 9318.603 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | | Count | % |
| | | Lower | Efficacy | | |
| 1 | 4000 | -2.963 | 7207 | 7207 | 7.207% |
| 2 | 8000 | -1.969 | 55185 | 92793 | 92.793% |
| Total | | | 62392 | 100000 | |
| % | | | 62.392% | | |

The null hypothesis was rejected 62,392 times in 100,000 trials for an overall power of 62.4%. The average sample size was 9318.60. In contrast, if there is no sample size increase, the power would be 57.2% and the average sample size would be 7703.1. Next, let us consider these results zone by zone. As discussed in Chapter 54, one of the major appeals of an adaptive design is the ability to invest in stages, with the additional sample size investment being required only if the interim result falls in the promising zone. From this point of view it is of interest to examine the power and expected sample size conditional on being in the unfavorable, promising and favorable zones. The bottom part of the simulation output shows that the trial falls into the promising zone, and thereby undergoes an adaptive sample size increase, in 24,944 of the of 100,000 simulations (24.94%). Moreover 87.98% of these simulated trials go on to reject the null hypothesis. This is a significant boost to the power of the study,

55 The Chen, DeMets and Lan Method

conditional on having a favorable interim outcome.

Zone-wise Averages

| | Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|---|-------------|--------------------------|----------|------------------------------|---------|-------------------|----------|---------------------|
| | | Count | Row % | Count | Row % | Count | Column % | |
| ⊖ | Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| | Unfavorable | 15813 | 35.075% | 29271 | 64.925% | 45084 | 45.084% | 8000 |
| | Promising | 21946 | 87.981% | 2998 | 12.019% | 24944 | 24.944% | 14441.961 |
| | Favorable | 17426 | 76.547% | 5339 | 23.453% | 22765 | 22.765% | 8000 |
| ⊖ | Efficacy | 7207 | 100.000% | 0 | 0.000% | 7207 | 7.207% | 4000 |
| | All Trials | 62392 | 62.392% | 37608 | 37.608% | 100000 | 100.000% | 9318.603 |
| Promising Zone defined as $0.5 \leq CP < 0.9$ | | | | | | | | |

The expected sample size of all the trials that undergo a sample size increase is 14,441.961. Although this is considerably greater than the overall average of 9,318.603, it is important to recognize that a sample size increase is only requested if a trial enters the promising zone at the interim look. In that case, the prospects of success become extremely promising (87.98% power) and hence, the sponsor or investor might be willing to make the additional investment. The alternative approach, to commit a large sample size at the very beginning, before any interim results have been observed, might not be as attractive. The simulation results are also displayed zone by zone as shown below.

Observe that trials fall into the Favorable + Efficacy zone (conditional power at least 90%) 29.97% of the time. For such trials the success rate is 82.19%, and no sample size increase is called for. Trials fall into the unfavorable zone 45.08% of the time and only 35.08% of such trials go on to succeed. In this design the adaptive option is invoked 24.94% of the time, and once invoked, it greatly improves the chances of success. This example has highlighted the importance of evaluating any proposed adaptive strategy by simulation before adopting it. One should look at the operating characteristics of the proposed adaptive design over the entire range of plausible parameter values in order to determine if the rules for sample size increase are acceptable. If the operating characteristics are not satisfactory it would be necessary to perform similar simulation experiments with a different adaptive strategy for sample size change. In this manner it is possible to converge to an acceptable design.

It is interesting to simulate the trial under the null hypothesis and verify that the type-1 error is indeed preserved. Accordingly set the **Proportion Under Treatment** cell in **Response Generation Info** tab to **Proportion Under Control**. The

other simulation parameters are unchanged.

Specify Proportion

Prop. under Control (π_c):

Prop. under Treatment (π_t):

The results based on 100,000 simulated trials are displayed below.

Zone-wise Averages

| | Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|---|-------------|--------------------------|----------|------------------------------|---------|-------------------|----------|---------------------|
| | | Count | Row % | Count | Row % | Count | Column % | |
| ⊖ | Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| | Unfavorable | 856 | 0.931% | 91067 | 99.069% | 91923 | 91.923% | 8000 |
| | Promising | 847 | 14.044% | 5184 | 85.956% | 6031 | 6.031% | 14912.324 |
| ⊖ | Favorable | 597 | 31.339% | 1308 | 68.661% | 1905 | 1.905% | 8000 |
| | Efficacy | 141 | 100.000% | 0 | 0.000% | 141 | 0.141% | 4000 |
| | All Trials | 2441 | 2.441% | 97559 | 97.559% | 100000 | 100.000% | 8411.242 |
| Promising Zone defined as 0.5 <= CP < 0.9 | | | | | | | | |

It is seen that only 2441 of the 100,000 trials rejected the null hypothesis, for an overall type-1 error of 2.44%. The type-1 error was thus preserved.

Suppose, in order to provide the maximum opportunity to increase the sample size we set the **Promising Zone**: **Min.CP** to 0, in addition to setting the response rate same for control and treatment. Let us keep the other simulation parameters are

55 The Chen, DeMets and Lan Method

unchanged.

| | | |
|--|--------------------------------------|--|
| Use Adaptation Method | | |
| <input type="radio"/> CHW | <input checked="" type="radio"/> CDL | <input type="radio"/> Müller and Schäfer |
| Adapt at: | Look # | 1 |
| Max. Sample Size if Adapt (multiplier; total #): | 2 | 16000 |
| Target CP for Re-estimating Sample Size: | 0.95 | |
| Use Wald Stat. if CP >= | 0.5 | |
| Promising Zone Scale: | Cond. Power | <input type="button" value="CP"/> |
| Promising Zone: | Min. CP: | 0 |
| | Max. CP: | 0.9 |
| CP Computation Based on: | Estimated (n_c, n_t) | |

The results based on 100,000 simulated trials are displayed below.

⊖ Zone-wise Averages

| | Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|---|-------------|--------------------------|----------|------------------------------|---------|-------------------|----------|---------------------|
| | | Count | Row % | Count | Row % | Count | Column % | |
| ⊖ | Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| | Unfavorable | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| | Promising | 1482 | 1.514% | 96408 | 98.486% | 97890 | 97.890% | 11936.073 |
| ⊖ | Favorable | 608 | 30.910% | 1359 | 69.090% | 1967 | 1.967% | 8000 |
| | Efficacy | 143 | 100.000% | 0 | 0.000% | 143 | 0.143% | 4000 |
| | All Trials | 2233 | 2.233% | 97767 | 97.767% | 100000 | 100.000% | 11847.302 |
| Promising Zone defined as $0 \leq CP < 0.9$ | | | | | | | | |

⊖ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 4000 |
| 2 | 11858.539 |
| Average | 11847.302 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | | Count | % |
| | | Lower | Efficacy | | |
| 1 | 4000 | -2.963 | 143 | 143 | 0.143% |
| 2 | 8000 | -1.969 | 2090 | 99857 | 99.857% |
| Total | | | 2233 | 100000 | |
| % | | | 2.233% | | |

It is seen that only 2233 of the 100,000 trials rejected the null hypothesis, for an overall type-1 error of 2.23%. The type-1 error was thus preserved.

Now suppose we disable the CDL constraint by changing the entry in the **Use Wald**

55 The Chen, DeMets and Lan Method

Stat. if CP (442) >= cell from 0.5 to 0.0 as shown below.

Use Adaptation Method

CHW CDL Müller and Schäfer

Adapt at: Look # 1

Max. Sample Size if Adapt (multiplier; total #): 2 16000

Target CP for Re-estimating Sample Size: 0.95

Use Wald Stat. if CP >= 0

Promising Zone Scale: Cond. Power CP

Promising Min. CP: 0

Zone: Max. CP: 0.9

CP Computation Based on: Estimated (n_c, n_t)

This time the type-1 error is not preserved.

Zone-wise Averages

| | Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|---|-------------|--------------------------|----------|------------------------------|---------|-------------------|----------|---------------------|
| | | Count | Row % | Count | Row % | Count | Column % | |
| ⊖ | Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| | Unfavorable | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| | Promising | 1878 | 1.917% | 96081 | 98.083% | 97959 | 97.959% | 11748.738 |
| ⊖ | Favorable | 599 | 31.493% | 1303 | 68.507% | 1902 | 1.902% | 8000 |
| | Efficacy | 139 | 100.000% | 0 | 0.000% | 139 | 0.139% | 4000 |
| | All Trials | 2616 | 2.616% | 97384 | 97.384% | 100000 | 100.000% | 11666.667 |

Promising Zone defined as 0 <= CP < 0.9

Of 100,000 simulated trials a total of 2616 rejected the null hypothesis, for a type-1 error of 2.62% which is slightly inflated. This shows that the CDL constraint is indeed necessary.

55.1.3 Survival Endpoint: Lung Cancer Trial

Let us re-visit the non-small cell lung cancer trial introduced in Section 54.5 of Chapter 54. This is a two-arm multi-center randomized clinical trial for subjects with advanced metastatic non-small cell lung cancer comparing the current standard second line therapy (docetaxel+cisplatin) to a new docetaxel containing combination regimen. The primary endpoint is overall survival (OS). The study is required to have one-sided $\alpha = 0.025$, and 90% power to detect an improvement in median survival, from 8

months on the control arm to 11.4 months on the experimental arm, which corresponds to a hazard ratio of 0.7. We shall first create a group sequential design for this study in East, and shall then show how the design may be improved by permitting an increase in the number of events and sample size at the time of the interim analysis.

Following the steps exactly as outlined in Section 54.5.1 of Chapter 54 we create a 2-look group sequential design with an efficacy boundary derived from the Lan and DeMets (1983) O'Brien-Fleming type spending function, a futility boundary derived from the γ -spending function of Hwang, Shih and DeCani (1990) with parameter $\gamma = -5$, and an interim analysis at 50% of the total information. It is planned to enroll subjects over 24 months and extend the follow-up for six additional months, thereby completing the study in 30 months. This design is created in East and displayed below

55 The Chen, DeMets and Lan Method

as Des1.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

| Test Parameters | |
|--------------------------------|-------------|
| Design ID | Des1 |
| Design Type | Superiority |
| Number of Looks | 2 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Attained α | 0.025 |
| Power | 0.901 |
| Model Parameters | |
| HR = λ_1/λ_0 | |
| Under H0 | 1 |
| Under H1 | 0.7 |
| Ratio of Medians | 1.429 |
| Var (Log HR) | Null |
| Allocation Ratio (n_1/n_0) | 1 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Futility Boundary | Gm (S) (NS) |
| Accrual/Dropout Parameters | |
| Accrual Duration | 24 |
| Max Study Duration | 30 |
| Dropout | No |

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|-------------------------------|-------------|---------------|---------|
| Sample Size (n) | | | |
| Maximum | 241 | 242 | 483 |
| Expected H1 | 226.836 | 226.836 | 453.671 |
| Expected H0 | 210.851 | 210.851 | 421.701 |
| Events (s) | | | |
| Maximum | 180 | 154 | 334 |
| Expected H1 | 162.827 | 135.155 | 297.982 |
| Expected H0 | 130.437 | 130.437 | 260.874 |
| Maximum Information (I): 83.5 | | | |

Accrual and Study Duration

| | Accrual Duration | Study Duration |
|-------------|------------------|----------------|
| Maximum | 24 | 29.984 |
| Expected H1 | 22.543 | 26.954 |
| Expected H0 | 20.954 | 22.92 |

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (s/s_max) | Events (s) | Cumulative α Spent | Cumulative β Spent | Boundaries | |
|--------|--------------------------|------------|---------------------------|--------------------------|------------|------------|
| | | | | | Efficacy Z | Futility Z |
| 1 | 0.5 | 167 | 0.002 | 0.008 | -2.963 | 0.126 |
| 2 | 1 | 334 | 0.025 | 0.099 | -1.969 | -1.969 |

Events, Sample Size, Pipeline and Analysis Times: Look by Look (Under H0)

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (s) | Pipeline (n-s) | Analysis Time | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|------------|----------------|---------------|---|----------|
| | | | | | | Efficacy | Futility |
| 1 | 0.5 | 348 | 167 | 181 | 17.251 | 0.002 | 0.45 |
| 2 | 1 | 483 | 334 | 149 | 27.583 | 0.023 | 0.625 |

Events, Sample Size, Pipeline and Analysis Times: Look by Look (Under H1)

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (s) | Pipeline (n-s) | Analysis Time | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|------------|----------------|---------------|---|----------|
| | | | | | | Efficacy | Futility |
| 1 | 0.5 | 372 | 167 | 205 | 18.455 | 0.255 | 0.008 |
| 2 | 1 | 483 | 334 | 149 | 29.984 | 0.645 | 0.092 |

Survival Information : Median Survival Times

| Median Survival Times | | Hazard Ratio |
|-------------------------|----------------|-----------------------------------|
| Control (λ_c) | Treatment (M1) | Alt. (λ_{M1}/λ_c) |
| 8 | 11.429 | 0.7 |

Accrual Information

| Period # | Starting at Time | Cum. % Accrued |
|----------|------------------|----------------|
| 1 | 24 | 100 |

Variable Follow-Up Design: All subjects are followed until failure, drop out or end of study.

Des1 requires an up-front commitment of 334 events to achieve 90% power. With an enrollment of 483 subjects over 24 months, the required 334 events are expected to arrive within 30 months. An interim analysis will be performed after 167 events are obtained (50% of the total information). Under the alternative hypothesis that the

hazard ratio is 0.7, the chance of crossing the efficacy boundary at the interim look is about 26% leading to an expected sample size of 454 subjects and an expected study duration of about 27 months.

⊖ Sample Size Information

| | Control Arm | Treatment Arm | Total |
|------------------------------|-------------|---------------|---------|
| Sample Size (n) | | | |
| Maximum | 241 | 242 | 483 |
| Expected H1 | 226.836 | 226.836 | 453.671 |
| Expected H0 | 210.851 | 210.851 | 421.701 |
| Events (s) | | | |
| Maximum | 180 | 154 | 334 |
| Expected H1 | 162.827 | 135.155 | 290.11 |
| Expected H0 | 130.437 | 130.437 | 258.635 |
| Maximum Information (I):83.5 | | | |

⊖ Accrual and Study Duration

| | Accrual Duration | Study Duration |
|-------------|------------------|----------------|
| Maximum | 24 | 29.984 |
| Expected H1 | 22.543 | 26.954 |
| Expected H0 | 20.954 | 22.92 |

Although Des1 is adequately powered to detect a hazard ratio of 0.7, its power deteriorates from 90% to below 68% if the true hazard ratio is 0.77, an effect that is still considered clinically meaningful. To see this let us simulate Des1 under HR=0.77.

Select Des1 in the **Library** and click the **S** icon. You will be taken to the usual simulation input window. This has four tabs as below:

Simulation Parameters Response Generation Info Accrual/Dropout Info Sample Size Re-estimation Simulation Control Info

The use of four tabs **Simulation Parameters**, **Response Generation Info**, **Accrual/Dropout Info** and **Simulation Control Info** is exactly same as that explained

55 The Chen, DeMets and Lan Method

in sections of CHW Simulations.

You may refer to Chapter 54, Section 54.5.3 for a complete description of their functioning. The fourth tab, **Sample Size Re-estimation**, is almost identical to the corresponding tab for CHW simulations but contains **one** additional input parameter that distinguishes the CDL method from the CHW method for adaptive design. We will assume for the remainder of this section that the user is familiar with the CHW simulation worksheet. If not, please refer to Section 54.5.3 of Chapter 54 where this worksheet was fully discussed with a worked example.

Observe that the **Response Generation Info** tab currently displays a hazard ratio of 0.7, since this was the value specified at the design stage.

| Survival Information | | | | |
|--|-------------|--------------|-----------|--------------|
| <input checked="" type="radio"/> Using Hazard Rates <input type="radio"/> Using Cum. % Survival | | | | |
| # of Hazard Pieces <input type="text" value="1"/> | | | | |
| Piece | Starting At | Hazard Rates | | Hazard Ratio |
| | | Control | Treatment | |
| 1 | 0.000 | 0.0866 | 0.061 | 0.700 |

We know from the design of Des1 that a hazard ratio of 0.7 will yield 90% power. But what if the true hazard ratio was 0.77? The resultant deterioration in power can be evaluated by simulation. Accordingly we shall alter the Treatment cell, containing the hazard 0.0607, by replacing it with $0.77 * 0.0866 = 0.0667$.

| Survival Information | | | | |
|--|-------------|--------------|-----------|--------------|
| <input checked="" type="radio"/> Using Hazard Rates <input type="radio"/> Using Cum. % Survival | | | | |
| # of Hazard Pieces <input type="text" value="1"/> | | | | |
| Piece | Starting At | Hazard Rates | | Hazard Ratio |
| | | Control | Treatment | |
| 1 | 0.000 | 0.0866 | 0.067 | 0.770 |

To run 10,000 simulations with a hazard ratio of 0.77, click on the **Simulate** button. The following simulation output is displayed.

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

| Simulation Parameters | |
|-------------------------------|---------------------|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 2 |
| Test Type | 1-Sided |
| Sample Size (n) | 483 |
| Fix at Each Look | Total No. of Events |
| Test Statistic | Logrank |
| Average Events | 310.804 |
| Total Accrual Duration | 24 |
| Avg. Power at Termination | 0.663 |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

⊖ Average Sample Size and Look Times

| Look # | Average Sample Size | Average Events | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|-------------------|-------------------|
| | | Control | Treatment | | |
| 1 | 364.81 | 90.63 | 76.37 | 18.105 | 6.002 |
| 2 | 483 | 177.378 | 156.622 | 29.252 | 9.101 |
| Average | 466.576 | 165.743 | 145.061 | 27.705 | 8.671 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Events | Boundaries | | Stopping For | | Total Simulations | |
|--------|--------|------------|----------|--------------|----------|-------------------|---------|
| | | Efficacy | Futility | Efficacy | Futility | Count | % |
| | | Lower | Upper | | | | |
| 1 | 167 | -2.963 | 0.126 | 1029 | 360 | 1389 | 13.890% |
| 2 | 334 | -1.969 | -1.969 | 5602 | 3009 | 8611 | 86.110% |
| Total | | | | 6631 | 3369 | 10000 | |
| % | | | | 66.310% | 33.690% | | |

⊖ Response Generation Parameters

No. of Hazard Pieces: 1
Input Method: Median Survival Times

| Piece # | Control | Treatment | Ratio |
|---------|---------|-----------|-------|
| 1 | 8 | 10.39 | 0.77 |

⊖ Accrual/Dropout Parameters

Sample Size: 483
Subjects are Followed: Until End of Study
Accrual Duration: 24
Accrual Input Method: Cum. Accrual %

| Period # | By Time | Cum. % Accrued |
|----------|---------|----------------|
| 1 | 24 | 100 |

The overall power is only 66.31% suggesting that it might be useful to consider an adaptive increase in the number of events and sample size at the interim look.

The impact of an adaptive increase in the number of events and sample size on power and study duration can be evaluated by simulation. Accordingly click on the **Sample Size Re-estimation** tab and select the option of **CDL** on this tab. This will take you to

55 The Chen, DeMets and Lan Method

the input parameters for performing the adaptive simulations using CDL method.

| | |
|--|--|
| Use Adaptation Method | |
| <input type="radio"/> CHW | <input checked="" type="radio"/> CDL |
| <input type="radio"/> Müller and Schäfer | |
| Adapt at: | Look # <input type="text" value="1"/> |
| Max. # of Events if Adapt (multiplier; total #): | <input type="text" value="1"/> <input type="text" value="334"/> |
| Max. Sample Size if Adapt (multiplier; total #): | <input type="text" value="1"/> <input type="text" value="483"/> |
| Upper Limit on Study Duration: | <input type="text" value="90"/> |
| Target CP for Re-estimating # of Events: | <input type="text" value="0.9"/> |
| Use Wald Stat. if CP >= | <input type="text" value="0.5"/> |
| Promising Zone Scale: | Cond. Power <input type="text" value="CP"/> |
| Promising Zone: | Min. CP: <input type="text" value="0.3"/> Max. CP: <input type="text" value="0.9"/> |
| CP Computation Based on: | <input type="text" value="Estimated HR"/> |
| Accrual Rate After Adaptation: | <input type="text" value="No Change"/> |

These inputs and output quantities (tables and charts) were fully described in Section 54.5.3 of Chapter 54. Thus, they will not be discussed again here with the exception of a single additional input parameter that appears on the tab when the CDL method is selected. This input is not a part of input parameters for the CHW simulations. This new parameter appears between the **Target CP for Re-estimating Events** field and the **Promising Zone Scale** field.

| | |
|--|---|
| Target CP for Re-estimating # of Events: | <input type="text" value="0.9"/> |
| Use Wald Stat. if CP >= | <input type="text" value="0.5"/> |
| Promising Zone Scale: | Cond. Power <input type="text" value="CP"/> |

Suppose the following values have been entered into the Input Parameters Table.

Use Adaptation Method CHW CDL Müller and Schäfer

Adapt at: Look #

Max. # of Events if Adapt (multiplier; total #):

Max. Sample Size if Adapt (multiplier; total #):

Upper Limit on Study Duration:

Target CP for Re-estimating # of Events:

Use Wald Stat. if CP >=

Promising Zone Scale:

Promising Zone: Min. CP: Max. CP:

CP Computation Based on:

Accrual Rate After Adaptation:

These inputs imply that there will be a 50% increase in the number of events for each simulation that enters the promising zone and up to a 50% increase in the sample size also. The promising zone is specified by conditional power (based on estimated HR) being between 0.3 and 0.9. These adaptation rules are the same as the adaptation rules applied in Section 54.5.3 of Chapter 54. However, the test statistic to be used for the final analysis will depend on the CP observed at the interim look. If this CP exceeds 0.5, the conventional Wald statistic (equation (54.3) in Chapter 54) will be used for the final analysis, whereas if this CP is below 0.5, the weighted CHW statistic (equation (54.2) in Chapter 54) will be used for the final analysis. Upon pressing the **Simulate** button the following outputs are obtained in the Table of Simulation Results by Zone

Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Rejecting H1 | | Total Simulations | | Average Sample Size | Average Number of Events | Average Accrual Duration | Average Study Duration |
|-------------|--------------------------|----------|--------------------------|----------|-------------------|----------|---------------------|--------------------------|--------------------------|------------------------|
| | Count | Row % | Count | Row % | Count | Column % | | | | |
| Futility | 0 | 0.000% | 357 | 100.000% | 357 | 3.570% | 364.134 | 167 | 18.042 | 18.095 |
| Unfavorable | 877 | 34.541% | 1662 | 65.459% | 2539 | 25.390% | 483 | 334 | 23.95 | 29.193 |
| Promising | 2984 | 86.795% | 454 | 13.205% | 3438 | 34.380% | 723.946 | 501 | 35.874 | 37.344 |
| Favorable | 2400 | 90.090% | 264 | 9.910% | 2664 | 26.640% | 483 | 334 | 23.95 | 29.261 |
| Efficacy | 1002 | 100.000% | 0 | 0.000% | 1002 | 10.020% | 365.519 | 167 | 18.068 | 18.119 |
| All Trials | 7263 | 72.630% | 2737 | 27.370% | 10000 | 100.000% | 549.822 | 368.719 | 27.249 | 30.508 |

Promising Zone defined as 0.3 <= CP < 0.9

55 The Chen, DeMets and Lan Method

and in the Table of Output for all Trials.

⊖ Zone-Wise Percentiles

| | Zone | Percentile | Number of Events | Sample Size | Accrual Duration | Study Duration |
|---|-------------|------------|------------------|-------------|------------------|----------------|
| + | Futility | Average | 167 | 364.134 | 18.042 | 18.095 |
| + | Unfavorable | Average | 334 | 483 | 23.95 | 29.193 |
| + | Promising | Average | 501 | 723.946 | 35.874 | 37.344 |
| + | Favorable | Average | 334 | 483 | 23.95 | 29.261 |
| + | Efficacy | Average | 167 | 365.519 | 18.068 | 18.119 |
| ⊖ | All Trials | 5.000% | 167 | 360 | 17.791 | 17.834 |
| | | 25.000% | 334 | 483 | 23.925 | 28.544 |
| | | 50.000% | 334 | 483 | 23.981 | 29.678 |
| | | 75.000% | 501 | 724 | 35.849 | 36.941 |
| | | 95.000% | 501 | 724 | 35.941 | 38.026 |
| | | Average | 368.719 | 549.822 | 27.249 | 30.508 |

These results are almost the same as were obtained by use of the CHW method.

The main advantage of using the CDL method is that one can dispense with the use of the non-standard, weighted CHW statistic (54.2) as long as the conditional power at the interim analysis exceeds 0.5. Therefore, if the minimum CP for the promising zone is itself 0.5, one can dispense with the use of the CHW statistic altogether and always use the conventional Wald statistic at the time of the final analysis.

To see that the CDL condition ($CP \geq 0.5$) is necessary for preserving the type-1 error if the Wald statistic is always used for increasing the number of events, consider the following simulation experiment based on 10,000 simulated trials.

Set the hazard ratio to 1, so as to simulate under the null hypothesis.

Survival Information

of Hazard Pieces Input Method:

Hazard Ratio

Hazard Ratio (λ_t / λ_c):

| Piece | Starting at Time | Hazard Rates | | Hazard Ratio |
|-------|------------------|--------------|-----------|--------------|
| | | Control | Treatment | |
| 1 | 0.000 | 0.087 | 0.087 | 1.000 |

55 The Chen, DeMets and Lan Method

Row 6 of the Table of Simulation Results by Zone displays the results for all trials, combined across zones. Thus Column 3 of Row 6 of this table displays the magnitude of the type-1 error, 0.0317 which is seen to exceed 0.025 even after accounting for Monte Carlo error.

To be sure, the entries in the **Sample Size Re-estimation** tab are rather extreme and unrealistic. However this example serves to illustrate the point that control of type-1 error cannot be guaranteed if the Wald statistic replaces the CHW statistic inappropriately. On the other hand suppose we set the **Use Wald Stat. if CP >=** to 0.5.

| | | |
|--|--------------------------------------|--|
| Use Adaptation Method | | |
| <input type="radio"/> CHW | <input checked="" type="radio"/> CDL | <input type="radio"/> Müller and Schäfer |
| Adapt at: | Look # | 1 |
| Max. # of Events if Adapt (multiplier; total #): | 10 | 3340 |
| Max. Sample Size if Adapt (multiplier; total #): | 10 | 4830 |
| Upper Limit on Study Duration: | 500 | |
| Target CP for Re-estimating # of Events: | 0.9 | |
| Use Wald Stat. if CP >= | 0.5 | |
| Promising Zone Scale: | Cond. Power | CP |
| Promising Zone: | Min. CP: | 0 |
| | Max. CP: | 0.9 |
| CP Computation Based on: | Estimated HR | |
| Accrual Rate After Adaptation: | No Change | |

With these inputs, the CHW statistic will be used for the final analysis if, at the interim analysis, $0 < CP(334) < 0.5$ and the conventional Wald statistic will be use if, at the interim analysis, $0.5 \leq CP(334) < 0.9$. Upon pressing the **Simulate** button, the

following results are displayed in the table of Simulation Results by Zone.

Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Rejecting H1 | | Total Simulations | | Average Sample Size | Average Number of Events | Average Accrual Duration | Average Study Duration | |
|------|--------------------------|-------|--------------------------|-------|-------------------|----------|---------------------|--------------------------|--------------------------|------------------------|---------|
| | Count | Row % | Count | Row % | Count | Column % | | | | | |
| ⊖ | Futility | 0 | 0.000% | 4529 | 100.000% | 4529 | 45.290% | 346.814 | 167 | 17.163 | 17.213 |
| | Unfavorable | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0 | 0 | 0 |
| | Promising | 177 | 3.362% | 5087 | 96.638% | 5264 | 52.640% | 2837.84 | 2613.607 | 140.858 | 141.241 |
| ⊕ | Favorable | 49 | 26.630% | 135 | 73.370% | 184 | 1.840% | 483 | 334 | 23.947 | 27.686 |
| | Efficacy | 23 | 100.000% | 0 | 0.000% | 23 | 0.230% | 349.913 | 167 | 17.086 | 17.134 |
| | All Trials | 249 | 2.490% | 9751 | 97.510% | 10000 | 100.000% | 1660.603 | 1457.967 | 82.401 | 82.694 |

Promising Zone defined as 0 <= CP < 0.9

This time 249 of the 10,000 simulations rejected the null hypothesis, for an overall type-1 error of 0.0249. Thus the type-1 error is controlled.

55.2 Extension of CDL Method

55.2.1 Underlying Theory

55.2.2 Normal Endpoint

55.2.3 Binomial Endpoint

55.2.4 Survival Endpoint

We now describe an extension to the CDL method in which the 0.5 probability limit above which one is permitted to substitute the conventional Wald statistic for the CHW statistic can be lowered. The amount by which the CDL criterion can be lowered will depend on the other design parameters of the trial, and must be computed separately for each specific trial design. We have provided a table of cut-off values from which one may extrapolate for this purpose. The underlying theory is discussed next and provides some insight into why both, the CDL method and the extended CDL method are able to protect the type-1 error.

55.2.1 Underlying Theory

The results in this section are only valid for one-sided tests, and only for a sample size increase, but not for a sample size decrease. For simplicity we confine the discussion to tests of $H_0: \delta = 0$ against the one-sided alternative $\delta > 0$. However these results apply equally to tests against the one-sided alternative $\delta < 0$.

The ability to relax the criterion for using the conventional Wald statistic in an adaptive trial is based on a result due to Gao, Ware and Mehta (2008). Using the notation introduced in Chapter 54, let (n_1, n_2) be the **pre-specified** cumulative sample sizes for look 1 and look 2, respectively, and let (b_1, b_2) be corresponding one-sided level- α boundaries. Let

$$Z_1 = \hat{\delta}_1 \sqrt{I_1}$$

be the observed value of the Wald statistic at look 1, where I_1 is the Fisher information about δ based on the n_1 observations available at the time of the interim analysis. After observing $Z_1 = z_1$ suppose that the cumulative sample size for the final analysis is increased from n_2 to n_2^* . Using the notation developed in Chapter 54, we define the incremental Wald statistic

$$Z^{*(2)} = \hat{\delta}^{*(2)} \sqrt{I^{*(2)}},$$

55 The Chen, DeMets and Lan Method

where $I^{*(2)}$ is the Fisher information about δ based only on the additional $n_2^* - n_1$ observations obtained after the interim analysis. The CHW statistic (54.2) can be expressed as

$$Z_{2,\text{chw}}^* = \sqrt{\frac{n_1}{n_2}} Z_1 + \sqrt{\left(1 - \frac{n_1}{n_2}\right)} Z^{*(2)}$$

while the conventional Wald statistic (54.4) can be expressed as

$$Z_{2,\text{wald}}^* = \sqrt{\frac{n_1}{n_2^*}} Z_1 + \sqrt{\left(1 - \frac{n_1}{n_2^*}\right)} Z^{*(2)}.$$

Since the CHW statistic preserves the type-1 error it is clear that

$$P_0(Z_1 \geq b_1) + P_0(Z_1 < b_1, Z_{\text{chw}}^* \geq b_2) = \alpha. \quad (55.1)$$

However, due to the data dependent sample size change at look 1,

$$P_0(Z_1 \geq b_1) + P_0(Z_1 < b_1, Z_{2,\text{wald}}^* \geq b_2) \neq \alpha.$$

Therefore using the conventional Wald statistic for the final analysis will not protect the type-1 error. Gao, Ware and Mehta (2008) have shown that if, upon observing $Z_1 = z_1$ and increasing the total sample size from n_2 to n_2^* we change the final critical boundary from b_2 to

$$b_2(z_1, n_2^*) = (n_2^*)^{-0.5} \left[\sqrt{\frac{n_2^* - n_1}{n_2 - n_1}} (b_2 \sqrt{n_2} - z_1 \sqrt{n_1}) + z_1 \sqrt{n_1} \right] \quad (55.2)$$

then

$$P_0(Z_1 \geq b_1) + P_0(Z_1 < b_1, Z_{2,\text{wald}}^* \geq b_2(z_1, n_2^*)) = \alpha. \quad (55.3)$$

Thus we can use the conventional Wald statistic for the final analysis and also protect the type-1 error provided we replace the final critical boundary value b_2 by $b_2(z_1, n_2^*)$. The extended CDL method follows from this result.

The Extended CDL Method: Whenever $b_2(z_1, n_2^*) \leq b_2$, we may reject the null hypothesis $H_0: \delta = 0$ in favor of the one sided alternative that $\delta > 0$ if

$$(Z_1 \geq b_1) \text{ or } (Z_1 < b_1, Z_{2,\text{wald}}^* \geq b_2) \quad (55.4)$$

and the type-1 error will not exceed α notwithstanding the data dependent sample size increase from n_2 to n_2^* at the interim analysis. This result holds because $b_2(z_1, n_2^*) \leq b_2$ implies that

$$\alpha = P_0(Z_1 \geq b_1) + P_0(Z_1 < b_1, Z_{2,\text{wald}}^* \geq b_2(z_1, n_2^*)) \quad (55.5)$$

$$\geq P_0(Z_1 \geq b_1) + P_0(Z_1 < b_1, Z_{2,\text{wald}}^* \geq b_2) \quad (55.6)$$

Recall that the regular CDL method satisfies (55.6) only if the conditional power at the interim look is at least 0.5. We shall see that the extended CDL method satisfies (55.6) over a wider range of conditional powers. To show this we must investigate the behavior of the adjusted boundary $b_2(z_1, n_2^*)$ as a function of z_1 and n_2^* . We first reduce the dimensionality of the investigation by making the increased sample size n_2^* a function of z_1 . This is achieved by imposing the requirement that the new sample size, n_2^* , should be such that the conditional power given z_1 , evaluated at $\hat{\delta}_1$, reaches some pre-specified target value, subject however to an upper limit on the magnitude of the sample size increase. To be specific, define

$$CP_{\hat{\delta}_1}(z_1, n_2^*) = P_{\hat{\delta}_1}(Z_{2,\text{chw}}^* \geq b_2|z_1).$$

Under the extended CDL method, we pre-specify a *target value* for $CP_{\hat{\delta}_1}(z_1, n_2^*)$, say $1 - \beta$, and attempt to reach it by altering the sample size from n_2 to n_2^* . The first step is to find the sample size $n_2'(z_1)$ for each possible value of z_1 such that

$$CP_{\hat{\delta}_1}(z_1, n_2'(z_1)) = 1 - \beta. \tag{55.7}$$

A simplification of Gao, Ware and Mehta(2008, equation (5)) shows that (55.7) is satisfied by the function

$$n_2'(z_1) = \left[\frac{n_1}{z_1^2} \right] \left[\frac{b_2\sqrt{n_2} - z_1\sqrt{n_1}}{\sqrt{n_2} - n_1} + z_\beta \right]^2 + n_1. \tag{55.8}$$

There are, however, restrictions on the range of sample size alterations that are allowable at the interim analysis. At the lower end, the CDL and extended CDL methods do not permit the sample size to be decreased below the original sample size n_2 . At the upper end there is usually a limit to the magnitude of the sample size increase that the sponsor will permit. Denote this upper limit by N_{max}^* . Then the new sample size at the time of the interim analysis is computed by the formula

$$n_2^* = \max\{n_2, \min(n_2'(z_1), N_{\text{max}}^*)\}. \tag{55.9}$$

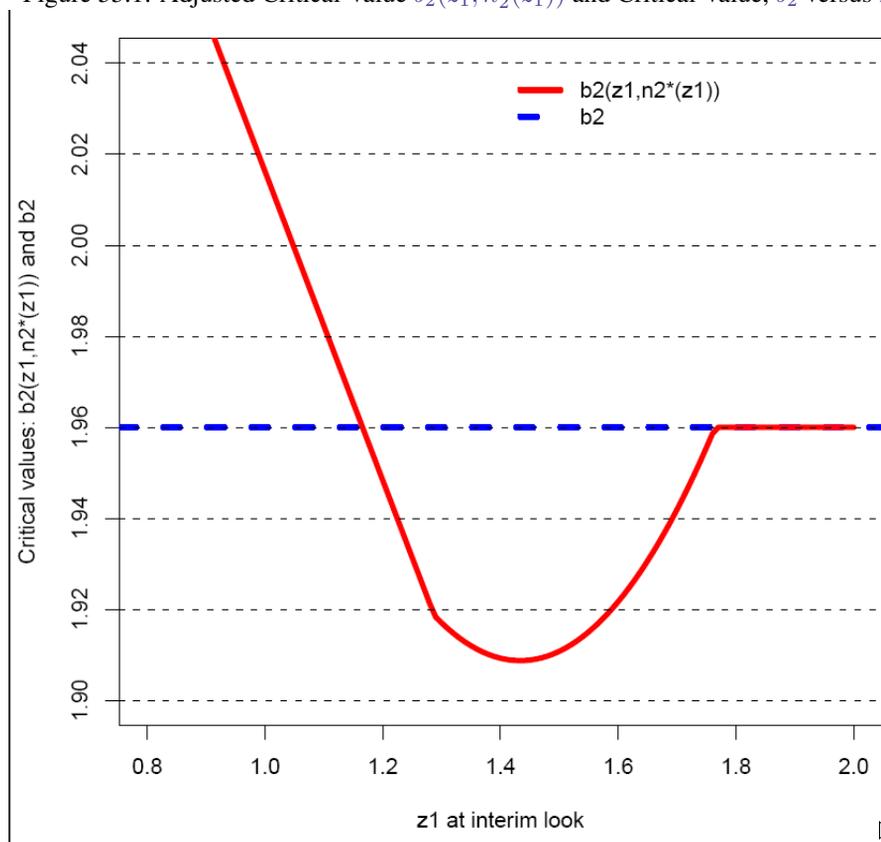
Note that $n_2^*(z_1)$ is a random variable at the start of the trial, its value being determined by the statistic z_1 obtained at the interim analysis.

By substituting (55.9) into (55.2) we can express the adjusted critical value $b_2(z_1, n_2^*)$ for the final analysis as a function of z_1 alone, and will hereafter denote it as $b_2(z_1, n_2^*(z_1))$ to show the explicit dependence of n_2^* on z_1 . Thus, we may use the criterion (55.4) for rejecting H_0 without inflating the type-1 error for the entire range of z_1 values that satisfy $b_2(z_1, n_2^*(z_1)) \leq b_2$, thereby utilizing the conventional group sequential hypothesis test at the final analysis despite a data dependent sample size

55 The Chen, DeMets and Lan Method

increase at the interim analysis. To obtain this range, it is convenient to plot $b_2(z_1, n_2^*(z_1))$ and b_2 against z_1 . Figure 55.1 displays such a plot for the two-look Schizophrenia trial that was discussed in Section 55.1.1. For this trial we have $n_1 = 208, n_2 = 442, N_{\max}^* = 884, b_1 = 5.25, b_2 = 1.96, \beta = 0.2$ and the sample size will be increased at look 1 from n_2 to $n_2^*(z_1)$ based on equation (55.9).

Figure 55.1: Adjusted Critical Value $b_2(z_1, n_2^*(z_1))$ and Critical Value, b_2 versus z_1



The curves of $b_2(z_1, n_2^*(z_1))$ and b_2 intersect at two places; at $z_{1,\min} = 1.1657$ and $z_{1,\max} = 1.7646$. Thus for all $1.0982 \leq z_1 \leq 1.7646$, we may use the conventional Wald test

$$Z_{2,\text{wald}}^* \geq b_2 \tag{55.10}$$

at the final analysis without inflating the type-1 error. To be sure we might lose some power because we are using (55.10) as our rejection criterion instead of using the less

restrictive rejection criterion

$$Z_{2,\text{wald}}^* \geq b_2(z_1, n_2^*(z_1)) \quad (55.11)$$

which also protects the type-1 error since, by (55.1) and (55.3),

$$\begin{aligned} & P_0(Z_1 \geq b_1) + P_0(Z_1 < b_1, Z_{2,\text{wald}}^* \geq b_2(z_1, n_2^*)) \\ &= P_0(Z_1 \geq b_1) + P_0(Z_1 < b_1, Z_{2,\text{chw}}^* \geq b_2) = \alpha. \end{aligned} \quad (55.12)$$

However, that is the price we must pay for using the conventional Wald test with guaranteed preservation of the type-1 error, instead of using the CHW test. In the next section we will show that the power loss is in fact negligible.

It is convenient re-scale the X-axis of Figure 55.1 in terms of conditional power. We can show that the conditional power given z_1 , evaluated at the estimated value $\hat{\delta}_1$, under the assumption that the final sample size remains unchanged at n_2 , is

$$\text{CP}_{\hat{\delta}_1}(z_1, n_2) = 1 - \Phi \left(\frac{b_2\sqrt{n_2} - z_1\sqrt{n_1}}{\sqrt{n_2 - n_1}} - \frac{z_1\sqrt{n_2 - n_1}}{\sqrt{n_1}} \right). \quad (55.13)$$

Accordingly we use equation (55.13) to transform the X-axis from z_1 to $\text{CP}_{\hat{\delta}_1}(z_1, n_2)$. Figure 55.2 is a plot of $b_2(z_1, n_2^*(z_1))$ and b_2 against $\text{CP}_{\hat{\delta}_1}(z_1, n_2)$. The curves intersect at two points which we denote as CP_{\min} and CP_{\max} . For the current example, $\text{CP}_{\min} = 0.36$ and $\text{CP}_{\max} = 0.8$. Thus for all $0.36 \leq \text{CP}_{\hat{\delta}_1}(z_1, n_2) \leq 0.8$, we may use the conventional Wald test (55.10) at the final analysis without inflating the type-1 error. The conventional Wald statistic may be used without inflating the type-1 error as long as $\text{CP}_{\hat{\delta}_1}(z_1, n_2) \geq \text{CP}_{\min}$, and the sample size is only permitted to increase (but never decreased) in accordance with (55.9).

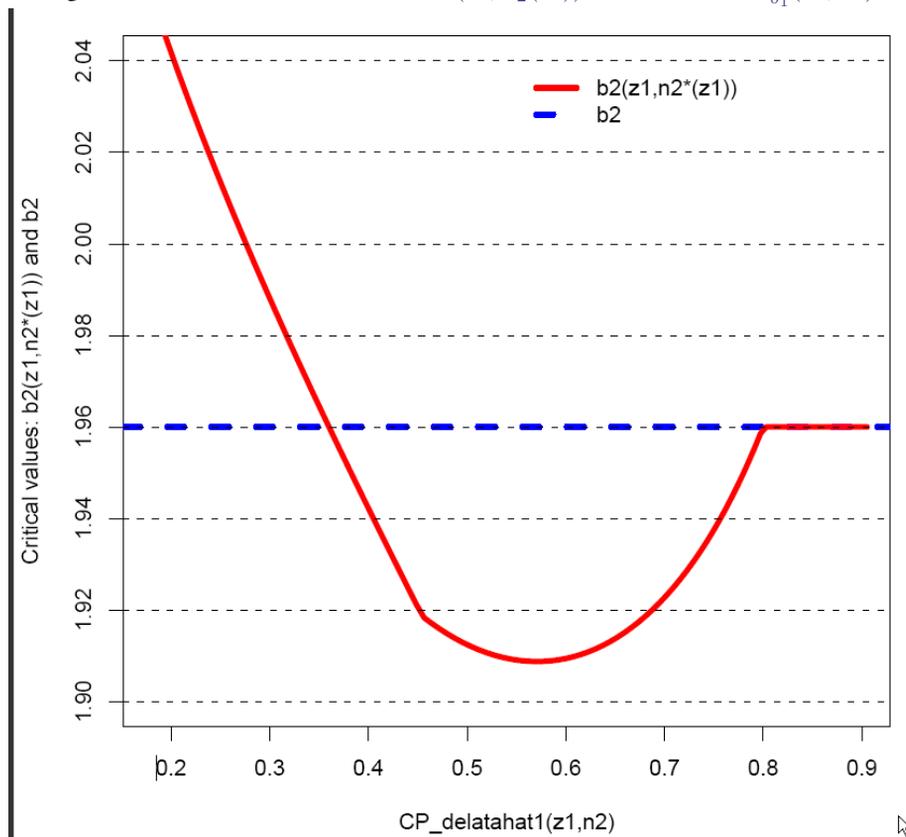
The extended CDL simulation module in the EastAdapt software accepts CP_{\min} as an input. The hypothesis test at the time of the final analysis of each simulated trial utilizes the conventional Wald criterion $Z_{2,\text{wald}}^* \geq b_2$ for rejecting H_0 if

$\text{CP}_{\hat{\delta}_1}(z_1, n_2) \geq \text{CP}_{\min}$ and utilizes the CHW criterion $Z_{2,\text{chw}}^* \geq b_2$ otherwise. Thus in all cases the type-1 error is preserved. The following is a summary of the extended CDL method:

1. Pre-specify the conditional power $1 - \beta$ that will be targeted at the time of the interim analysis
2. For a wide range of z_1 values, compute the new sample size $n_2^*(z_1)$ that would be needed to achieve the targeted conditional power, using equation (55.9)
3. Substitute $n_2^*(z_1)$ into equation (55.2) to obtain $b_2(z_1, n_2^*(z_1))$

55 The Chen, DeMets and Lan Method

Figure 55.2: Plots of Critical Values $b_2(z_1, n_2^*(z_1))$ and b_2 versus $CP_{\hat{\delta}_1}(z_1, n_2)$



4. Transform each z_1 into a corresponding conditional power $CP_{\hat{\delta}_1}(z_1, n_2)$ using equation (55.13)
5. Plot $b_2(z_1, n_2^*(z_1))$ and b_2 versus $CP_{\hat{\delta}_1}(z_1, n_2)$ and determine the value CP_{\min} where $b_2(z_1, n_2^*(z_1))$ first intersects with b_2 as shown in Figure 55.2.
6. Under the extended CDL method we can use the conventional Wald criterion $Z_{2, \text{wald}}^* \geq b_2$ to reject H_0 at the final analysis whenever $CP_{\hat{\delta}_1}(z_1, n_2) \geq CP_{\min}$.

For the convenience of the user we have pre-computed CP_{\min} cut-offs for some common two-stage, adaptive designs with no early stopping, and have displayed them in Table 55.1. All table entries are expressed as multiples of the initially proposed sample size n_2 and do not depend on the actual value of n_2 specified in the design.

One may conveniently refer to this table for suitable cut-offs instead of calculating them through the six-step procedure outlined above. For values of (n_1/n_2) or

Table 55.1: CP_{\min} Cut-Off Values for Some Typical Two-Stage Adaptive Designs with no Early Stopping either for Efficacy or Futility

| Sample Size Ratios | | CP_{\min} Values for Targeted Conditional Powers | | |
|---------------------------------------|--------------------------------|--|------|------|
| Maximum Allowed (N_{\max}^*/n_2) | At Interim Look (n_1/n_2) | 80% | 90% | 95% |
| 1.5 | 0.25 | 0.42 | 0.42 | 0.42 |
| 1.5 | 0.5 | 0.41 | 0.41 | 0.41 |
| 1.5 | 0.75 | 0.38 | 0.38 | 0.38 |
| 2 | 0.25 | 0.37 | 0.37 | 0.37 |
| 2 | 0.5 | 0.36 | 0.36 | 0.36 |
| 2 | 0.75 | 0.33 | 0.33 | 0.33 |
| 3 | 0.25 | 0.32 | 0.32 | 0.32 |
| 3 | 0.5 | 0.31 | 0.31 | 0.30 |
| 3 | 0.75 | 0.30 | 0.27 | 0.27 |
| ∞ | 0.25 | 0.32 | 0.28 | 0.26 |
| ∞ | 0.5 | 0.31 | 0.27 | 0.25 |
| ∞ | 0.75 | 0.30 | 0.25 | 0.23 |

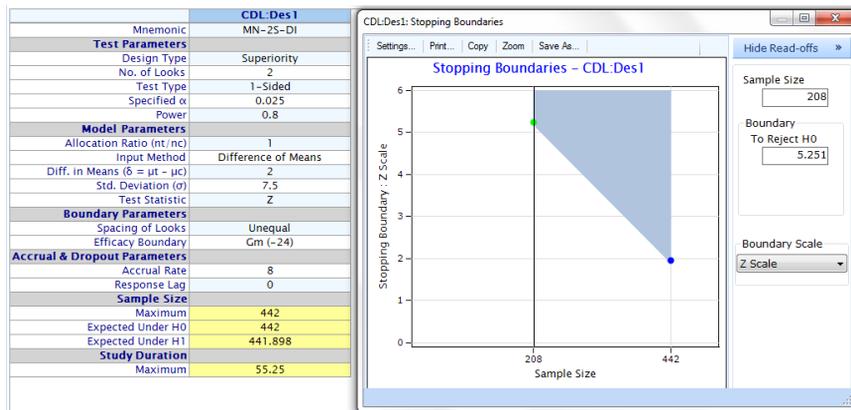
(N_{\max}^*/n_2) not included in the table, one may utilize the closest available cut-off value that guarantees conservative preservation of type-1 error. For example, for the Schizophrenia trial, $(N_{\max}^*/n_2) = 2$ and $(n_1/n_2) = 0.47$. Table 55.1 shows that the cut-off value $CP_{\min} = 0.37$ will preserve the type-1 error conservatively for targeted conditional powers of 80%, 90% or 95%. Observe that $CD_{\min} < 0.5$ for all the entries in Table 55.1 thus demonstrating that the extended CDL method is a relaxation of the original CDL method.

55.2.2 Normal Endpoint: Schizophrenia Trial

Consider again the Schizophrenia example introduced in Section 55.1.1. This is a two-look design with an initially specified sample size $n_2 = 442$ and one interim look after seeing data on $n_1 = 208$ completers. The stopping boundaries at the interim and final look are one-sided level-0.025 efficacy boundaries derived from the $\gamma(-24)$ spending function, which for all practical purposes implies that there will be no early stopping for efficacy. There is no futility boundary. This trial has slightly over 80% power to detect $\delta = 2$, given a standard deviation of $\sigma = 7.5$. The East design

55 The Chen, DeMets and Lan Method

screenshot is reproduced below.



Suppose that at the time of the interim analysis the sample size may be increased up to a maximum of $N_{\max}^* = 884$ in an attempt to attain a target conditional power of 95%. Assume that the sample size will only be increased (never decreased) if at the interim analysis the observed z_1 is such that $0.5 \leq CP_{\hat{\delta}_1}(z_1, 442) < 0.8$, identified as the *promising zone* for the interim results.

We will simulate the trial under different assumptions about δ and σ , using the extended CDL criterion instead of the original CDL criterion. To do this we need to know CP_{\min} , the value of $CP_{\hat{\delta}_1}(z_1, n_2)$ at which the adjusted critical value $b_2(z_1, n_2^*(z_1))$ starts to dip below the critical value $b_2 = 1.96$. To obtain the exact cut-off value we would have to manually execute the six-step procedure outlined at the end of Section 55.2.1. An easier alternative is to use the cut-off values provided in Table 55.1 for the standard two-stage designs. The difference in the operating characteristics of the design produced the two methods is negligible. Here $(N_{\max}^*/n_2) = 2$, $(n_1/n_2) = 0.47$ and the targeted conditional power is 0.95. Since there is no entry in Table 55.1 for this choice of parameters we use the more conservative choices, $(N_{\max}^*/n_2) = 2$ and $(n_1/n_2) = 0.25$, whereupon $CP_{\min} = 0.37$ for a targeted conditional power of 95%.

Suppose we wish to obtain the power of the above adaptive design under $\delta = 1.6$ and $\sigma = 7.5$. We therefore save the design in **Library**, insert Simulations for this design and enter the following parameters into different tabs: The **Response Generation Info**

tab:

| | |
|---------------------------------------|---|
| Generate Data Using: Individual Means | <input checked="" type="checkbox"/> Common Standard Deviation |
| Mean Control (μ_c): 0 | SD Control (σ_c): 7.5 |
| Mean Treatment (μ_t): 1.6 | SD Treatment (σ_t): 7.5 |

The **Sample Size Re-estimation** tab:

| | |
|--|--------------------------------------|
| Use Adaptation Method | |
| <input type="radio"/> CHW | <input checked="" type="radio"/> CDL |
| <input type="radio"/> Müller and Schäfer | |
| Adapt at: Look # | 1 |
| Max. Sample Size if Adapt (multiplier; total #): | 2 884 |
| Upper Limit on Study Duration: | 165.75 |
| Target CP for Re-estimating Sample Size: | 0.95 |
| Use Wald Stat. if CP >= | 0.37 |
| Promising Zone Scale: | Cond. Power CP |
| Promising Zone: | Min. CP: 0.37 |
| | Max. CP: 0.9 |
| CP Computation Based on: | Estimated δ/σ |
| Accrual Rate After Adaptation: | No Change |

and the **Simulation Control Info** tab:

| | |
|--|---|
| Number of Simulations: 100000 | Output Options |
| Refresh Frequency: 10000 | Output Type: Case Data |
| Random Number Seed | <input type="checkbox"/> Save summary statistics for every simulation run |
| <input checked="" type="radio"/> Clock | <input type="checkbox"/> Save subject-level data for [] simulation runs |
| <input type="radio"/> Fixed 100 | Note: Max. 100,000 records will be saved. |

We will now run 100,000 simulations at $\delta = 1.6$ and $\sigma = 7.5$. Upon clicking the

55 The Chen, DeMets and Lan Method

Simulate button, the simulations are activated. The results are shown below.

⊖ Average Sample Size and Look Times

| Look # | Average Sample Size (n) | Average Look Time |
|---------|-------------------------|-------------------|
| 1 | 208.001 | 25.939 |
| 2 | 557.057 | 69.478 |
| Average | 557.001 | 69.471 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | | Count | % |
| 1 | 208 | 5.251 | 16 | 16 | 0.016% |
| 2 | 442 | 1.96 | 68603 | 99984 | 99.984% |
| Total | | | 68619 | 100000 | |
| % | | | 68.619% | | |

The null hypothesis was rejected a total of 68619 times in 100,000 trials for an overall power of 68.6%. The average sample size was 557.0. The top part of the simulation output, shows the zone by zone results and the results conditional on falling in the promising zone and thereby undergoing a sample size increase. This occurred 30,890 times out of 100,000 simulations. Moreover 27,805 of these trials rejected the null hypothesis for a power of 90.013%. The expected sample size of all trials that underwent a sample size increase was 814.413.

⊖ Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|-------------|--------------------------|----------|------------------------------|---------|-------------------|----------|---------------------|
| | Count | Row % | Count | Row % | Count | Column % | |
| Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| Unfavorable | 11035 | 30.572% | 25060 | 69.428% | 36095 | 36.095% | 442 |
| Promising | 27805 | 90.013% | 3085 | 9.987% | 30890 | 30.890% | 814.413 |
| Favorable | 29763 | 90.194% | 3236 | 9.806% | 32999 | 32.999% | 442 |
| Efficacy | 16 | 100.000% | 0 | 0.000% | 16 | 0.016% | 208 |
| All Trials | 68619 | 68.619% | 31381 | 31.381% | 100000 | 100.000% | 557.001 |

Promising Zone defined as $0.37 \leq CP < 0.9$

As before, it is of interest to verify that the type-1 error is preserved by the extended CDL method. Accordingly we set $\delta = 0$ in the **Mean Treatment** μ_t . Rest all

parameters are unchanged.

| | |
|---------------------------------------|---|
| Generate Data Using: Individual Means | <input checked="" type="checkbox"/> Common Standard Deviation |
| Mean Control (μ_c): 0 | SD Control (σ_c): 7.5 |
| Mean Treatment (μ_t): 0 | SD Treatment (σ_t): 7.5 |

The results from 100,000 simulations are shown below.

⊖ Average Sample Size and Look Times

| Look # | Average Sample Size (n) | Average Look Time |
|---------|-------------------------|-------------------|
| 1 | 208.001 | 25.939 |
| 2 | 479.899 | 59.852 |
| Average | 479.899 | 59.852 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|----------|
| | | Efficacy | | Efficacy | Count |
| 1 | 208 | 5.251 | 0 | 0 | 0.000% |
| 2 | 442 | 1.96 | 2356 | 100000 | 100.000% |
| Total | | | 2356 | 100000 | |
| % | | | 2.356% | | |

It is seen that only 2356 of the 100,000 simulations were able to reject the null hypothesis, for a type-1 error of 0.02356.

Now in addition to setting $\delta = 0$ in the **Difference of Means**, we set the **Promising Zone: Min CP** to zero as well, so as to provide the simulations with the largest possible opportunity to increase the sample size and thereby inflate the

55 The Chen, DeMets and Lan Method

type-1 error.

Use Adaptation Method
 CHW CDL Müller and Schäfer

Adapt at: Look # 1

Max. Sample Size if Adapt (multiplier; total #): 2 884

Upper Limit on Study Duration: 165.75

Target CP for Re-estimating Sample Size: 0.95

Use Wald Stat. if CP >= 0.37

Promising Zone Scale: Cond. Power CP

| | | |
|-----------------|----------|-----|
| Promising Zone: | Min. CP: | 0 |
| | Max. CP: | 0.9 |

CP Computation Based on: Estimated δ/σ

Accrual Rate After Adaptation: No Change

The results from 100,000 simulations are shown below.

⊖ Average Sample Size and Look Times

| Look # | Average Sample Size (n) | Average Look Time |
|---------|-------------------------|-------------------|
| 1 | 208.001 | 25.939 |
| 2 | 648.779 | 80.914 |
| Average | 648.779 | 80.914 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|----------|
| | | Efficacy | | Efficacy | Count |
| 1 | 208 | 5.251 | 0 | 0 | 0.000% |
| 2 | 442 | 1.96 | 2267 | 100000 | 100.000% |
| Total | | | 2267 | 100000 | |
| % | | | 2.267% | | |

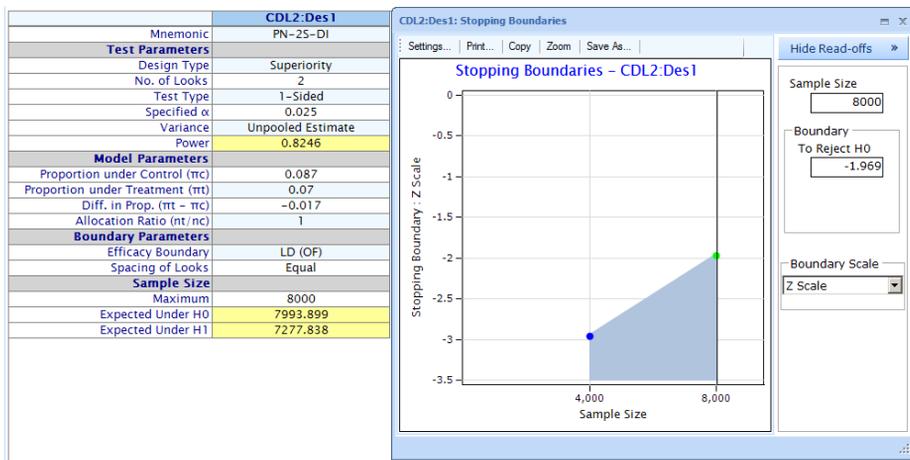
It is seen that only 2267 of the 100,000 simulations were able to reject the null hypothesis, for a type-1 error of 0.02267, preserving type-1 error of 0.025.

55.2.3 Binomial Endpoint: Acute Coronary Syndromes Trial

Consider again a two-arm, placebo controlled randomized clinical trial for subjects with acute cardiovascular disease undergoing percutaneous coronary intervention (PCI), which we discussed in Section 54.4. The primary endpoint in this study is a composite of death, myocardial infarction or ischemia-driven revascularization during the first 48 hours after randomization. We assume on the basis of prior knowledge that the event rate for the placebo arm is 8.7%. The investigational drug is expected to reduce the event rate by at least 20%. The investigators are planning to randomize a total of 8000 subjects in equal proportions to the two arms of the study.

As explained in the beginning of this chapter, for applying CDL method, a 2 look group sequential design will suffice, without loss of generality.

It is easy to show that a group sequential design enrolling a total of 8000 subjects with an interim look after 4000 subjects are enrolled (50% of total information), will have 82% power to detect a 20% risk reduction with a one-sided level-0.025 test of significance, and early stopping efficacy boundary derived from the Lan and DeMets (1983) O’Brien-Fleming type error spending function.



Suppose that at the time of the interim analysis the sample size may be increased up to a maximum of $N_{\max}^* = 16000$ in an attempt to attain a target conditional power of 95%. Assume that the sample size will only be increased (never decreased) if at the interim analysis the observed z_1 is such that $0.5 \leq CP_{\delta_1}(z_1, 8000) < 0.9$, identified as the *promising zone* for the interim results. We will simulate the trial under different assumptions about δ and σ , using the extended CDL criterion instead of the original

55 The Chen, DeMets and Lan Method

CDL criterion. To do this we need to know CP_{\min} , the value of $CP_{\delta_1}(z_1, n_2)$ at which the adjusted critical value $b_2(z_1, n_2^*(z_1))$ starts to hover above the critical value $b_2 = -1.9686$. To obtain the exact cut-off value we would have to manually execute the six-step procedure outlined at the end of Section 55.2.1. An easier alternative is to use the cut-off values provided in Table 55.1 for the standard two-stage designs. The difference in the operating characteristics of the design produced the two methods is negligible. Here $(N_{\max}^*/n_2) = 2$, $(n_1/n_2) = 0.50$ and the targeted conditional power is 0.95. There is an entry in Table 55.1 for this choice of parameters, $(N_{\max}^*/n_2) = 2$ and $(n_1/n_2) = 0.50$, whereupon $CP_{\min} = 0.36$ for a targeted conditional power of 95%. Suppose, we wish to obtain the power of the above adapted design, for example, at risk reduction $\rho = 0.15$ and to increase the sample size only if the conditional power at the interim analysis under the original sample size is between 0.36 and 0.9. And in that case suppose that we wish to increase the sample size by just the right amount so that the conditional power is boosted to 0.95. Furthermore suppose that the re-estimated sample size is constrained to remain between 8000 and 16000 subjects. To run the simulations with these specifications we would change the entries in simulation tabs as shown below.

The **Response Generation Info** tab:

| Specify Proportion | |
|------------------------------------|-------|
| Prop. under Control (π_c): | 0.087 |
| Prop. under Treatment (π_t): | 0.074 |

The **Sample Size Re-estimation** tab:

| | | |
|--|--------------------------------------|--|
| Use Adaptation Method | | |
| <input type="radio"/> CHW | <input checked="" type="radio"/> CDL | <input type="radio"/> Müller and Schäfer |
| Adapt at: | Look # | 1 |
| Max. Sample Size if Adapt (multiplier; total #): | 2 | 16000 |
| Target CP for Re-estimating Sample Size: | 0.95 | |
| Use Wald Stat. if CP >= | 0.36 | |
| Promising Zone Scale: | Cond. Power | CP |
| Promising Zone: | Min. CP: | 0.36 |
| | Max. CP: | 0.9 |
| CP Computation Based on: | Estimated (n_c, n_t) | |

and the **Simulation Control Info** tab:

| | |
|---|-------------------|
| Number of Simulations: | 100000 |
| Refresh Frequency: | 10000 |
| Random Number Seed | |
| <input checked="" type="radio"/> Clock | |
| <input type="radio"/> Fixed | 100 |
| Output Options | |
| Output Type: | Case Data |
| <input type="checkbox"/> Save summary statistics for every simulation run | |
| <input type="checkbox"/> Save subject-level data for | 1 simulation runs |
| Note: Max. 100,000 records will be saved. | |

We run the simulations by pressing the **Simulate** button. The results are as shown

55 The Chen, DeMets and Lan Method

below.

⊖ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 4000 |
| 2 | 10332.703 |
| Average | 9877.509 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | | Count | % |
| | | Lower | Efficacy | | |
| 1 | 4000 | -2.963 | 7188 | 7188 | 7.188% |
| 2 | 8000 | -1.969 | 57851 | 92812 | 92.812% |
| Total | | | 65039 | 100000 | |
| % | | | 65.039% | | |

The null hypothesis was rejected 65,039 times in 100,000 trials for an overall power of 65 %. The average sample size was 9877.5. In contrast, if there is no sample size increase, the power would be 57% and the average sample size would be 8000.

The top part of the simulation output displayed below, shows the zone by zone results as well as results conditional on falling in the promising zone and thereby undergoing a sample size increase. This occurred 31,855 times out of 100,000 simulations. Moreover 28,086 of these trials rejected the null hypothesis for a power of 88.2%. The expected sample size of all trials that underwent a sample size increase was 14,796.5.

⊖ Zone-wise Averages

| | Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|---|-------------|--------------------------|----------|------------------------------|---------|-------------------|----------|---------------------|
| | | Count | Row % | Count | Row % | Count | Column % | |
| ⊖ | Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| | Unfavorable | 9702 | 25.604% | 28191 | 74.396% | 37893 | 37.893% | 8000 |
| | Promising | 28086 | 88.168% | 3769 | 11.832% | 31855 | 31.855% | 14796.511 |
| ⊖ | Favorable | 20063 | 86.988% | 3001 | 13.012% | 23064 | 23.064% | 8000 |
| | Efficacy | 7188 | 100.000% | 0 | 0.000% | 7188 | 7.188% | 4000 |
| | All Trials | 65039 | 65.039% | 34961 | 34.961% | 100000 | 100.000% | 9877.509 |

Promising Zone defined as $0.36 \leq CP < 0.9$

As before, it is of interest to verify that the type-1 error is preserved by the extended

CDL method. Accordingly we set treatment proportion same as control proportion =0.087, thereby making $\delta = 0$. Rest all parameters are unchanged.

Specify Proportion

Prop. under Control (π_c):

Prop. under Treatment (π_t):

The results from 100,000 simulations are shown below.

⊖ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 4000 |
| 2 | 8678.491 |
| Average | 8670.958 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | | Count | % |
| | | Lower | Efficacy | | |
| 1 | 4000 | -2.963 | Efficacy | 161 | 0.161% |
| 2 | 8000 | -1.969 | Efficacy | 99839 | 99.839% |
| Total | | | Efficacy | 100000 | |
| % | | | Efficacy | 2320 | 2.320% |

It is seen that only 2320 of the 100,000 simulations were able to reject the null hypothesis and hence the simulated type-1 error is 0.0232.

On the other hand, suppose we perform the very same simulations but set the **Use Wald Stat. if CP(8000) >=** parameter to zero and **Promising Zone: Min CP** to zero as well. There is now no protection against type-1 error inflation. As seen below, 2679 of 100,000 simulations with this change rejected the null hypothesis

55 The Chen, DeMets and Lan Method

giving us a type-1 error of 0.02679.

⊖ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 4000 |
| 2 | 11690.227 |
| Average | 11679.23 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|---------|
| | | Efficacy | | Efficacy | Count |
| | | Lower | | | |
| 1 | 4000 | -2.963 | 143 | 143 | 0.143% |
| 2 | 8000 | -1.969 | 2536 | 99857 | 99.857% |
| Total | | | 2679 | 100000 | |
| % | | | 2.679% | | |

55.2.4 Survival Endpoint: Lung Cancer Trial

The statistical methodology described in Section 55.2.1 for normal and binomial endpoints applies also to survival endpoints with appropriate changes in notation as described in Chapter 54, Section 54.2. To see this, carry out CDL simulations of Des 1, the lung cancer example discussed earlier in Section 55.1.3 of this chapter.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

| Test Parameters | |
|---|--------------|
| Design ID | Des1 |
| Design Type | Superiority |
| Number of Looks | 2 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Attained α | 0.025 |
| Power | 0.901 |
| Model Parameters | |
| HR = λ_1/λ_0 | |
| Under H0 | 1 |
| Under H1 | 0.7 |
| Ratio of Medians: λ_1/λ_0 | 1.429 |
| Var (Log HR) | Null |
| Allocation Ratio (n_1/n_0) | 1 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Futility Boundary | Gm (-5) (VB) |
| Accrual/Dropout Parameters | |
| Accrual Duration | 24 |
| Max Study Duration | 30 |
| Dropout | No |

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|------------------------------|-------------|---------------|---------|
| Sample Size (n) | | | |
| Maximum | 241 | 242 | 483 |
| Expected H1 | 226.836 | 226.836 | 453.671 |
| Expected H0 | 210.851 | 210.851 | 421.701 |
| Events (e) | | | |
| Maximum | 180 | 154 | 334 |
| Expected H1 | 162.827 | 135.155 | 297.982 |
| Expected H0 | 130.437 | 130.437 | 260.874 |
| Maximum Information (i) 83.9 | | | |

Accrual and Study Duration

| | Accrual Duration | Study Duration |
|-------------|------------------|----------------|
| Maximum | 24 | 29.984 |
| Expected H1 | 22.543 | 26.954 |
| Expected H0 | 20.954 | 22.92 |

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (s/s_max) | Events (e) | Cumulative α Spent | Cumulative β Spent | Boundaries | |
|--------|--------------------------|------------|---------------------------|--------------------------|------------|------------|
| | | | | | Efficacy Z | Futility Z |
| 1 | 0.5 | 167 | 0.002 | 0.008 | -2.963 | 0.126 |
| 2 | 1 | 334 | 0.025 | 0.099 | -1.969 | -1.969 |

Events, Sample Size, Pipeline and Analysis Times: Look by Look (Under H0)

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (e) | Pipeline (n-s) | Analysis Time | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|------------|----------------|---------------|---|----------|
| | | | | | | Efficacy | Futility |
| 1 | 0.5 | 348 | 167 | 181 | 17.251 | 0.002 | 0.45 |
| 2 | 1 | 493 | 334 | 149 | 27.583 | 0.023 | 0.525 |

Events, Sample Size, Pipeline and Analysis Times: Look by Look (Under H1)

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (e) | Pipeline (n-s) | Analysis Time | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|------------|----------------|---------------|---|----------|
| | | | | | | Efficacy | Futility |
| 1 | 0.5 | 372 | 167 | 205 | 18.455 | 0.255 | 0.008 |
| 2 | 1 | 493 | 334 | 149 | 29.984 | 0.645 | 0.092 |

Survival Information: Median Survival Times

| Median Survival Times | | Hazard Ratio |
|-----------------------|----------------|--------------|
| Control (Ac) | Treatment (AT) | Alt. (AT)/Ac |
| 8 | 11.429 | 0.7 |

Accrual Information

| Period # | Starting at Time | Cum. % Accrued |
|----------|------------------|----------------|
| 1 | 24 | 100 |

Variable Follow-Up Design: All subjects are followed until failure, drop out or end of study.

We will simulate this design using CDL method. To do this, insert simulations for this design and add the **Sample Size Re-estimation** tab.

In the **Response Generation Info** tab, set the hazard ratio to 1 as shown below, so as

55 The Chen, DeMets and Lan Method

to simulate under the null hypothesis.

| Survival Information | | | | |
|--|------------------|---------------|-----------|--------------|
| # of Hazard Pieces | 1 | Input Method: | | Hazard Rates |
| <input checked="" type="checkbox"/> Hazard Ratio | | | | |
| Hazard Ratio (λ_t/λ_c): | | 1 | | |
| Piece | Starting at Time | Hazard Rates | | Hazard Ratio |
| | | Control | Treatment | |
| 1 | 0.000 | 0.0866 | 0.0866 | 1.000 |

Now go to the **Sample Size Re-estimation** tab and enter the following values:

| | | | |
|--|--------------|---|-----------|
| Use Adaptation Method | | <input type="radio"/> CHW <input checked="" type="radio"/> CDL <input type="radio"/> Müller and Schäfer | |
| Adapt at: | Look # | 1 | |
| Max. # of Events if Adapt (multiplier; total #): | 10 | 3340 | |
| Max. Sample Size if Adapt (multiplier; total #): | 10 | 4830 | |
| Upper Limit on Study Duration: | 500 | | |
| Target CP for Re-estimating # of Events: | 0.9 | | |
| Use Wald Stat. if CP >= | 0 | | |
| Promising Zone Scale: | Cond. Power | | CP |
| Promising Zone: | Min. CP: | 0 | |
| | Max. CP: | 0.9 | |
| CP Computation Based on: | Estimated HR | | |
| Accrual Rate After Adaptation: | No Change | | |

We can verify by simulating 10,000 times with these input parameters that the type-1 error will not be preserved because the input parameter **Use Wald Statistic if CP >=** has been set to 0 instead of being at 0.5. Consequently the CDL condition, required for preserving the type-1 error if the conventional Wald statistic is being used with a data dependent increase in the number of events, is not satisfied. Hit the **Simulate** button to obtain the following output:

Row 6 of the table of Zone-wise Averages displays the results for all trials, combined across zones. Thus Column 4 of Row 6 of this table displays the magnitude of the type-1 error, 0.029 which is seen to exceed 0.025 even after accounting for Monte Carlo error.

⊖ Zone-wise Averages

| | Zone | Simulations Rejecting H0 | |
|---|-------------|--------------------------|----------|
| | | Count | Row % |
| ⊖ | Futility | 0 | 0.000% |
| | Unfavorable | 0 | 0.000% |
| | Promising | 211 | 3.996% |
| ⊖ | Favorable | 61 | 30.964% |
| | Efficacy | 17 | 100.000% |
| | All Trials | 289 | 2.890% |

Table 55.1 shows cut-off values of conditional power below 0.5 at which the use of the conventional Wald statistic will preserve the type-1 error. There is no entry in this table for a Sample Size Ratio (i.e., event multiplier) of 10. However a 10-fold multiplier is for all practical purposes the same as an infinite multiplier. Table 55.1 shows that for a Sample Size Ratio equal to ∞ (infinite multiplier), the cut-off for a trial powered at 90% with and interim analysis at 50% of the information is 0.27. Let us therefore use a

55 The Chen, DeMets and Lan Method

cut-off of 0.27 for the simulations instead of a zero cut-off as was done previously.

Use Adaptation Method CHW CDL Müller and Schäfer

Adapt at:

Max. # of Events if Adapt (multiplier; total #):

Max. Sample Size if Adapt (multiplier; total #):

Upper Limit on Study Duration:

Target CP for Re-estimating # of Events:

Use Wald Stat. if CP >=

Promising Zone Scale:

Promising Zone:

CP Computation Based on:

Accrual Rate After Adaptation:

Now run the 10,000 simulations once again. This time it is seen that the type-1 error is preserved. The simulated alpha is 0.0208.

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration (CDL Simulation)

| Simulation Parameters | |
|--------------------------------------|---------------------------------|
| Simulation ID | CDLSim2 |
| Design Type | Superiority |
| Number of Looks | 2 |
| Test Type | 1-Sided |
| Sample Size (n) | 483 |
| Fix at Each Look | Total No. of Events |
| Test Statistic | Logrank |
| Average Events | 1369.405 |
| Total Accrual Duration | 24 |
| Avg. Power at Termination | 0.021 |
| Sample Size Re-estimation Parameters | |
| Method of Adaptation | Chen-DeMets-Lan |
| Adapt At Look No. | 1 |
| Max. # of Events if Adapt | Multiplier: 10 Total #: 3340 |
| Max. Sample Size if Adapt | Multiplier: 10 Total #: 4830 |
| Upper Limit on Study Duration | 500 |
| Target CP | 0.9 |
| Use Wald Stat. if CP >= | 0.27 |
| Promising Zone Scale | Cond. Power |
| Min. CP | 0 |
| Max. CP | 0.9 |
| CP Computation Based on | Estimated HR |
| Accrual Rate After Adaptation | No Change |
| Simulation Control Parameters | |
| Starting Seed | Click |
| Number of Simulations | 10000 |

Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Rejecting H1 | | Total Simulations | | Average Sample Size | Average Number of Events | Average Accrual Duration | Average Study Duration |
|-------------|--------------------------|----------|--------------------------|----------|-------------------|----------|---------------------|--------------------------|--------------------------|------------------------|
| | Count | Row % | Count | Row % | Count | Column % | | | | |
| Futility | 0 | 0.000% | 4532 | 100.000% | 4532 | 45.320% | 347.124 | 167 | 17.185 | 17.234 |
| Unfavorable | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 | 0 | 0 | 0 |
| Promising | 127 | 2.422% | 5116 | 97.578% | 5243 | 52.430% | 2566.804 | 2339.098 | 121.381 | 121.788 |
| Favorable | 70 | 32.710% | 144 | 67.290% | 214 | 2.140% | 483 | 334 | 23.955 | 27.528 |
| Efficacy | 11 | 100.000% | 0 | 0.000% | 11 | 0.110% | 352.636 | 167 | 17.388 | 17.464 |
| All Trials | 208 | 2.080% | 5792 | 97.920% | 10000 | 100.000% | 1513.816 | 1309.465 | 71.96 | 72.272 |

Promising Zone defined as $0 <= CP < 0.9$

Average Sample Size and Look Times

| Look # | Average Sample Size | Control | Treatment | Average Look Time | Average Follow up |
|---------|---------------------|----------|-----------|-------------------|-------------------|
| 1 | 347 | 83.64 | 83.36 | 17.225 | 5.548 |
| 2 | 2485.098 | 1131.066 | 1129.401 | 118.091 | 9.925 |
| Average | 1513.816 | 653.698 | 655.719 | 72.272 | 7.939 |

Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Events | Boundaries | | Stopping For | | Total Simulations | |
|--------|--------|------------|----------|--------------|----------|-------------------|---------|
| | | Efficacy | Futility | Efficacy | Futility | Count | % |
| 1 | 167 | -2.963 | 0.126 | 11 | 4532 | 4543 | 45.430% |
| 2 | 334 | -1.969 | -1.969 | 197 | 5260 | 5457 | 54.570% |
| Total | 208 | | | 208 | 9792 | 10000 | |
| % | | | 2.080% | 97.920% | | | |

Thus the extended CDL method permits a lower cut-off than 0.5 and may be used to design studies with a wider range of promising zones while permitting the use of a conventional Wald statistic for the final analysis without type-1 error inflation.

55.3 Efficiency Considerations

At the beginning of this chapter we cited some theoretical results by Tsiatis and Mehta (2003) and Jennison and Turnbull (2006) who demonstrated that the use of the CHW statistic instead of the conventional Wald statistic to perform hypothesis tests in a group sequential clinical trial with sample size changes can lead to loss of efficiency. These results, however, involved extremely large sample size increases (up to tenfold) and numerous interim looks at the accruing data. It would thus be of interest to determine whether the CHW statistic also loses power relative to the conventional Wald statistic for the more common situation of a two-stage clinical trial with at most a doubling of the sample size if the interim results fall in a promising zone. The CHW and CDL simulation worksheets provides us with the tools to make the relevant comparisons. We will accordingly compare the operating characteristics of the two-stage schizophrenia trial when the CHW test, the CDL test and the conventional Wald test are utilized for the final analysis. The design specifications for this trial were provided at the beginning of Section 55.1.1 of this chapter. The trial has a planned enrollment of 442 subjects and an interim analysis after seeing data on 208 completers. The main purpose of the interim analysis is to decide whether to increase the sample size, not to stop early for efficacy. Consequently the conservative $\gamma(-24)$ error spending function is utilized at the interim analysis. The sample size may be increased up to a maximum of 884 subjects so as to recover a target conditional power of 80%, provided the interim results fall in a promising zone. The promising zone is defined by $0.3 \leq \text{CP}(442) < 0.8$ where $\text{CP}(442)$ is the conditional power at the interim look (based on the estimated value of δ/σ) assuming no change in the initially specified sample size of 442 subjects.

We shall compare power and expected sample size of all three methods (CHW, CDL, conventional Wald) for $\delta = 0, 1, 1.6, 1.8, 2$ assuming $\sigma = 7.5$. The CHW method utilizes the CHW simulation worksheet. The following are the simulation parameters for simulating under $\delta = 0$.

The **Response Generation Info** tab:

| | | |
|-----------------------------|------------------|---|
| Generate Data Using: | Individual Means | <input checked="" type="checkbox"/> Common Standard Deviation |
| Mean Control (μ_c): | 0 | SD Control (σ_c): 7.5 |
| Mean Treatment (μ_t): | 0 | SD Treatment (σ_t): 7.5 |

55 The Chen, DeMets and Lan Method

The **Sample Size Re-estimation** tab:

| | | | |
|--|---------------------------|--|-----|
| Use Adaptation Method | | Weights... | |
| <input checked="" type="radio"/> CHW | <input type="radio"/> CDL | <input type="radio"/> Müller and Schäfer | |
| Adapt at: | Look # | 1 | |
| Max. Sample Size if Adapt (multiplier; total #): | | 2 | 884 |
| Target CP for Re-estimating Sample Size: | | 0.8 | |
| Promising Zone Scale: | | Cond. Power | CP |
| Promising Zone: | Min. CP: | 0.3 | |
| | Max. CP: | 0.8 | |
| CP Computation Based on: | | Estimated δ/σ | |

and the **Simulation Control Info** tab:

| | |
|---|---------|
| Number of Simulations: | 1000000 |
| Refresh Frequency: | 100000 |
| Random Number Seed | |
| <input checked="" type="radio"/> Clock | |
| <input type="radio"/> Fixed | 100 |
| <input type="checkbox"/> Suppress All Intermediate Output | |
| <input type="checkbox"/> Pause after Refresh | |
| <input checked="" type="checkbox"/> Stop At End | |

The results for 1,000,000 simulations are shown below.

Simulation: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means (CHW Simulation)

| Simulation Parameters | |
|--------------------------------------|---------------------------|
| Simulation ID | CHWSim1 |
| Design Type | Superiority |
| Number of Looks | 2 |
| Test Type | 1-Sided |
| Sample Size (n) | 442 |
| Variance | Equal |
| Test Statistic | 1 |
| Avg. Power at Termination | 0.026 |
| Response Generation Parameters | |
| Generate Data Using | Individual Means |
| Mean Control (μ_c) | 0 |
| Mean Treatment (μ_t) | 0 |
| SD Control (σ_c) | 7.5 |
| SD Treatment (σ_t) | 7.5 |
| Sample Size Re-estimation Parameters | |
| Method of Adaptation | Cui-Hung-Wang |
| Adapt At Look No. | 1 |
| Max. Sample Size if Adapt | |
| Multiplier | 2 |
| Total # | 884 |
| Target CP | 0.8 |
| Promising Zone Scale | Cond. Power |
| Min. CP | 0.3 |
| Max. CP | 0.8 |
| CP Computation Based on | Estimated δ/σ |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 1000000 |

☉ Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|-------------|--------------------------|---------|------------------------------|---------|-------------------|----------|---------------------|
| | Count | Row % | Count | Row % | Count | Column % | |
| ☉ Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| Unfavorable | 6437 | 0.748% | 854161 | 99.252% | 860598 | 86.060% | 442 |
| Promising | 8634 | 8.632% | 91394 | 91.368% | 100028 | 10.003% | 741.733 |
| ☉ Favorable | 10471 | 26.594% | 28903 | 73.406% | 39374 | 3.937% | 442 |
| Efficacy | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| All Trials | 25542 | 2.554% | 974458 | 97.446% | 1000000 | 100.000% | 471.982 |

Promising Zone defined as $0.3 \leq CP < 0.8$

☉ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 208 |
| 2 | 471.982 |
| Average | 471.982 |

☉ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|----------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 208 | 5.251 | 0 | 0 | 0.000% |
| 2 | 442 | 1.96 | 25542 | 1000000 | 100.000% |
| Total | | | 25542 | 1000000 | |
| % | | | 2.554% | | |

The null hypothesis was rejected 25542 times in 1,000,000 trials, comfortably within the range of Monte Carlo accuracy for a level 0.025 test. Simulation results for other values of δ are displayed in Table 55.2.

The CDL method utilizes the following simulation parameters for simulating under

55 The Chen, DeMets and Lan Method

$$\delta = 0.$$

| Use Adaptation Method | | |
|--|--------------------------------------|--|
| <input type="radio"/> CHW | <input checked="" type="radio"/> CDL | <input type="radio"/> Müller and Schäfer |
| Adapt at: | Look # | 1 |
| Max. Sample Size if Adapt (multiplier; total #): | 2 | 884 |
| Target CP for Re-estimating Sample Size: | 0.8 | |
| Use Wald Stat. if CP >= | 0.5 | |
| Promising Zone Scale: | Cond. Power | <input type="button" value="CP"/> |
| Promising Zone: | Min. CP: | 0.3 |
| | Max. CP: | 0.8 |
| CP Computation Based on: | Estimated δ/σ | |

Notice that the CDL parameters input tab stipulates that the conventional Wald statistic will be used if $CP \geq 0.5$. This is the CDL criterion (Chen, DeMets and Lan, 2004) for guaranteeing that the type-1 error will be preserved. The following are the results for

1,000,000 simulated trials under $\delta = 0$.

Simulation: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means (CDL Simulation)

| Simulation Parameters | |
|--------------------------------------|---------------------------|
| Simulation ID | CDLSim1 |
| Design Type | Superiority |
| Number of Looks | 2 |
| Test Type | 1-Sided |
| Sample Size (n) | 442 |
| Variance | Equal |
| Test Statistic | t |
| Avg. Power at Termination | 0.025 |
| Response Generation Parameters | |
| Generate Data Using | Individual Means |
| Mean Control (μ_c) | 0 |
| Mean Treatment (μ_t) | 0 |
| SD Control (σ_c) | 7.5 |
| SD Treatment (σ_t) | 7.5 |
| Sample Size Re-estimation Parameters | |
| Method of Adaptation | Chen-DeMets-Lan |
| Adapt At Look No. | 1 |
| Max. Sample Size if Adapt | |
| Multiplier | 2 |
| Total # | 884 |
| Target CP | 0.8 |
| Use Wald Stat. if CP >= | 0.5 |
| Promising Zone Scale | Cond. Power |
| Min. CP | 0.3 |
| Max. CP | 0.8 |
| CP Computation Based on | Estimated δ/σ |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 1000000 |

☰ Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|---------------|--------------------------|----------|------------------------------|---------|-------------------|-----------|---------------------|
| | Count | Row % | Count | Row % | Count | Column % | |
| ☹ Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| ☹ Unfavorable | 6335 | 0.736% | 854209 | 99.264% | 860544 | 86.054% | 442 |
| ☹ Promising | 8084 | 8.067% | 92122 | 91.933% | 100206 | 10.021% | 752.664 |
| ☹ Favorable | 10509 | 26.775% | 28740 | 73.225% | 39249 | 3.925% | 442 |
| ☹ Efficacy | 1 | 100.000% | 0 | 0.000% | 1 | 1.000E-4% | 208 |
| All Trials | 24929 | 2.493% | 975071 | 97.507% | 1000000 | 100.000% | 473.13 |

Promising Zone defined as $0.3 \leq CP < 0.8$

☰ Average Sample Size

| Look # | Average Sample Size |
|---------|---------------------|
| 1 | 208 |
| 2 | 473.13 |
| Average | 473.13 |

☰ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|-----------|
| | | Efficacy | | Efficacy | Count |
| 1 | 208 | 5.251 | 1 | 1 | 1.000E-4% |
| 2 | 442 | 1.96 | 24928 | 999999 | 100.000% |
| Total | | | 24929 | 1000000 | |
| % | | | 2.493% | | |

The null hypothesis was rejected 24929 times in 1,000,000 trials, comfortably just the range of Monte Carlo accuracy for a level 0.025 test. Simulation results for other values of δ are displayed in Table 55.2.

The conventional Wald method also utilizes the CDL simulation worksheet, but it disables the CDL criterion by setting the cell titled

55 The Chen, DeMets and Lan Method

Use Wald Stat. if CP \geq to zero as shown below.

| | | |
|--|--------------------------------------|--|
| Use Adaptation Method | | |
| <input type="radio"/> CHW | <input checked="" type="radio"/> CDL | <input type="radio"/> Müller and Schäfer |
| Adapt at: | Look # | 1 |
| Max. Sample Size if Adapt (multiplier; total #): | 2 | 884 |
| Target CP for Re-estimating Sample Size: | 0.8 | |
| Use Wald Stat. if CP \geq : | 0 | |
| Promising Zone Scale: | Cond. Power | <input type="checkbox"/> CP |
| Promising Zone: | Min. CP: | 0.3 |
| | Max. CP: | 0.8 |
| CP Computation Based on: | Estimated δ/σ | |

By setting this CDL parameter to zero we have ensured that the conventional Wald statistic will be used for the final analysis all the time. In principle this should inflate the type-1 error. However, because the sample size is only increased in the promising zone, it is possible that the type-1 error might not be inflated in the current setting. This

turns out to be the case as is shown below for 1,000,000 simulated trials under $\delta = 0$.

Simulation: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means (CDL Simulation)

| Simulation Parameters | |
|--------------------------------------|---------------------------|
| Simulation ID | CDLSim2 |
| Design Type | Superiority |
| Number of Looks | 2 |
| Test Type | 1-Sided |
| Sample Size (n) | 442 |
| Variance | Equal |
| Test Statistic | t |
| Avg. Power at Termination | 0.025 |
| Response Generation Parameters | |
| Generate Data Using | Individual Means |
| Mean Control (μ_c) | 0 |
| Mean Treatment (μ_t) | 0 |
| SD Control (σ_c) | 7.5 |
| SD Treatment (σ_t) | 7.5 |
| Sample Size Re-estimation Parameters | |
| Method of Adaptation | Chen-DeMets-Lan |
| Adapt At Look No. | 1 |
| Max. Sample Size if Adapt | Multiplier |
| | 2 |
| Total # | 884 |
| Target CP | 0.8 |
| Use Wald Stat. if CP >= | 0 |
| Promising Zone Scale | Cond. Power |
| | Min. CP |
| | Max. CP |
| | 0.8 |
| CP Computation Based on | Estimated δ/σ |
| Simulation Control Parameters | |
| Starting Seed | Clock |
| Number of Simulations | 1000000 |

⊖ Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|-------------|--------------------------|----------|------------------------------|---------|-------------------|-----------|---------------------|
| | Count | Row % | Count | Row % | Count | Column % | |
| ⊖ Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| Unfavorable | 6394 | 0.744% | 853308 | 99.256% | 859702 | 85.970% | 442 |
| Promising | 8325 | 8.290% | 92101 | 91.710% | 100426 | 10.043% | 756.318 |
| ⊖ Favorable | 10686 | 26.801% | 29185 | 73.199% | 39871 | 3.987% | 442 |
| Efficacy | 1 | 100.000% | 0 | 0.000% | 1 | 1.000E-4% | 208 |
| All Trials | 25406 | 2.541% | 974594 | 97.459% | 1000000 | 100.000% | 473.565 |

Promising Zone defined as 0.3 <= CP < 0.8

⊖ Average Sample Size

| Look # | Average Sample Size (n) |
|---------|-------------------------|
| 1 | 208 |
| 2 | 473.566 |
| Average | 473.565 |

⊖ Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|--------|-----------------|------------|--------------|-------------------|-----------|
| | | Efficacy | Efficacy | Count | % |
| 1 | 208 | 5.251 | 1 | 1 | 1.000E-4% |
| 2 | 442 | 1.96 | 25405 | 999999 | 100.000% |
| Total | | | 25406 | 1000000 | |
| % | | | 2.541% | | |

The null hypothesis was rejected 25406 times in 1,000,000 trials allowing for Monte Carlo accuracy for a level 0.025 test. Simulation results for other values of δ are displayed in Table 55.2.

Having established that the CHW, CDL and conventional Wald tests have all preserved the type-1 error, it is now possible to have a meaningful comparison of their respective operating characteristics for other values of δ . These results are displayed in Table 55.2 for $\delta = 0, 1, 1.6, 1.8$ and 2. As noted above, the results for $\delta = 0$ were based on 1,000,000 simulated trials so as to leave no doubt that the type-1 error is preserved. The other results in Table 55.2 are all based on 100,000 simulated trials, which easily produces Monte Carlo accuracy to the nearest percentage point. The operating characteristics of the 442-subject fixed sample non-adaptive trial are also displayed so as to provide a benchmark for the comparisons.

Table 55.2 shows that all three adaptive methods preserve the type-1 error and are practically indistinguishable with respect to power or expected sample size for non-zero values of δ . This interesting finding suggests that for practical applications of adaptive sample size re-estimation in two-stage designs there is no loss of efficiency

55 The Chen, DeMets and Lan Method

Table 55.2: Operating Characteristics of Fixed Sample and Adaptive (CHW, CDL and Conventional Wald) Adaptive Designs

| Value of δ | Fixed Sample | | Adaptive-CHW | | Adaptive-CDL | | Adaptive-Wald | |
|-------------------|--------------|-----|--------------|------|--------------|------|---------------|------|
| | Power | N | Power | E(N) | Power | E(N) | Power | E(N) |
| 2.0 | 80.0% | 442 | 84.2% | 500 | 84.2% | 503 | 84.1% | 503 |
| 1.8 | 71.3% | 442 | 76.5% | 505 | 77% | 509 | 76.6% | 510 |
| 1.6 | 61.1% | 442 | 67.0% | 509 | 67% | 514 | 67.0% | 514 |
| 1.0 | 28.8% | 442 | 33.0% | 507 | 33.1% | 510 | 33.2% | 511 |
| 0.0 | 2.5% | 442 | 2.5% | 472 | 2.45% | 473 | 2.5% | 474 |

due to the use of the CHW statistic, notwithstanding the theoretical results of Tsiatis and Mehta (2003) or Jennison and Turnbull (2006). Further investigation of this conjecture would be desirable.

56

Muller and Schafer Method

This chapter discusses the Müller and Schäfer (2001) method for adaptive design. This is the most general of the three methods provided by EastAdapt and permits many different types of data dependent changes to a study design in addition to sample size changes. These include data-dependent changes in the error spending function, changes in the number and spacing of the interim looks, and population enrichment via the selection of prospectively identified subgroups. The actual decision rule for making an adaptive change at an interim look can be selected after examining the data available at that look. Indeed the adaptation may be made on the basis of either internal data from the trial, externally available data at the time of the interim look, or a combination of the two. Furthermore, these adaptive changes can be made more than once in any group sequential design. The method is based on preserving the conditional type-1 error in effect at the time of the adaptive change. One can show that if the type-1 error is preserved conditionally for all possible interim results, then it is also preserved unconditionally. P-values, point estimates and confidence intervals adjusted for the adaptive change are produced by extending the work of Müller and Schäfer (2001). We have developed two methods for this extension. Method 1 generalizes the repeated confidence intervals of Jennison and Turnbull (2000, Chapter 9) and was developed by Mehta, Bauer, Posch and Brannath (2007). We refer to it as the RCI method. It is more general than the RCI method discussed in Chapter 54 in that it is valid with any type of adaptive design change whereas the latter is only valid for sample size changes. When only sample size changes are involved, the two RCI methods are the same. Both RCI methods produce confidence intervals with conservative coverage of the unknown δ . Method 2 is BWCI (Backward Image Confidence Interval) method, developed by Gao, Liu and Mehta (2013), provided for computing a two-sided confidence interval having exact coverage, along with a point estimate that is median unbiased for the primary efficacy parameter in a two-arm adaptive group sequential design. The possible adaptations are not only confined to sample size alterations but also include data-dependent changes in the number and spacing of interim looks and changes in the error spending function. The procedure is based on mapping the final test statistic obtained in the modified trial into a corresponding backward image in the original trial. This is an advance on previously available methods, which either produced conservative coverage and no point estimates or provided exact coverage for one-sided intervals only.

In Section 56.1 we provide a quick review of the theory underlying the Müller and Schäfer method for preserving the type-1 error and its extension for parameter estimation by the RCI and BWCI methods. For more details, refer to Müller and Schäfer (2001), Mehta, Bauer, Posch and Brannath (2007), and Gao, Liu and Mehta (2013). In Section 56.2, we illustrate the methods through a worked example using the

56 Muller and Schafer Method

EastAdapt software.

56.1 Statistical Method

56.1.1 Hypothesis Testing

56.1.2 Parameter Estimation

The original method published by Müller and Schäfer (2001) only provided a solution for the problem of preserving the type-1 error in an adaptive hypothesis test. Subsequently the method was extended by Gao, Liu and Mehta (2013) to cover the related inference problem of computing the point estimate, confidence intervals and p-value. Accordingly in Section 56.1.1 we will discuss hypothesis testing based on the original Müller and Schäfer (2001) method. In Section 56.1.2 we will generalize the approach so as to cover parameter estimation and p-value computation based on the method of Gao, Liu and Mehta (2013).

56.1.1 Hypothesis Testing

To understand how the Müller and Schäfer (2001) method works let us consider a one-sided, level- α test of the null hypothesis

$$H_0: \delta = 0$$

versus the one-sided alternative hypothesis

$$H_1: \delta > 0$$

for a two-arm randomized clinical trial. We assume that this is a group sequential trial, designed for K looks at the information fractions t_1, t_2, \dots, t_K . Let $\alpha_j, j = 1, 2, \dots, K$, denote the amount of type-1 error to be spent at the j th look. Let the corresponding stopping boundaries be denoted by $\{b_j: j = 1, 2, \dots, K\}$.

Now suppose that at some interim look L the investigators, having already seen the results for the first L looks, wish to alter one or more design parameters for the future course of the study. Such data-dependent alterations might include a change in the maximum sample size, a change in the rate of error spending for the remainder of the trial, a change in the number and spacing of the future interim looks, and even a refinement of the eligibility criteria for enrolling additional patients into the trial. Müller and Schäfer have shown that all such changes are permissible provided the remainder of the trial preserves the *conditional rejection probability* (CRP), or conditional probability of rejecting H_0 , that are in effect at look L . This needs further explanation. Let Z_j be the Wald statistic at any look j and suppose that z_L is its observed value at look L . Then the CRP, denoted by ϵ_0 , is the conditional probability given z_L that, under the null hypothesis H_0 , Z_j will cross the stopping boundary at some future look. Specifically,

$$\begin{aligned} \epsilon_0 = & P_0\{Z_{L+1} \geq b_{L+1}|z_L\} + P_0\{Z_{L+1} < b_{L+1} \text{ and } Z_{L+2} \geq b_{L+2}|z_L\} + \dots \\ & \dots + P_0\left\{\bigcap_{j=L+1}^{K-1} Z_j < b_j \text{ and } Z_K \geq b_K|z_L\right\} \end{aligned} \quad (56.1)$$

(56.2)

This CRP is calculated by applying the recursive integration algorithm of Armitage, McPherson and Rowe (1969).

Müller and Schäfer (2001) have shown that, no matter what data dependent changes one makes at look L , the overall **unconditional** type-1 error of the entire trial with respect to all possible trial modifications at look L will be preserved provided the CRP for the modified trial beyond look L , under H_0 , remain fixed at ϵ_0 . Moreover, as the trial proceeds, the same process can be repeated again with further trial modifications that also preserve the CRP of the remainder of the trial.

For practical implementation in East one would conduct the adaptive trial as though it consisted of two trials; one primary and the other secondary. The initial design, prior to any adaptation is known as the primary trial. Suppose that at some look L of the primary trial the decision is taken to make an adaptive change in the design. At that point one would invoke East's conditional power calculator from the interim monitoring worksheet to obtain ϵ_0 (56.1.1). One would then use East to design a one-sided secondary trial with ϵ_0 as the significance level. This secondary trial would incorporate all the desired adaptive changes such a sample size change, spending function change, etc. The secondary trial would then be monitored as though it were a completely a separate trial with no relationship to the primary trial except for carrying over the significance level ϵ_0 . Acceptance or rejection of the null hypothesis in the secondary trial would imply acceptance or rejection of the null hypothesis overall. We shall illustrate this approach with the help of a detailed example in Section 56.2

56.1.2 Parameter Estimation

The material in this section summarizes the paper by Mehta, Bauer, Posch and Brannath (2007). It is fairly technical and may be skipped if you simply wish to design, monitor and simulate an adaptive trial by the Müller and Schäfer method. In that case you may proceed directly to Section 56.2. A careful study of this section will, however, provide you with a deeper appreciation of the difficulties of parameter estimation.

We will only consider parameter estimation for adaptive designs with one-sided hypothesis testing and no futility boundaries, since this is the only setting in which

56 Muller and Schafer Method

East currently provides point and interval estimates by the extended Müller and Schäfer method. (For the two-sided case one may use the repeated confidence intervals and the repeated p-values discussed in Chapter 54, Section 54.1.2). Accordingly we consider a level- α test of

$$H_0: \delta = 0 \tag{56.3}$$

versus the one-sided alternative hypothesis that $\delta > 0$. We shall be interested in estimating $\underline{\delta}$, the lower confidence bound of the $100 \times (1 - \alpha)\%$ confidence set

$$C_\alpha = (\underline{\delta}, \infty) .$$

We shall also be interested in estimating $\tilde{\delta}$, a point estimate for δ , and p_1 , a one sided p-value for the test of H_0 .

A general way to construct a $100 \times (1 - \alpha)\%$ confidence set C_α , applicable to both non-adaptive and adaptive group sequential trials is by performing a level- α test of the hypothesis

$$H_h: \delta = h \tag{56.4}$$

versus the one-sided alternative hypothesis that $\delta > h$. The confidence set C_α will then consist of all values h having the property that the hypothesis (56.4) cannot be rejected by a level- α one-sided hypothesis test. The lower limit of the confidence set C_α is therefore the supremum of the set of all h for which (56.4) is rejected by a level- α one-sided hypothesis test . It remains only to find a way to perform such a test in the adaptive setting.

Let us first review the Müller and Schäfer method for performing the one-sided test of the null hypothesis (56.3) that $\delta = 0$ at level α in the adaptive setting. For $j = 1, 2, \dots K$, let Z_j denote the Wald statistics and b_j denote the efficacy boundaries of a K -look one-sided level- α group sequential test. At some interim look L , where $Z_L = z_L$, it is decided to alter the future course of the trial through an adaptive change. In order to preserve the type-1 error of the trial despite the adaptive change, the following steps must be followed:

1. Compute the conditional rejection probability

$$\epsilon = P_0 \left\{ \bigcup_{j=L+1}^K (Z_j \geq b_j) | z_L \right\} . \tag{56.5}$$

2. Use the ϵ so obtained as the significance level of a $K^{(2)}$ -look secondary trial with Wald statistics $Z_j^{(2)}$ and efficacy boundaries $b_j^{(2)}$, $j = 1, 2, \dots K^{(2)}$, in

which all the adaptive changes have been incorporated. Thus

$$P_0\{\cup_{j=1}^{K^{(2)}} (Z_j^{(2)} \geq b_j^{(2)})\} = \epsilon ,$$

where all quantities associated with the secondary trial are tagged with the superscript ⁽²⁾.

3. Monitor the secondary trial until it is terminated at some stage $L^{(2)} \leq K^{(2)}$. Compute the stage wise adjusted p-value (see for example, Jennison and Turnbull 2000, page 179)

$$p^{(2)} = P_0\left(\bigcup_{j=1}^{L^{(2)}-1} \{Z_j^{(2)} \geq b_j^{(2)}\} \cup \{Z_{L^{(2)}}^{(2)} \geq z_{L^{(2)}}^{(2)}\}\right) . \quad (56.6)$$

4. Reject H_0 if $p^{(2)} \leq \epsilon$. By the Müller and Schäfer principle this is a level- α test of H_0 .

Now consider how the procedure might be extended to produce a level- α test of H_h . Analogous to (56.5) and (56.6) we must compute the conditional rejection probability $\epsilon(h)$ and the secondary trial p-value $p^{(2)}(h)$ under the hypothesis that $\delta = h$. The expression for $p^{(2)}(h)$ is a straightforward extension of (56.6) and is given by

$$p^{(2)}(h) = P_h\left(\bigcup_{j=1}^{L^{(2)}-1} \{Z_j^{(2)} \geq b_j^{(2)}\} \cup \{Z_{L^{(2)}}^{(2)} \geq z_{L^{(2)}}^{(2)}\}\right) \quad (56.7)$$

where $P_h(\cdot)$ denotes probability under H_h .

RCI Method We have shown in Mehta, Bauer, Posch and Brannath (2007) that

$$\epsilon(h) = P_h \bigcup_{j=L+1}^K (Z_j - h\sqrt{I_j} \geq b_j | z_L) \quad (56.8)$$

where I_j is the Fisher information at look j .

BWCI method Please refer to Gao, Liu and Mehta (2013) for details of BWCI method.

56 Muller and Schafer Method

56.2 Implementation of Hypothesis Testing

56.2.1 Designing the Primary

Primary

56.2.2 Monitoring the Primary

Primary

56.2.3 Primary Trial

56.2.4 Secondary Trial

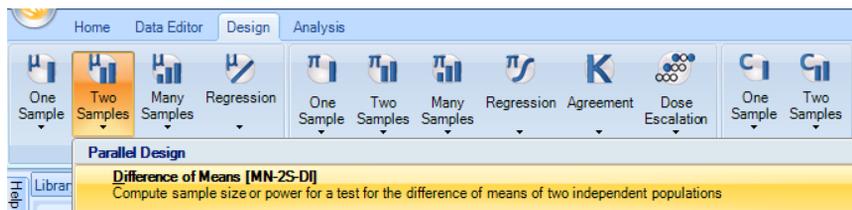
56.2.5 Combining Trial

56.2.6 Simulation

We illustrate the Müller and Schäfer method in this section through a worked example that includes the design of the trial, its adaptive re-design, and verification of its operating characteristics by simulation. Parameter estimation and p-value computation are presented separately in Section 56.3 since these capabilities are only available for one-sided tests.

56.2.1 Designing the Primary Trial

We begin with a one-sided, level 0.025, three look, group sequential design, with $LD(OF)$ spending function, for testing the difference of means, δ , in a two arm randomized clinical trial with a normally distributed primary endpoint. The study is designed to have 90% power to detect $\delta = 15$ at $\sigma = 50$. To design this study using EastAdapt, select as shown in the screen below.



Change the Number of Looks to 3. You will see a new tab **Boundary Info** added. Before we go to this tab, change **Input Method** to **Difference of Means, Diff. in Means** to 15 and **Std.Deviation** to 50. Keep other default selections without any change. Now, the input dialog box will look as shown below.

| | | | | | |
|---|---------------|--|---------------------|------------------------------|----|
| Test Type: | 1-Sided | Input Method: | Difference of Means | Test Statistic: | Z |
| Type I Error (α): | 0.025 | Diff. in Means ($\delta = \mu_1 - \mu_2$): | 15 | Std. Deviation (σ): | 50 |
| Power: | 0.9 | | | | |
| Sample Size (n): | Computed | | | | |
| Allocation Ratio: | 1 | | | | |
| | (n_1/n_2) | | | | |
| <input type="checkbox"/> Assurance (Probability of Success) | | | | | |

Click on the tab **Boundary Info**. Keep all the default selections in this tab without any

change. This tab inputs will look as shown below.

| | | | |
|---|----------------|---|-------------------|
| Efficacy Boundary Family: <input type="text" value="Spending Functions"/> Spending Function: <input type="text" value="Lan-DeMets"/> Parameter: <input type="text" value="OF"/> Type I Error (α): 0.025 | | Futility Boundary Family: <input type="text" value="None"/> | |
| Spacing of Looks: <input checked="" type="radio"/> Equal <input type="radio"/> Unequal | | Efficacy Boundary: <input type="text" value="Z Scale"/> | |
| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
| 1 | 0.333 | 0.000 | 3.710 |
| 2 | 0.667 | 0.006 | 2.511 |
| 3 | 1.000 | 0.025 | 1.993 |

Click **Compute** and the outputs for the design will be displayed in the **Output Preview** window in a newly added row.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | Input Method | δ | μ | Mean Treatment (Alt.) | σ | Test Statistic | |
|------|-------------|--------------|-----------|--------------------|-------|-------|------------------|-------------------|-------------|------------------|------------------|------------------|----------|-------|-----------------------|----------|----------------|---|
| Des1 | Superiority | 3 | 1-Sided | 0.025 | 0.9 | 1 | Equal | LD (OF) | 473 | 472.032 | 379.185 | Individual Means | 15 | 0 | | 15 | 50 | Z |

Now you can add the design output to the library workbook by clicking on the . This action saves the design Des1 as a node under the workbook Wbk1 in the library.



You can click on the icon in the library  to get the output summary as shown

56 Muller and Schafer Method

below.

| Wbk1:Des1 | |
|---|------------------|
| Mnemonic | MN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Input Method | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 15 |
| Mean Control (μ_c) | 0 |
| Mean Treatment (μ_t) | 15 |
| Std. Deviation (σ) | 50 |
| Test Statistic | Z |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Sample Size | |
| Maximum | 473 |
| Expected Under H0 | 472.032 |
| Expected Under H1 | 379.185 |

We see that the study will achieve the desired power at a maximum sample size of 473 subjects. However, the values of δ and σ on which these calculations rest were selected after considerable discussion and disagreement amongst the investigators. There was a scarcity of reliable data from previous studies about the treatment arm, the patient population and the primary endpoint. Thus the sample size of 473 was selected as a compromise, with the understanding that this important design parameter would be re-assessed at the first interim look, using data from the trial and possibly other external information that might become available at that time.

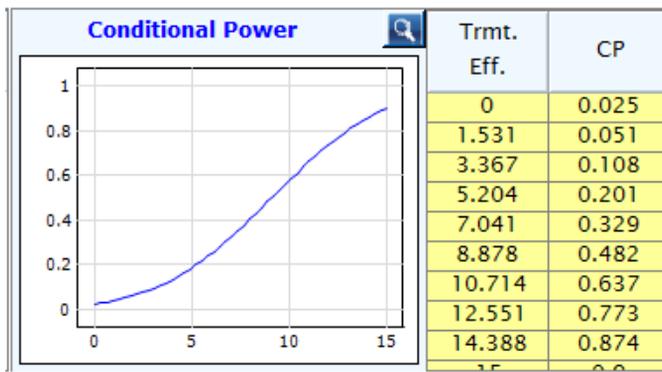
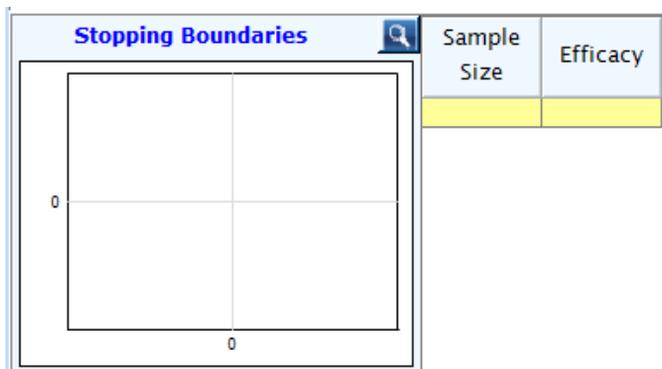
56.2.2 Monitoring the Primary Trial

To monitor this trial click on the 'Create Interim Monitoring' icon  to invoke the interim monitoring worksheet.

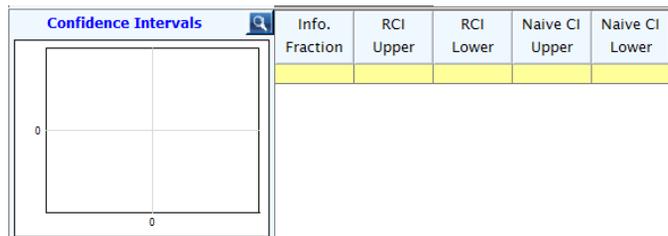
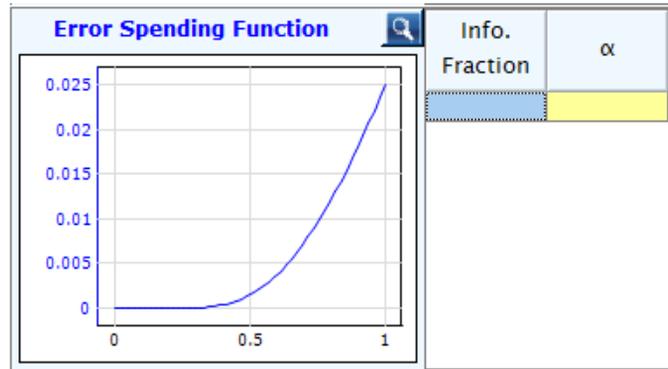
The parts of this sheet are shown, for visual clarity, in separate screen shots displayed

below.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 97.5% RCI for δ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|------------------------|-------|------------------|----|------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | | | | | | | | | | | |
| 2 | | | | | | | | | | | |
| 3 | | | | | | | | | | | |



56 Muller and Schafer Method



The top portion of the Interim Monitoring sheet is where the inputs for the interim looks will be entered and is displayed here again.

In the IM sheet you are ready to enter values for Look 1. Click on the button

Enter Interim Data

to see the Test Statistic Calculator dialog box displayed as shown

below.

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Cumulative Sample Size: 158

Input for Normal end point

Estimate of δ : 15
 $\delta = (\mu_t - \mu_c)$

Standard Error of Estimate of δ : 7.956

Output

Test Statistic: 1.885

Recalc OK Cancel

Some default values for Look 1 are already estimated and displayed in the calculator. You are free to change these values depending on your actual data. Suppose the first interim look is taken when data are available on $n_1 = 158$ subjects. Further, suppose that the observed difference of means is $\hat{\delta}_1 = 8$ and the observed standard deviation is $\hat{\sigma}_1 = 55$. Enter the value 8 for the estimate of δ and enter the square root of $(4 \times \hat{\sigma}_1^2 / 158) = 4 \times 55^2 / 158 = 8.751$ as the standard error of estimate of δ into the appropriate cells of this calculator. (Note that you can either type in a numerical value or a formula into the cells of any dialog box that accepts numerical values.)

56 Muller and Schafer Method

Click the **Recalc** button to see the output of test statistic value in the calculator.

Now click on **OK** to post these entries into the interim monitoring worksheet. The current information fraction is $t_1 = 158/473 = 0.334$. East populates the first row of the interim monitoring worksheet with the observed test statistic $z_1 = \hat{\delta}_1 / \text{se}(\hat{\delta}_1) = 0.914$, the corresponding efficacy stopping boundary = 3.706, the repeated 97.5% confidence interval limits for δ , and the repeated p-value.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 97.5% RCI for δ | | Repeat ... p-value | CP | Predicti... Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|------------------------|---------|--------------------|-------|-------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | 0.334 | 158 | 0.914 | 8 | 8.751 | 3.706 | Infinity | -24.432 | 0.439 | 0.311 | 0.388 |
| 2 | | | | | | | | | | | |
| 3 | | | | | | | | | | | |

56.2.3 Making Adaptive Changes to Primary Trial

The observed value of the Wald statistic at the first look, is $z_1 = 0.914$ whereas the critical value for rejecting H_0 is $b_1 = 3.706$. Thus a traditional group sequential trial would continue on to the next interim monitoring time point. Here, however, we have built in the flexibility to re-assess the adequacy of the sample size specified at the start of the trial. How should this be done? There are two aspects to this question; logistical

and scientific. We have already mentioned the logistical difficulties in Section 53.5 of Chapter 53 and will not discuss them further here. The scientific question is, how should one decide on the new sample size? The observed treatment difference $\hat{\delta}_1 = 8$ is considerably smaller than the value $\delta = 15$ at which the trial was powered.

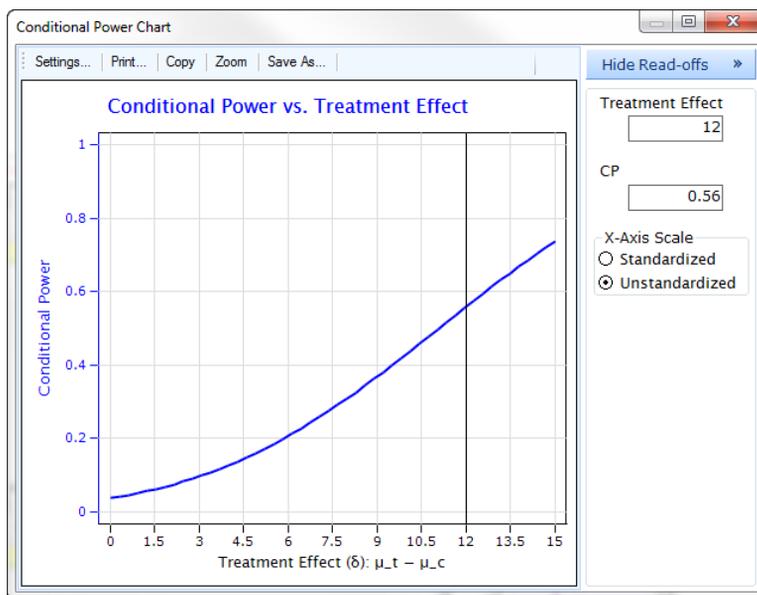
For an estimate of the conditional power that at any future look, that the test statistic value will cross the stopping boundary, click on the icon . You will see the following Conditional Power Calculator, displaying the conditional power as 0.311.

| Conditional Power Calculator | |
|--|-------|
| Input | |
| Look #: | 1 |
| Cumulative Sample Size: | 158 |
| Test Statistic: | 0.914 |
| Input/Output | |
| <input type="checkbox"/> Value of δ : | |
| Value of σ : | |
| <input checked="" type="checkbox"/> Value of δ/σ : | 0.145 |
| Computed Conditional Power: | 0.311 |
| Sample Size (Overall): | 473 |
| <input type="button" value="Recalc"/> <input type="button" value="Plot"/> <input type="button" value="Details..."/> <input type="button" value="Close"/> | |

You can also see the conditional power value for different assumed values of δ from

56 Muller and Schafer Method

the conditional power chart as shown below.

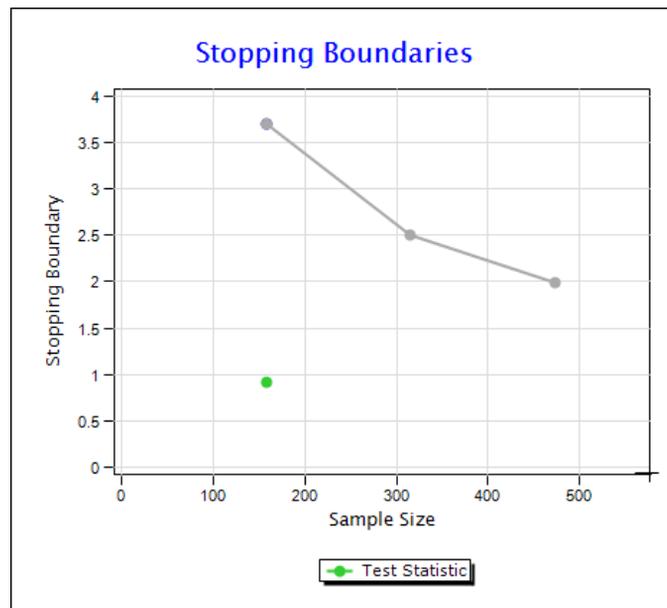


We shall shortly discuss the important role that conditional power plays in making an adaptive modification to the trial.

We would like to increase the sample size and thereby boost up the conditional power. The computation of conditional power, however, requires us to input a value for δ . Now, of course, the true value of δ cannot be known. While considerable weight should be given to the point estimate $\delta = 8$ obtained at the interim analysis, it might be wise to retain the flexibility to use this estimate in conjunction with other data from the trial, and other externally available data. The Müller and Schäfer method gives you this flexibility. You can revise the sample size in any manner that seems appropriate at the time of the interim analysis, without having to pre-specify a particular decision rule for determining the new sample size. (You may also decide that no sample size increase is warranted.) We will assume that the trial investigators have taken advantage of this flexibility to review the interim data, as well as all relevant external data, and have finally determined that the clinically meaningful value at which to power the study should be revised downwards to $\delta = 10$. Based on the observed data, the investigators continue to assume that $\sigma = 50$. Suppose therefore that they wish to increase the sample size so as to increase the conditional power to 90% at $\delta = 10$ and

$\sigma = 50$, while simultaneously preserving the type-1 error at 0.025 despite the data dependent change. The Müller and Schäfer method achieves this goal through a re-designed secondary trial as shown next.

The trial has so far only proceeded to the first interim look with 158 subjects enrolled, and the current value of the test statistic is $z_1 = 0.914$. Its current status can be depicted graphically in East by clicking on the ‘yellow up-arrow’ at the top of the thumbnail chart titled **Stopping Boundaries** in the interim monitoring worksheet, and checking off **Show Design** check box in the expanded chart that appears.



This chart displays the status quo. It shows us the current position of the test statistic in relation to the current and future stopping boundaries. Our objective is to re-design the continuation of this trial with appropriate changes to the sample size and stopping boundaries, and possibly also to the spending function, the number of remaining looks and their spacing. In effect, we wish to capitalize on having taken an unblinded look at the data from the 158 subjects already enrolled to re-design the trial so that it has a better chance of success and utilizes the data yet to be collected more efficiently. At the same time we do not wish to ignore the data already obtained when we perform the final analysis, and we do not want this trial to lose its pivotal status by failing to

56 Muller and Schafer Method

preserve the overall type-1 error. The Müller and Schäfer method makes this possible.

We stated in Section 56.1 that the unconditional type-1 error, over all possible design modifications, is preserved provided that each time a design modification is made, the remainder of the trial preserves the CRP, ϵ_0 . Here in our example with one-sided test, the first step is to compute ϵ_0 using equation (56.1.1). For example, in the above nominal critical point chart, $z_1 = 0.914$, the boundary at sample size $n_2 = 315$ is $(b_2) = 2.513$ and the boundary at sample size $n_3 = 473$ is $(b_3) = 1.993$. Therefore

$$\begin{aligned} \epsilon_0 &= P_0\{Z_2 \geq 2.513 | z_1 = 0.914\} \\ &\quad + P_0\{Z_2 < 2.513 \text{ and } Z_3 \geq 1.993 | z_1 = 0.914\} . \end{aligned}$$

The Müller and Schäfer calculator provided by EastAdapt can evaluate this CRP. With the cursor in any cell of the interim monitoring worksheet of Plan1, click on the icon



. The following Müller and Schäfer calculator dialog box appears,

revealing that at the first interim look the sample size is 158, the observed value of the test statistic is $z = 0.914$. The calculator permits the user to enter values for δ and σ , or for δ/σ , and computes conditional power assuming no change in the future course of the current group sequential design having a maximum sample size 473. The default values of δ and σ when this calculator is first invoked are the values that were entered into the interim monitoring worksheet through the test statistic calculator for the current look. In this case, the values were $\delta = 8$ and $\sigma = 55$, resulting in $\delta/\sigma = 0.145$. The conditional power if the trial proceeds without any design modification is shown to be 0.311. To obtain the conditional type-1 error (or conditional rejection

probability) we enter the value 0 in the δ/σ edit box and press the **Recalc** button.

| Conditional Power Calculator | |
|--|-------|
| Input | |
| Look #: | 1 |
| Cumulative Sample Size: | 158 |
| Test Statistic: | 0.914 |
| Input/Output | |
| <input type="checkbox"/> Value of δ : | |
| Value of σ : | |
| <input checked="" type="checkbox"/> Value of δ/σ : | 0 |
| Computed Conditional Power: | 0.038 |
| Sample Size (Overall): | 473 |
| <input type="button" value="Recalc"/> <input type="button" value="Plot"/> <input type="button" value="Details..."/> <input type="button" value="Close"/> | |

The conditional type-1 error is seen to be $\epsilon_0 = 0.038$.

We can make any desired modifications to the remainder of the trial, such as changing the remaining sample size, or the number of future interim look and their locations, provided we preserve the conditional rejection probability. Accordingly, it is decided that the trial should continue to be extended to two further looks but should utilize the **LD (PK)** (Pocock) spending function instead of the current **LD (OF)** (O'Brien-Fleming) spending function for the stopping boundary, so as to increase the chance of early termination. Additionally, the sample size should be increased appropriately to make the conditional power, given $z_1 = 0.914$ at $\delta = 10$ and $\sigma = 50$, equal to 90%. In keeping with the Müller and Schäfer principle, the new stopping boundaries should be such that if in fact $\delta = 0$, the probability of crossing the boundary is 0.038. That is, we must preserve the conditional type-1 error of the original unmodified trial in the modified trial.

56.2.4 Implementing Adaptive Changes through Secondary Trial

56 Muller and Schafer Method

At first sight, it appears complicated to modify the boundaries of the on-going trial in which $z_1 = 0.914$ so as to fulfill conditional type-1 error requirement, $\epsilon_0 = 0.038$, with a Pocock **LD (PK)** spending function for the boundary. The solution, however, is rather simple and can be accomplished very naturally within EastAdapt. The approach, proposed by Müller and Schäfer, is to step away from the actual trial (hereafter referred to as the primary trial) at its look $L = 1$, (where an adaptive change has been requested), and to instead design an independent secondary two-look trial that has 90% power to detect $\delta = 10$ at $\sigma = 50$, and utilizes the Pocock **LD (PK)** spending function to generate the boundary, with $\alpha = 0.038$. Note that this error probability is the only statistic that we are required to carry forward from the primary trial into the design of the secondary trial. The further progress of the primary trial may then be conveniently monitored by entering the observed values of the test statistic, **computed only from incremental data generated after trial modification** into the interim monitoring worksheet of this secondary trial. In particular, the value $z_1 = 0.914$ from the primary trial plays no role in the interim monitoring of the secondary trial, since this value was already factored into the computation of ϵ_0 . We illustrate below.

To design the secondary trial, click on Des1 node in the library, and then click on the icon . In the ensuing input dialog box, enter Number of Looks as 2, Test Type as 1-sided, Type I Error(α) as 0.038, power as 0.9, and Mean and SD values under alternative as 10 and 50 respectively as shown below.

| | | | | | |
|----------------------------|---------------|--|---------------------|------------------------------|----|
| Test Type: | 1-Sided | Input Method: | Difference of Means | Test Statistic: | Z |
| Type I Error (α): | 0.038 | Diff. in Means ($\delta = \mu_t - \mu_c$): | 10 | Std. Deviation (σ): | 50 |
| Power: | 0.9 | | | | |
| Sample Size (n): | Computed | | | | |
| Allocation Ratio: | 1 | | | | |
| | (n_t/n_c) | | | | |

Click on **Boundary Info** tab and change the efficacy boundary as PK, as shown below.

| | | | |
|--|-----------------------|---|--------------------------|
| Efficacy Boundary Family: <input type="text" value="Spending Functions"/> Spending Function: <input type="text" value="Lan-DeMets"/> Parameter: <input type="text" value="PK"/> Type I Error (α): 0.038 | | Futility Boundary Family: <input type="text" value="None"/> | |
| Spacing of Looks: <input checked="" type="radio"/> Equal <input type="radio"/> Unequal | | Efficacy Boundary: <input type="text" value="Z Scale"/> | |
| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
| 1 | 0.500 | 0.024 | 1.985 |
| 2 | 1.000 | 0.038 | 2.015 |

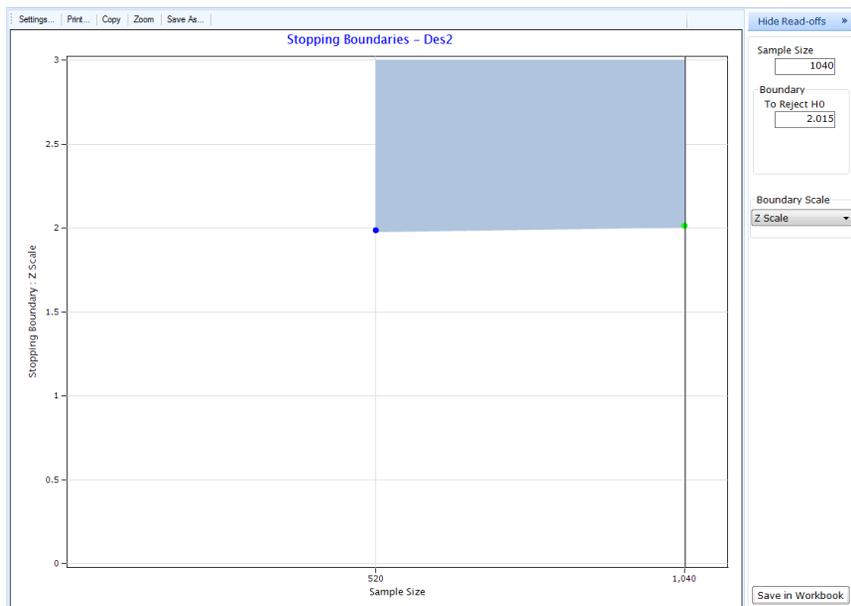
Now click on **Compute**. A new row will be added in the Output Preview window. Click on the row and save it in the library as Des2 node. Now select Des1 and Des2 nodes by holding ctrl key and then click Output Summary icon . You will see the following screen shot displaying the results for Des1 and Des2 side by side.

| | Wbk1:Des1 | Wbk1:Des2 |
|---|---------------------|---------------------|
| Mnemonic | MN-25-DI | MN-25-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 3 | 2 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.038 |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Input Method | Difference of Means | Difference of Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 15 | 10 |
| Std. Deviation (σ) | 50 | 50 |
| Test Statistic | Z | Z |
| Boundary Parameters | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (PK) |
| Sample Size | | |
| Maximum | 473 | 1040 |
| Expected Under H0 | 472.032 | 1027.747 |
| Expected Under H1 | 379.185 | 719.637 |

Des2 requires a maximum sample size of 1040 subjects and calls for two equally

56 Muller and Schafer Method

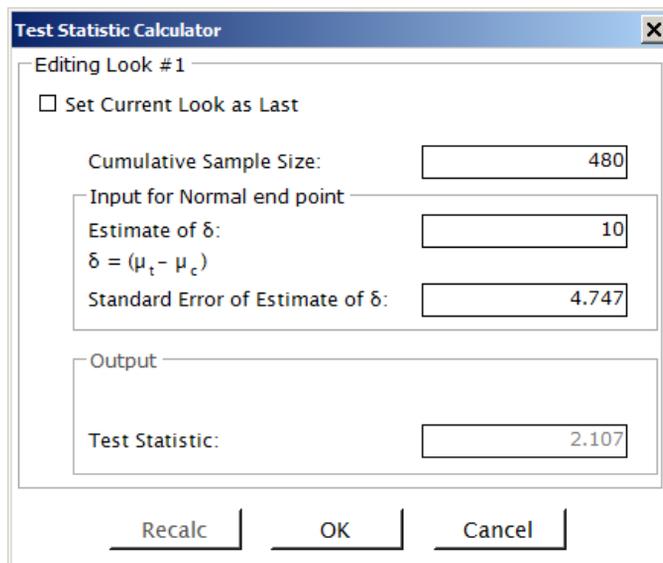
spaced looks, with a **LD (PK)** stopping boundary, as displayed below.



The further progress of the modified trial is now monitored on the interim monitoring worksheet of Plan2. Click on the **IM** tool to invoke the interim monitoring worksheet.

Suppose the data are monitored after 480 **new** subjects enter the trial. This corresponds to a total enrollment of $158 + 480 = 638$ subjects in the primary and secondary trials combined together. The secondary trial, however, only monitors the incremental data obtained from the 480 new subjects. Let us assume that these 480 new subjects provide the estimates $\hat{\delta} = 10$ and $\hat{\sigma} = 52$, so that $se(\hat{\delta}) = \sqrt{4 \times 52^2 / 480} = 4.747$. Enter these values into the Plan2 interim monitoring worksheet in the usual manner as shown

below.



The screenshot shows a dialog box titled "Test Statistic Calculator" with a close button (X) in the top right corner. The dialog is divided into several sections:

- Editing Look #1**: Contains a checkbox labeled "Set Current Look as Last" which is currently unchecked.
- Input for Normal end point**: A section containing three input fields:
 - "Cumulative Sample Size:" with the value 480.
 - "Estimate of δ :" with the value 10. Below this field is the formula $\delta = (\mu_t - \mu_c)$.
 - "Standard Error of Estimate of δ :" with the value 4.747.
- Output**: A section containing one output field:
 - "Test Statistic:" with the value 2.107.

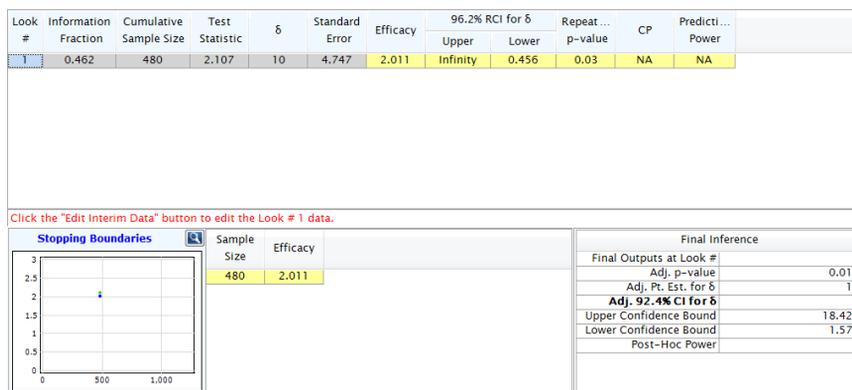
At the bottom of the dialog, there are three buttons: "Recalc", "OK", and "Cancel".

Click on **OK** to post these numbers into the interim monitoring worksheet. Now the observed value of the test statistic is 2.107 whereas the upper stopping boundary to reject H_0 is 2.011. Therefore you'll be notified by East that the stopping boundary has been crossed.

Click on the **Stop** button to terminate the trial and the Final Inference details are

56 Muller and Schafer Method

displayed as shown below.



Since the test statistic has crossed the upper boundary, the null hypothesis $\delta = 0$ is rejected. The requirement that the CRP be preserved no matter how the original trial is modified ensures that the unconditional type-1 error of the primary trial, taken over all possible trial modifications within any family of modifications under consideration, will always be preserved. We shall verify this fact through simulation in Section 56.2.6.

It should be noted that the confidence interval, point estimate and p-value displayed on the interim monitoring worksheet of the secondary trial are not valid for the overall trial. Those inferences must be made using the adaptive extension of the Müller and Schäfer (2001) procedure as described in Section 56.1.2. The implementation in EastAdapt is shown in Section 56.3.

56.2.5 Reconstructing a Combined Trial from the Primary and Secondary Trials

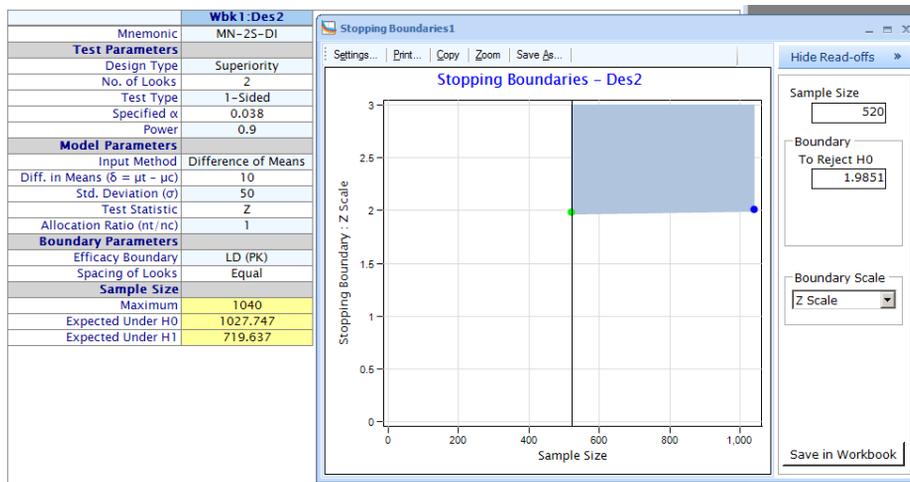
The secondary trial was terminated after a single look, taken at a sample size of 480 subjects. The interim monitoring worksheet of the secondary trial showed that the stopping boundary at this look was 2.011. Although not strictly necessary, it is instructive to transform these boundaries appropriately and attach them to the primary trial thereby recreating the combined trial in one piece. The path traced out by the test statistic in the secondary trial can likewise be appropriately transformed and attached to the test statistic generated in the primary trial before the trial was modified. The reconstruction is helpful for clarifying that it is the combined trial and not the secondary trial that is actually being monitored after an adaptive change in the design. The secondary trial is an artificial construct; a convenient way to obtain new stopping boundary satisfying the specification of the conditional rejection probability in the

combined trial.

Let us illustrate by reconstructing the combined trial for our example. In the discussion that follows, we shall distinguish between data from the primary and secondary trials by labeling the test statistics, stopping boundaries and sample sizes with superscripts. For example the sample size at look 1 in the primary trial is denoted by $n_1^{(1)}$ while the sample size of the secondary trial at look 1 is denoted by $n_1^{(2)}$. Now recall that the primary trial only proceeded up to the first interim look with a sample size $n_1^{(1)} = 158$ and corresponding stopping boundary 3.706. The mean and standard error of δ at look 1 were $\hat{\delta}_1^{(1)} = 8$ and $se(\hat{\delta}_1^{(1)}) = 8.751$, leading to the Wald statistic $z_1^{(1)} = 0.914$. At this point we implemented an adaptive change in the primary trial with the following requirements:

- Conditional rejection probability at $\delta = 0$ should be 0.038
- Conditional power at $\delta = 10, \sigma = 50$ should be 90%
- Two equally spaced additional looks with the **LD (PK)** spending function spending the type-1 error according to the CRP.

These requirements were incorporated into a secondary trial displayed below as Des2.



Although Des2 was designed for two equally spaced looks, at sample sizes 520 and 1040 respectively, the first look was actually taken at a sample size of $n_1^{(2)} = 480$. The information fraction at this look was $t_1^{(2)} = 480/1040 = 0.462$. By spending the appropriate amount of error at this information fraction the stopping boundaries was

56 Muller and Schafer Method

obtained as 2.011. We observed $\hat{\delta}_1^{(2)} = 10$, and $\text{se}(\hat{\delta}_1^{(2)}) = 4.747$, resulting in the Wald statistic $z_1^{(2)} = 2.107$.

We now show how to represent the stopping boundaries and test statistic values of the primary and secondary trials through a single combined trial. Suppose that a K -look primary trial is monitored up to and including look $L < K$, at which point an adaptive change takes effect. Suppose that all new data obtained after the adaptive change are monitored through a $K^{(2)}$ -look secondary trial which terminates at some look $K' \leq K^{(2)}$. It is possible to prove that monitoring the primary and secondary trials separately, as was done above, is equivalent to monitoring a single combined trial consisting of $L + K'$ looks. The stopping boundaries and test statistic values for the first L looks of this combined trial are identical to the corresponding values of the primary trial. The value of the test statistic at look $L + j$, $j = 1, 2, \dots, K'$, of the combined trial is

$$z_{L+j}^{(c)} = \frac{z_L^{(1)} \sqrt{n_L^{(1)}} + z_j^{(2)} \sqrt{n_j^{(2)}}}{\sqrt{n_L^{(1)} + n_j^{(2)}}}. \quad (56.9)$$

The value of the stopping boundary at look $L + j$ of the combined trial is

$$b_{L+j}^{(c)} = \frac{z_L^{(1)} \sqrt{n_L^{(1)}} + b_j^{(2)} \sqrt{n_j^{(2)}}}{\sqrt{n_L^{(1)} + n_j^{(2)}}}. \quad (56.10)$$

For more general settings, such as binomial or survival data, we would replace sample size by Fisher information in each of the above formulae.

Applying these formulae to the example under consideration we have $L = 1$ and $K' = 1$ so that the combined trial consists of $L + K' = 2$ looks. The boundaries and test statistics for the first look of the combined trial are identical to the corresponding values of the primary trial. The upper stopping boundary of the second look of the combined trial is obtained from equation (56.10) to be

$$b_2^{(c)} = \frac{0.914 \times \sqrt{158} + 2.011 \times \sqrt{480}}{\sqrt{158 + 480}} = 2.199$$

The value of the test statistic at the second look of the combined trial is obtained from equation (56.9) to be

$$z_2^{(c)} = \frac{0.9142 \times \sqrt{158} + 2.107 \times \sqrt{480}}{\sqrt{158 + 480}} = 2.282$$

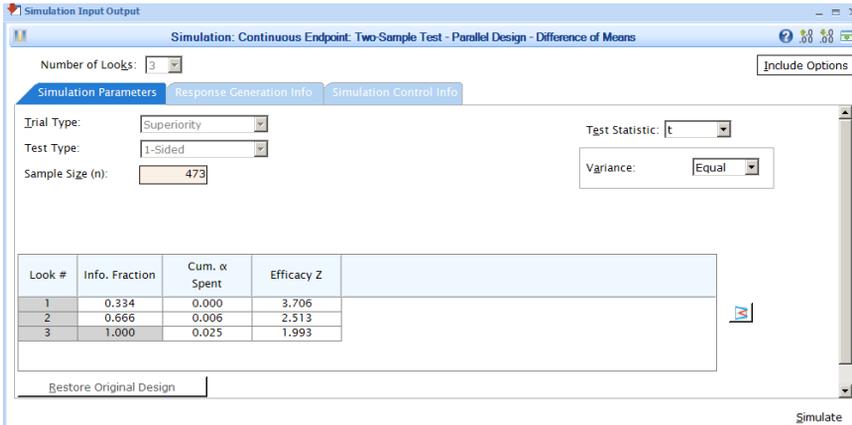
Since $z_2^{(c)} > b_2^{(c)}$, the combined trial is terminated.

56.2.6 Verifying Operating Characteristics by Simulation

Simulation is a very valuable tool for making adaptive decisions that suit the needs of the study. For example one might want to place an upper bound on the magnitude of the sample size increase following an adaptive look at the interim results, or one might want to place a lower bound on the estimated value of δ such that no sample size increase would be permitted should the estimate fall below the lower bound. These and other similar restrictions will affect the power of the study as well as the expected sample size in ways that might not be analytically tractable. One can, however, easily estimate power and expected sample size for various adaptive designs through simulation.

There is a second important reason for including a simulation tool in EastAdapt. We have made a major claim that by preserving the CRP after any type of adaptation, we will automatically preserve the unconditional type-1 error, taken over all possible adaptations, as well. A convincing way to demonstrate that this claim is correct is through simulation.

To illustrate how to use the simulation tool in EastAdapt, let us consider once again Des1 that we created in Section 56.2.1. With the cursor on Des1 node in the library, click on . You will see the following simulation input/output dialog box with three tabs **Simulation Parameters**, **Response Generation Info**, and **Simulation Control Info**. These are the same tabs you would have come across in the earlier chapters of the manual.



Simulation Input Output

Simulation: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means

Number of Looks: 3 Include Options

Simulation Parameters | Response Generation Info | Simulation Control Info

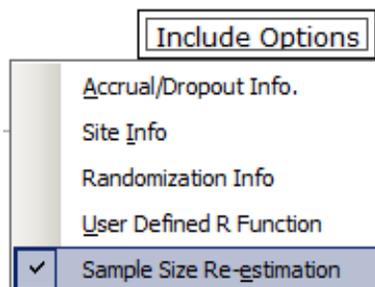
Trial Type: Superiority Test Statistic: t
 Test Type: 1-Sided Variance: Equal
 Sample Size (n): 473

| Look # | Info. Fraction | Cum. alpha Spent | Efficacy Z |
|--------|----------------|------------------|------------|
| 1 | 0.334 | 0.000 | 3.706 |
| 2 | 0.666 | 0.006 | 2.513 |
| 3 | 1.000 | 0.025 | 1.993 |

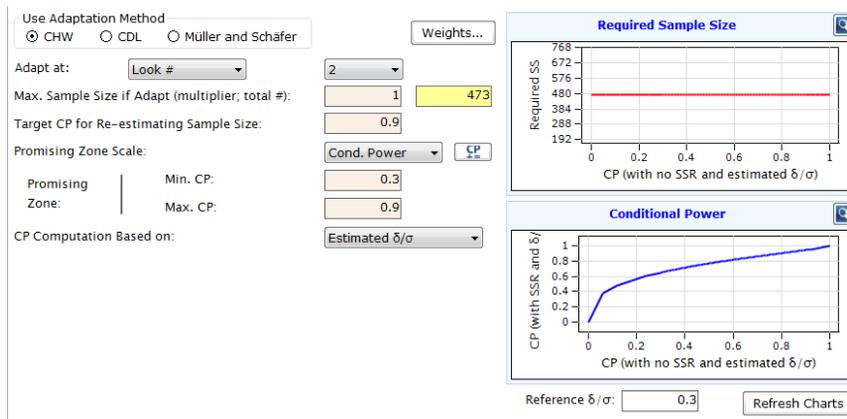
Restore Original Design Simulate

56 Muller and Schafer Method

Now click on the button **Include Options** and choose the item **Sample Size Re-estimation**.

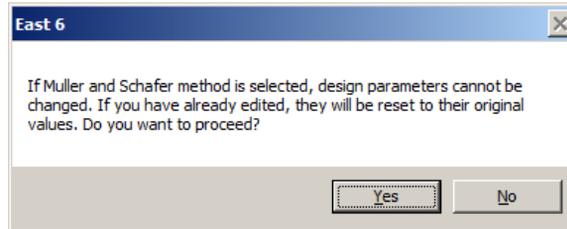


This will add a fourth tab, bearing the same name, to the dialog box.



Select the radio button against Müller and Schäfer . You will get the following dialog

box



Click on the button 'Yes'. In the resulting dialog box, Specify **Adapt at Look #** as 2 and **Max. Sample Size if Adapt (multiplier, total #)** as 2. You will see the max.sample size is computed and displayed as 946. Keep other specifications at the default values.

Use Adaptation Method
 CHW CDL Müller and Schäfer

Adapt at:

Max. Sample Size if Adapt (multiplier, total #):

Promising Zone Scale:
 Promising Zone:

CP Computation Based on:

Estimation Method:

Specify Stage II Design

Cond. Power

| | Stage I |
|---|---------------------|
| Mnemonic | MN-25-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Input Method | Difference of Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 15 |
| Mean Control (μ_c) | |
| Mean Treatment (μ_t) | |
| Std. Deviation (σ) | 50 |
| Test Statistic | Z |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Sample Size | |
| Maximum | 473 |

Now click on the button **Specify Stage II Design**. In the resulting dialog box, specify the Stage II Design details as described below.

The above dialog box has three sections.

Number of Looks Specify number of looks as 2.

Specification of α for Stage II This section of the dialog box asks you to specify how

56 Muller and Schafer Method

EastAdapt is to obtain the type-1 error for creating each simulated design of the Stage II or secondary trial . There are two choices. The default item is **Conditional Type-1 Error from Stage-1**. If you select this option, EastAdapt computes the conditional rejection probability ϵ_0 from each simulation of the primary trial at its look L , where an adaptive change has been requested. The secondary trial is then designed so as to spend $\alpha = \epsilon_0$. If you choose the **User Specified** item, then, you will have to specify how much α you would like to spend for the secondary trial. Ordinarily the default option should be selected as it ensures that the overall type-1 error of the adaptive trial will be preserved.

Specification of δ for Stage II This query of the dialog box asks you to specify how EastAdapt is to obtain the value of δ at which to power each simulation of the Stage II (secondary) trial. If you choose the **Estimated from Stage I** item, EastAdapt will use the value of δ estimated from the primary trial at its look L , where an adaptive change has been requested. If you choose the **User Specified** radio button, you will have to specify the value of δ at which to power the secondary trial. We stated in Section 56.2.3 that the secondary trial will be powered at $\delta = 10$. Therefore we select the **User Specified** radio button.

Specification of σ for Stage II This query of the dialog box asks you to specify how EastAdapt is to obtain the value of σ for each simulation of the Stage II (secondary) trial. If you choose the **Estimated from Stage I** radio button, EastAdapt will use the value of σ estimated from the primary trial at its look L , where an adaptive change has been requested. If you choose the **User Specified** radio button, you will have to specify the value of σ in a subsequent dialog box. We stated in Section 56.2.3 that the secondary trial will be powered at $\sigma = 50$. Therefore we select the **User Specified** radio button.

The dialog boxes will look as shown below.

Design Type: Superiority Number of Looks: 2 Specification of alpha: Cond. Type I Error From Stage-1
 Specification of delta: User Specified
 Specification of sigma: User Specified

Design Parameters **Boundary Info**

Test Type: 1-Sided Input Method: Individual Means Test Statistic: Z
 Type I Error (α): Computed
 Power: 0.9
 Sample Size (n): Computed
 Allocation Ratio: 1
 (n₁ / n₂)

Specify Mean Responses
 Mean Control (μ_c): 0
 Mean Treatment (μ_t): 10
 Std. Deviation (σ): 50

Efficacy Boundary Family: Spending Functions
 Spending Function: Lan-DeMets
 Parameter: PK
 Type I Error (α): Computed
 Spacing of Looks: Equal Unequal

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|--------------|-------------------|
| 1 | 0.500 | Computed | Computed |
| 2 | 1.000 | Computed | Computed |

Futility Boundary Family: None

Click OK and you will get the following dialog box where the summary details of Stage I and Stage II designs are displayed side by side.

Simulation Parameters **Response Generation Info** **Sample Size Re-estimation** **Simulation Control Info**

Use Adaptation Method
 CHW CDL Müller and Schäfer

Adapt at: Look #
 Max. Sample Size if Adapt (multiplier; total #): 2 946
 Promising Zone Scale: Cond. Power 0.3
 Promising Zone: Min. CP: Max. CP: 0.9
 CP Computation Based on: Estimated δ/σ
 Estimation Method: None

Specify Stage II Design

| | Stage I | Stage II |
|---|---------------------|------------------|
| Mnemonic | MN-25-DI | MN-25-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 3 | 2 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | Computed |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Allocation Ratio (n _t /n _c) | 1 | 1 |
| Input Method | Difference of Means | Individual Means |
| Diff. in Means (δ = μ _t - μ _c) | 15 | 10 |
| Mean Control (μ _c) | | 0 |
| Mean Treatment (μ _t) | | 10 |
| Std. Deviation (σ) | 50 | 50 |
| Test Statistic | Z | Z |
| Boundary Parameters | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (PK) |
| Sample Size | | |
| Maximum | 473 | Computed |

Now we are ready for carrying out our simulation of the trial. Let us call this as 'Experiment 1'.

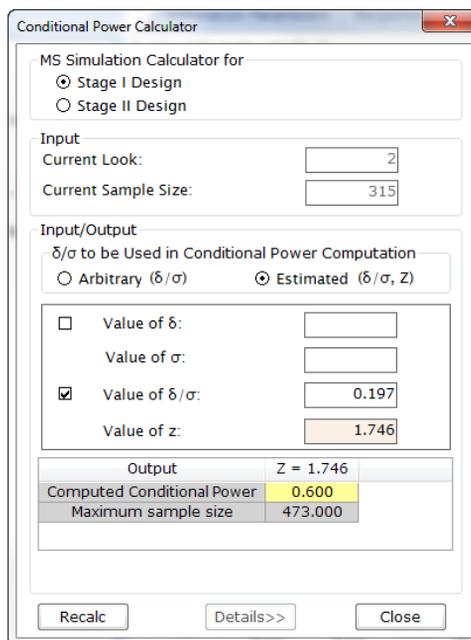
56 Muller and Schafer Method

Experiment 1: Explaining the Basics of the Simulation Tool

Suppose we specify simulation parameters as described below.

- Data for each simulation of the primary trial will be generated from a normal population with a difference of means $\delta = 12$ and population standard deviation $\sigma = 50$.
- In each simulation, the primary trial will proceed through $L = 2$ looks, each look being equally spaced with $473/3 = 158$ subjects. After look 2, there may be an adaptive change, depending on the simulated data obtained at look 2.
- At the end of the second look, when the sample size is 316, the conditional power will be computed. This computation will utilize the estimates $\hat{\delta}$ and $\hat{\sigma}$ obtained from the simulated data up to look L . The value of the conditional power estimate in relation to the re-design criteria **Min.CP** and **Max.CP** will determine whether or not the primary trial should undergo an adaptive change. If the conditional power obtained under the current design falls between 30% and 90%, then an adaptive change will be made to the primary trial. Alternatively, you can specify Promising Zone range in terms of Test Statistic or Estimated δ/σ , by making the choice in the drop down box. If adaptive change is decided, in that case:
 - As explained previously, the adaptive change to the primary trial will be implemented **indirectly** by invoking a secondary trial whose plan details are shown as **Specify Stage II Design** on this screen..
 - The secondary design is one-sided and spends its $\alpha = \epsilon_0$. The value assumed by this conditional rejection probability depends on the value of $z_L^{(1)}$ obtained in the primary trial.
 - There will be two equally spaced looks in the secondary trial with both α being spent according to the **LD (PK)** (Pocock) spending function.
 - The sample size of the secondary trial will be computed so that this trial can achieve $(1 - \beta) = 0.9$ power under the alternative hypothesis $\delta_1 = 10$ with $\sigma = 50$.
 - This indirect approach corresponds to modifying the primary trial in such a way that the **conditional power** for the remainder of the trial, given the observed value of $z_L^{(1)}$ is 90%.
- We have stated that EastAdapt will compute the sample size required in order for the secondary trial to achieve $(1 - \beta) = 0.9$ power at significance level of $\alpha = \epsilon_0$. Denote this sample size by $N_{\max}^{(2)}$, and denote the combined sample size, to be utilized by both the primary and secondary trials, by $N_{\max}^{(c)}$. In the present example, since $L = 2$, we must have $N_{\max}^{(c)} = 316 + N_{\max}^{(2)}$. More generally $N_{\max}^{(c)} = n_L^{(1)} + N_{\max}^{(2)}$

- In the present example, the field titled **Max. Sample Size if Adapt** has been set to 946 through a multiplier of 2 on primary trial max. sample size which is 473. This means that:
 - If $N_{\max}^{(c)} < 473$, EastAdapt will **extend** the combined sample size to $N_{\max}^c = 473$. In that case the sample size of the secondary trial will be correspondingly increased to $N_{\max}^{(2)} = 473 - 158 = 316$.
 - If $N_{\max}^c > 946$, East Adapt will **truncate** the combined sample size to $N_{\max}^{(c)} = 946$. In that case the sample size of the secondary trial will be correspondingly truncated to $N_{\max}^{(2)} = 946 - 316 = 630$.
- There is also a Conditional Power calculator available in this dialog box, which you can access by clicking on the button . This calculator will be useful to understand the simulation parameters and their impact on the simulation results. The calculator has two functions one for Stage I Design and the other for Stage II Design. By default, Stage I Design will appear selected as shown below.



You may enter any input values involving δ , σ , and z and can get the computed conditional power for the Stage I Design. In this example, the default values for δ/σ and z of 0.197 and 1.746 are displayed corresponding to conditional power of 0.6 which is a mid-value in the range specified for promising zone - 0.3 to

56 Muller and Schafer Method

0.9. You may change the input values to see their impact on computed conditional power in Stage I. Now select the radio button against Stage II Design as shown below.

Conditional Power Calculator

M5 Simulation Calculator for

Stage I Design

Stage II Design

Input

Current Look in Stage I:

Current Sample Size in Stage I:

Input/Output

δ/σ to be Used in Conditional Power Computation

Arbitrary (δ/σ) Estimated ($\delta/\sigma, Z$)

Value of δ :

Value of σ :

Value of δ/σ :

Value of z :

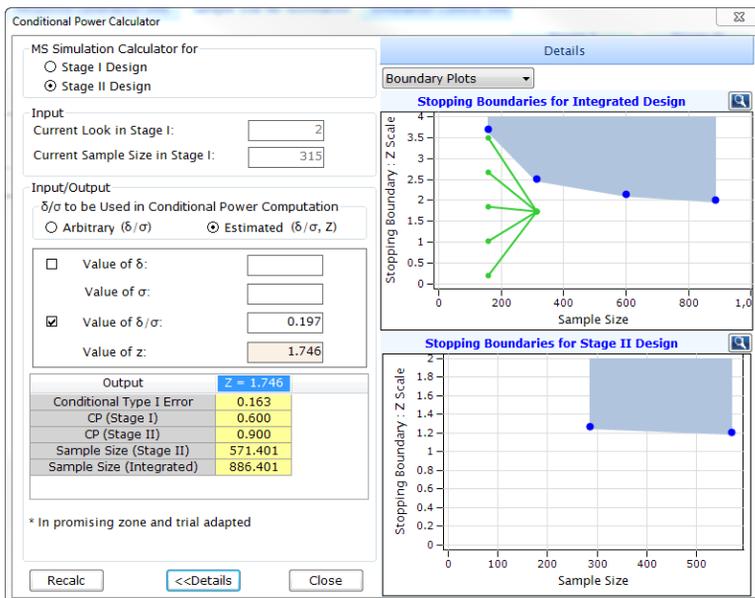
| Output | |
|--------------------------|---------|
| Conditional Type I Error | 0.163 |
| CP (Stage I) | 0.600 |
| CP (Stage II) | 0.900 |
| Sample Size (Stage II) | 571.401 |
| Sample Size (Integrated) | 886.401 |

* In promising zone and trial adapted

Recalc Details>> Close

Since at Stage I, the computed conditional power of 0.6 is in the Promising Zone, adaptation takes place. Further, the computations show that the maximum sample size for Stage II design is 571 and that for the integrated trial is 886 in order to achieve 90% power in Stage II. The implication is that we can choose a multiplier less than 2 in the specification for maximum sample size ($886/473 = 1.87$), provided the Stage I results assumption holds good. Now click on the button 'Details'. You will see the two Boundary Plots as shown

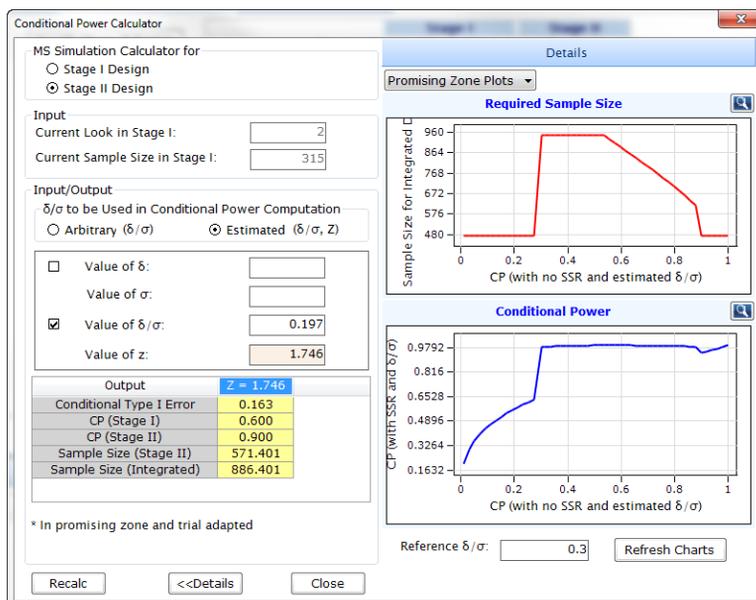
below.



In the Boundary Plot for the integrated design, the first and second looks boundaries correspond to those of Stage I design. The point plotted below the second look boundary value correspond to the z value estimated at that look. You may reach this point by different routes from the first look z value. For illustration, five different routes are shown, all joining the second look z value. Next, choose **Promising Zone Plots** in the drop down box under Details. You

56 Muller and Schafer Method

will see two plots -**Required Sample Size** and **Conditional Power** Plots.



The Required Sample Size plot shows that required sample size increases to a maximum of 946 at the start of promising zone, that is at CP=0.3 and gradually reduces when reaching the end of promising zone, that is at CP=0.9. Outside the promising zone, the required sample size remains steady at the stage I maximum sample size value of 473.

The Conditional Power plot shows the relationship between the conditional power without SSR and the conditional power with SSR, under a reference value of δ/σ . The conditional power with SSR increases to maximum values in the promising zone.

- There will be 10000 simulated trials and the screen will be refreshed after every 1000 simulations, and the starting seed for the simulations will be 100.

To run 10000 simulations of this adaptive design click on the **Simulate** button. After 10000 simulations are done click **Close**. East will add the results in a new row in Output Preview Window. Click on this row and add it to the library node under Des1. If you double-click on this node you will see the simulation results displayed in several small tables. You can collapse or expand each of these tables by clicking on down arrow or right arrow buttons at the top left hand side in each table.

First let us look at the table on the far left side of the screen.

| Stage I Design Parameters | |
|--------------------------------------|---------------------------|
| Simulation ID | MSSim1 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 473 |
| Variance | Equal |
| Test Statistic | t |
| Response Generation Parameters | |
| Generate Data Using | Individual Means |
| Mean Control (μ_c) | 0 |
| Mean Treatment (μ_t) | 12 |
| SD Control (σ_c) | 50 |
| SD Treatment (σ_t) | 50 |
| Sample Size Re-estimation Parameters | |
| Method of Adaptation | Müller and Schäfer |
| Adapt At Look No. | 2 |
| Max. Sample Size if Adapt | |
| Multiplier | 2 |
| Total # | 946 |
| Promising Zone Scale | Cond. Power |
| Min. CP | 0.3 |
| Max. CP | 0.9 |
| CP Computation Based on | Estimated δ/σ |

56 Muller and Schafer Method

| Stage II Design Parameters | |
|--------------------------------|------------------|
| Design Type | Superiority |
| Number of Looks | 2 |
| Test Type | 1-Sided |
| Type I Error (α) | Computed |
| Power | 0.9 |
| Test Statistic | Z |
| Input Method | Individual Means |
| Mean Control (μ_c) | 0 |
| Mean Treatment (μ_t) | 10 |
| Std. Deviation (σ) | 50 |
| Allocation Ratio (n_t/n_c) | 1 |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (PK) |

The above tables show parameters of stage-I design, parameters for sample-size re-estimation, and the parameters for stage-II design.

Now let us look at the Tables on the right side.

Zone-wise Averages The first tables on the right-side is displayed below.

⊖ Zone-wise Averages

| | Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|---|-------------|--------------------------|----------|------------------------------|---------|-------------------|----------|---------------------|
| | | Count | Row % | Count | Row % | Count | Column % | |
| ⊖ | Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| | Unfavorable | 563 | 24.900% | 1698 | 75.100% | 2261 | 22.610% | 473 |
| | Promising | 2944 | 94.541% | 170 | 5.459% | 3114 | 31.140% | 625.946 |
| ⊖ | Favorable | 987 | 90.968% | 98 | 9.032% | 1085 | 10.850% | 473 |
| | Efficacy | 3540 | 100.000% | 0 | 0.000% | 3540 | 35.400% | 308.436 |
| | All Trials | 8034 | 80.340% | 1966 | 19.660% | 10000 | 100.000% | 462.372 |

Promising Zone defined as $0.3 \leq CP < 0.9$

The above results show the classification of simulations by the criterion used under **Sample Size Re-estimation parameters** section in the simulation sheet. Out of 10000 simulations carried out, as there was no futility

rule in the design, no simulation was stopped for futility. In 3540 simulations, H_0 was rejected at the first or second look itself (Efficacy). In 2261 simulations, the CP was less than 0.30 (unfavorable zone) and in 1085 simulations, the CP was greater than 0.90 (favorable zone) and in both these cases, no adaption was needed. In the remaining 3114 simulations, where CP was between 0.3 and 0.9 (promising zone), adaption might be needed.

You may also notice what eventually happened to the simulations under each of the three Zones. Of the 2261 simulations that fell under unfavorable zone where no adaption was made to the sample size, in 563 simulations (24.9%), H_0 was rejected eventually. In all the 1085 simulations that were classified into favorable zone, in 987 simulations (91.0%), H_0 was rejected. Compared to these, in the 3114 simulations that were in the promising zone and where adaption in sample size was made, in as much as 2944 (94.5%) simulations, H_0 was rejected. This result illustrates the positive impact, the sample size adaption can bring about in a trial.

Simulation Results for Integrated Trial This table for the integrated trial, shows, look by look, information on the average sample size, the number of simulations in which the boundary for efficacy was crossed and the total number of simulations. The last row shows that the power attained in the integrated trial is 80.34%.

☰ **Simulation Results for Integrated Trial**

| Look # | Average Sample Size (n) | Stopping For | Total Simulations | |
|--------|-------------------------|--------------|-------------------|---------|
| | | Efficacy | Count | % |
| 1 | 158 | 148 | 148 | 1.480% |
| 2 | 315 | 3392 | 3392 | 33.920% |
| 3 | 517.079 | 3902 | 5698 | 56.980% |
| 4 | 818.373 | 592 | 762 | 7.620% |
| Total | 462.372 | 8034 | 10000 | |
| % | | 80.340% | | |

Zone-Wise Percentiles The table below shows the distribution of sample sizes in each

56 Muller and Schafer Method

zone in terms the percentiles.

⊖ Zone-Wise Percentiles

| | Zone | Percentile | Sample Size |
|---|-------------|------------|-------------|
| ⊖ | Futility | 5.000% | 0 |
| | | 25.000% | 0 |
| | | 50.000% | 0 |
| | | 75.000% | 0 |
| | | 95.000% | 0 |
| | | Average | 0 |
| ⊖ | Unfavorable | 5.000% | 473 |
| | | 25.000% | 473 |
| | | 50.000% | 473 |
| | | 75.000% | 473 |
| | | 95.000% | 473 |
| | | Average | 473 |
| ⊖ | Promising | 5.000% | 468 |
| | | 25.000% | 527 |
| | | 50.000% | 612 |
| | | 75.000% | 631 |
| | | 95.000% | 946 |
| | | Average | 625.946 |
| ⊖ | Favorable | 5.000% | 473 |
| | | 25.000% | 473 |
| | | 50.000% | 473 |
| | | 75.000% | 473 |
| | | 95.000% | 473 |
| | | Average | 473 |
| ⊖ | Efficacy | 5.000% | 315 |
| | | 25.000% | 315 |
| | | 50.000% | 315 |
| | | 75.000% | 315 |
| | | 95.000% | 315 |
| | | Average | 308.436 |
| ⊖ | All Trials | 5.000% | 315 |
| | | 25.000% | 315 |
| | | 50.000% | 473 |
| | | 75.000% | 510 |
| | | 95.000% | 751 |
| | | Average | 462.372 |

Simulation Results for Stage II Trial the table shown below gives the simulation results for Stage II alone, look by look.

⊖ Simulation Results for Stage II Trial

| Look # | Average Sample Size (n) | Stopping For | Total Simulations | |
|--------|-------------------------|--------------|-------------------|---------|
| | | Efficacy | Count | % |
| 1 | 249.443 | 2352 | 2352 | 75.530% |
| 2 | 503.373 | 592 | 762 | 24.470% |
| Total | 310.946 | 2944 | 3114 | |
| % | | 94.541% | | |

Simulation Boundaries for Stage I Design The last table shown below gives the

details of simulation boundaries for Stage I Design.

⊖ Simulation Boundaries for Stage I Design

| Look # | Sample Size (n) | Boundaries |
|--------|-----------------|------------|
| | | Efficacy |
| 1 | 158 | 3.706 |
| 2 | 315 | 2.513 |
| 3 | 473 | 1.993 |

Experiment 2: No Sample Size Increase

In Experiment 1, we permitted a sample size increase up to a maximum of 946 subjects. This sample size increase enabled the clinical trial to recover power even though the simulations were performed with $\delta = 12$ whereas the primary trial was actually a design to detect $\delta = 15$ with 90% power. If we were to impose the restriction that there should be no sample size increase in these simulations, we would expect to lose power. To see this, re-run the simulations with the **Simulation Parameters** as shown below.

Use Adaptation Method
 CHW CDL Müller and Schäfer

Adapt at: Look # 2

Max. Sample Size if Adapt (multiplier; total #): 1 473

Promising Zone Scale:

| | | |
|-----------------|---|--|
| Promising Zone: | Min. CP: 0.3 | Cond. Power ▼ <input type="button" value="CP"/> |
| | Max. CP: 0.9 | |

CP Computation Based on: Estimated δ/σ ▼

Estimation Method: None ▼

Notice that the sample size is not permitted to change from the initially specified value of 473 in these simulations. Click on the **Simulate** button and observe the results

56 Muller and Schafer Method

shown below.

Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Not Rejecting H0 | | Total Simulations | | Average Sample Size |
|-------------|--------------------------|----------|------------------------------|---------|-------------------|----------|---------------------|
| | Count | Row % | Count | Row % | Count | Column % | |
| Futility | 0 | 0.000% | 0 | 0.000% | 0 | 0.000% | 0 |
| Unfavorable | 588 | 25.995% | 1674 | 74.005% | 2262 | 22.620% | 473 |
| Promising | 2282 | 71.738% | 899 | 28.262% | 3181 | 31.810% | 435.524 |
| Favorable | 953 | 91.459% | 89 | 8.541% | 1042 | 10.420% | 473 |
| Efficacy | 3515 | 100.000% | 0 | 0.000% | 3515 | 35.150% | 307.853 |
| All Trials | 7338 | 73.380% | 2662 | 26.620% | 10000 | 100.000% | 403.03 |

Promising Zone defined as $0.3 \leq CP < 0.9$

This time only 7338 of the 10000 simulations of the combined trial rejected H_0 , yielding 73.38% unconditional power. Of the 10000 simulated trials, 3181 required a sample size increase and therefore activated the secondary trial. However, since no sample size increase was forthcoming, only 2282 of these trials were able to reject H_0 , resulting in 71.74% conditional power.

Experiment 3: Preserving the Unconditional Type-1 Error The statistical validity of the Müller and Schäfer adaptive procedure hinges on the claim that, despite making data dependent changes to the primary trial, the unconditional type-1 error is always preserved so long as the conditional rejection probability in effect at the time of the adaptive re-design is preserved. To verify this claim, edit Des1 by changing the type-1 error from 0.025 to 0.05, and by changing the spending function from **LD (OF)** (O’Brien-Fleming) to **LD (PK)** (Pocock), and save the edited design as Des3. These changes will exaggerate any possible inflation of type-1 error, and will thereby provide stronger empirical evidence for the validity of the Müller and Schäfer procedure. Des3 is displayed below as a group sequential design with three equally spaced looks and a

maximum sample size of 442 subjects.

| Wbk1:Des3 | |
|---|------------------|
| Mnemonic | MN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Input Method | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 15 |
| Mean Control (μ_c) | 0 |
| Mean Treatment (μ_t) | 15 |
| Std. Deviation (σ) | 50 |
| Test Statistic | Z |
| Allocation Ratio (nt/nc) | 1 |
| Boundary Parameters | |
| Efficacy Boundary | LD (PK) |
| Spacing of Looks | Equal |
| Sample Size | |
| Maximum | 442 |
| Expected Under H0 | 433.04 |
| Expected Under H1 | 269.098 |

Suppose we decided to convert Des3 into an adaptive design in the following manner:

1. Proceed with the primary group sequential trial up to and including look 2.
2. Let $z_2^{(1)}$ denote the observed value of the Wald statistic $Z_2^{(1)}$ at look 2 of the primary trial, and let $\hat{\delta}_2^{(1)}$ be the estimate of δ at look 2 of the primary trial.
3. Compute the sample size $N_{\max}^{(2)}$ for the secondary trial (i.e., for the remainder of the combined trial) so as to make the conditional power, given the observed value of $z_2^{(1)}$, equal to 90% under the assumption that $\delta = \hat{\delta}_2^{(1)}$.

Now let us simulate this adaptive design 10,000 times in two ways. First we will simulate the design **without** preserving the conditional rejection probability, ϵ_0 , obtained at the end of look 2 of the primary trial. We will, instead, run each simulation of the secondary trial at the 0.1 level.

56 Muller and Schafer Method

Fill in the ensuing dialog boxes as shown below.

The image displays three sequential dialog boxes from the EastAdapt software interface, configured for the Muller and Schafer adaptive design method.

Dialog 1: Response Generation Info
 - **Generate Data Using:** Individual Means (dropdown)
 - **Mean Control (μ_c):** 0 (input field)
 - **Mean Treatment (μ_t):** 0 (input field)
 - **Common Standard Deviation:**
 - **SD Control (σ_c):** 50 (input field)
 - **SD Treatment (σ_t):** 50 (input field)

Dialog 2: Adaptation Method
 - **Use Adaptation Method:** CHW, CDL, Müller and Schäfer
 - **Adapt at:** Look # (dropdown), 2 (dropdown)
 - **Max. Sample Size if Adapt (multiplier; total #):** 10 (input field), 4420 (input field)
 - **Promising Zone Scale:** Cond. Power (dropdown), CP (button)
 - **Promising Zone:** Min. CP: 0 (input field), Max. CP: 0.9 (input field)
 - **CP Computation Based on:** Estimated δ/σ (dropdown)
 - **Estimation Method:** None (dropdown)

Dialog 3: Specification of Stage II Design
 - **Design Type:** Superiority (dropdown)
 - **Number of Looks:** 1 (dropdown)
 - **Specification of alpha:** User Specified (dropdown)
 - **Specification of delta:** Estimated from Stage-I (dropdown)
 - **Specification of sigma:** Estimated from Stage-I (dropdown)

Dialog 4: Design Parameters
 - **Test Type:** 1-Sided (dropdown)
 - **Type I Error (α):** 0.05 (input field)
 - **Power:** 0.9 (input field)
 - **Sample Size (n):** Computed (input field)
 - **Allocation Ratio:** 1 (input field), (n_t/n_c)
 - **Input Method:** Individual Means (dropdown)
 - **Test Statistic:** Z (dropdown)
 - **Specify Mean Responses:** Mean Control (μ_c): Computed (input field), Mean Treatment (μ_t): Computed (input field)
 - **Std. Deviation (σ):** Computed (input field)

By selecting the **User Specified** radio button for the **Specification of alpha**, we have informed EastAdapt **not** to use the conditional rejection probability for the secondary trial. By selecting **Estimated from Stage-I** for δ and σ , we have asked EastAdapt to make data dependent sample size changes to the trial based on estimates of these parameters obtained from the primary trial. Click on **Simulate** button to generate 10,000 simulations of this adaptive design. Save the results in the library node. By double-clicking on the node or by clicking on 'Details' button, you

will the following results along with other results.

⊖ Simulation Results for Integrated Trial:

| Look # | Average Sample Size (n) | Stopping For | Total Simulations | |
|--------|-------------------------|--------------|-------------------|---------|
| | | Efficacy | Count | % |
| 1 | 147 | 229 | 229 | 2.290% |
| 2 | 295 | 152 | 152 | 1.520% |
| 3 | 1965.018 | 245 | 9619 | 96.190% |
| Total | 1898.002 | 626 | 10000 | |
| % | | 6.26 | | |

⊖ Simulation Results for Stage II Trial:

| Look # | Average Sample Size (n) | Stopping For | Total Simulations | |
|--------|-------------------------|--------------|-------------------|----------|
| | | Efficacy | Count | % |
| 1 | 3271.315 | 245 | 4689 | 100.000% |
| Total | 3271.315 | 245 | 4689 | |
| % | | 5.225 | | |

The **Simulation Results for Integrated Trial** panel shows that that 626 of the 10,000 simulations rejected the null hypothesis. Thus the type-1 error rate was 0.0626 which is excessive, even allowing for Monte Carlo error. We conclude that in these simulations the type-1 error was inflated.

Next we will simulate the adaptive design while also preserving the conditional rejection probability, ϵ_0 , obtained at the end of look 2 of the primary trial. Keeping cursor on the last simulation node, click on 'Edit' button and make selection for alpha as shown below.

56 Muller and Schafer Method

By selecting the **Conditional Type-I Error from Stage-I** radio button in the **Specification of alpha from Stage-I**, we are asking EastAdapt to use the conditional rejection probabilities obtained at the time of the adaptive change, for re-designing the trial. Therefore we would expect the unconditional type-1 error to be preserved. To see this, click on **OK** and then on **Simulate** to run 10,000 simulations. Now you will see the results as shown below.

⊖ Simulation Results for Integrated Trial:

| Look # | Average Sample Size (n) | Stopping For | Total Simulations | |
|--------|-------------------------|--------------|-------------------|---------|
| | | Efficacy | Count | % |
| 1 | 147 | 228 | 228 | 2.280% |
| 2 | 295 | 178 | 178 | 1.780% |
| 3 | 2031.773 | 118 | 9594 | 95.940% |
| Total | 1957.886 | 524 | 10000 | |
| % | | 5.24 | | |

⊖ Simulation Results for Stage II Trial:

| Look # | Average Sample Size (n) | Stopping For | Total Simulations | |
|--------|-------------------------|--------------|-------------------|----------|
| | | Efficacy | Count | % |
| 1 | 3391.477 | 118 | 4701 | 100.000% |
| Total | 3391.477 | 118 | 4701 | |
| % | | 2.51 | | |

The **Simulation Results for Integrated Trial** table shows that only 524 of the 10,000 simulations rejected H_0 , for an overall unconditional type-1 error rate of 0.0524. This demonstrates that the type-1 error of 0.05 was preserved up to Monte Carlo accuracy.

56.3 Implementation of Parameter Estimation

56.3.1 Parkinson's Disease

56.3.2 BWCI versus RCI

In this section we show how the generalization of the Müller and Schäfer (2001) method to the problem of parameter estimation has been implemented in EastAdapt. Results are presented for both the RCI method developed by Mehta, Bauer, Posch and Brannath (2007) and the BWCI (Backward Image Confidence Interval) method developed by Gao, Liu and Mehta (2013).

We shall see that the BWCI method has some advantages over the RCI method.

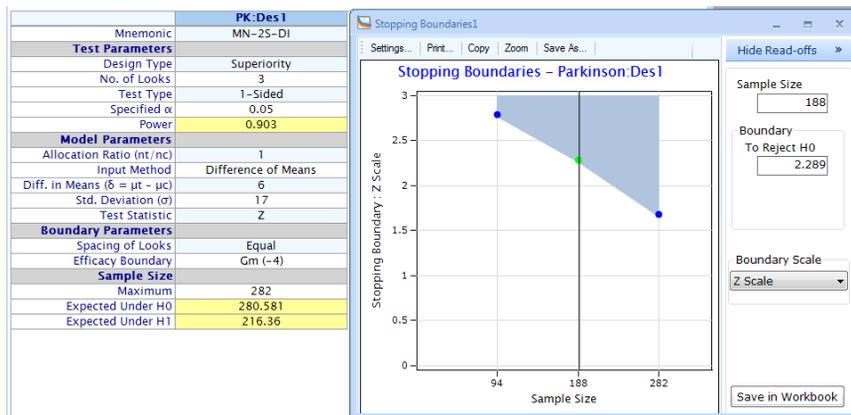
- The BWCI method produces confidence intervals with exact coverage, whereas the RCI method produces conservative coverage. The procedure is based on mapping the final test statistic obtained in the modified trial into a corresponding backward image in the original trial.
- The BWCI method produces a median unbiased point estimate, whereas the RCI method provides point estimates that can be severely negatively biased.

56.3.1 Parkinson's Disease Example

To illustrate how parameter estimation has been implemented, we consider a slight modification of an example discussed in Müller and Schäfer (2001). Müller and Schäfer consider a clinical trial comparing deep brain stimulation to conventional treatment for Parkinson's disease. The main outcome variable was the quality of life as measured by the 39-item Parkinson's Disease Questionnaire (the PDQ-39). Since no prior PDQ-39 data on deep brain stimulation were available, the study was planned based on the data from the pallidotomy trial of Martinez-Martin (2000). This led to the assumption of an improvement by $\delta = 6$ points in PDQ-39 for the treatment arm relative to the control arm. The standard deviation, also subject to considerable uncertainty, was assumed to be 17. We shall assume here that the trial was initially planned as a three-look group sequential design at the one-sided 0.05 level to test $H_0: \delta = 0$. A sample size of 282 subjects was selected with equally spaced interim monitoring after $n_1^{(1)} = 94$, $n_2^{(1)} = 188$, and $n_3^{(1)} = 282$ subjects, using the $\gamma(-4)$ error spending function of Hwang, Shih, and DeCani (1990). Upon entering these parameters into East we obtain the following design (Plan1) with slightly over 90% power to detect $\delta = 6$, and Wald stopping boundaries given by $b_1^{(1)} = 2.794$,

56 Muller and Schafer Method

$$b_2^{(1)} = 2.289, \text{ and } b_3^{(1)} = 1.680.$$



To illustrate our estimation procedure we implement a hypothetical (but realistic) scenario in which the first interim analysis is followed by an adaptive change to the design. Suppose that at the first interim analysis, when 94 subjects have been evaluated, the estimate of δ is $\hat{\delta}^{(1)} = 4.5$ with estimated standard deviation $\hat{\sigma} = 20$. We invoke the interim monitoring worksheet by pressing the **IM** icon.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 95% RCI for δ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|----------------------|-------|------------------|----|------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | | | | | | | | | | | |
| 2 | | | | | | | | | | | |
| 3 | | | | | | | | | | | |

Next click on 'Enter Interim Data' to bring up the test statistic calculator

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Cumulative Sample Size: 94

Input for Normal end point

Estimate of δ : 6
 $\delta = (\mu_1 - \mu_c)$

Standard Error of Estimate of δ : 3.507

Output

Test Statistic: 1.711

Recalc OK Cancel

Keep the cumulative sample size as 94. Enter $\hat{\delta} = 4.5$ and $se(\hat{\delta}) = \sqrt{4 \times 20^2 / 94}$ into

56 Muller and Schafer Method

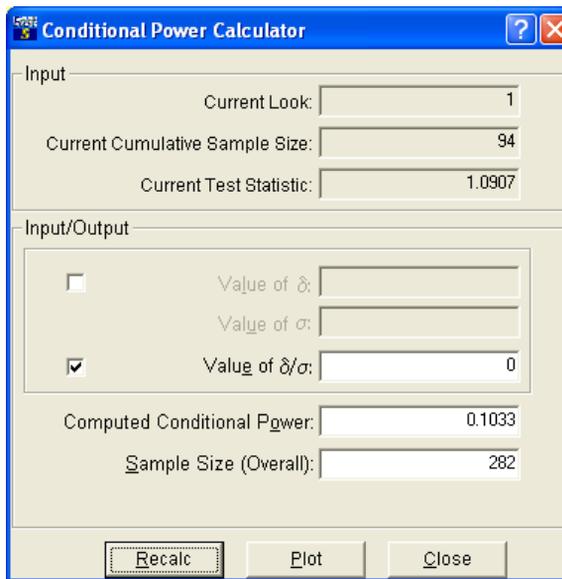
the test statistic calculator. Next, hit the **Recalc** button.

Now click **OK**. This completes the data entry for the first interim look.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 95% RCI for δ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|----------------------|--------|------------------|-------|------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | 0.333 | 94 | 1.091 | 4.5 | 4.126 | 2.794 | Infinity | -7.026 | 1 | 0.606 | 0.562 |
| 2 | | | | | | | | | | | |
| 3 | | | | | | | | | | | |

At this point it is decided to increase the sample size since, if in truth $\delta = 4.5$ and $\sigma = 20$, the conditional power is only about 60%, whereas we would prefer to proceed with at least 80% conditional power. The conditional rejection probability for the remainder of the trial is 0.1033. This can be seen by invoking the conditional power

calculator icon CP, setting $\delta/\sigma = 0$, and clicking on 'Recalc' button.



You can click on 'Close' and close the calculator. We may construct any suitable secondary trial to take over from the primary trial at the present look, as long as the significance level of the secondary trial is $\epsilon = 0.1033$.

How should the secondary trial be designed? The real benefit of an adaptive trial lies in the fact that all aspects of the original design can be re-visited at an interim look. All the observed efficacy and safety data, rather than just the summary statistics $\hat{\delta}$ and $\hat{\sigma}$, could be reviewed alongside any new external information that may also become available. Suitable design changes can then be made to the primary trial. In the present case we will assume that as a result of this type of review the investigators have determined that $\delta = 5$ rather than $\delta = 6$ would still constitute a clinically meaningful treatment benefit. Suppose then that the sponsor decides to re-design the study under the now more accurate assumption that $\delta = 5$ and $\sigma = 20$. To this end they decided to adopt a three-look secondary trial with $\gamma(-2)$ spending function and 80% power. The $\gamma(-2)$ spending function was selected because, under the new alternative hypothesis $\delta = 5$, it provides a reasonable chance of terminating for efficacy at the first or second interim looks. In keeping with the Müller and Schäfer principle the α for the secondary trial must be 0.1033. This secondary trial is constructed as shown below and displayed as Des2.

56 Muller and Schafer Method

Enter $\alpha = 0.1033$, 3 looks, 0.8 power, $\delta = 5$ and $\sigma = 20$ into the first dialog box of the design wizard.

Design Type: Superiority Number of Looks: 3

Design Parameters Boundary Info

Test Type: 1-Sided Input Method: Difference of Means Test Statistic: Z

Type I Error (α): 0.1033

Power: 0.8

Sample Size (n): Computed

Allocation Ratio: 1
(n_1/n_2)

Diff. in Means ($\delta = \mu_1 - \mu_2$): 5 Std. Deviation (σ): 20

Enter the $\gamma(-2)$ spending function into the second dialog box of the design wizard.

Design Parameters Boundary Info

Efficacy

Boundary Family: Spending Functions

Spending Function: Gamma Family

Parameter (γ): -2

Type I Error (α): 0.103

Spacing of Looks: Equal Unequal

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.333 | 0.015 | 2.162 |
| 2 | 0.667 | 0.045 | 1.781 |
| 3 | 1.000 | 0.103 | 1.351 |

Click on the **Compute** button to complete the design of the secondary trial. Save the design output into the library. Select Des1 and Des2 nodes and click on output

summary icon

| | PK:Des1 | PK:Des2 |
|---|---------------------|---------------------|
| Mnemonic | MN-2S-DI | MN-2S-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 3 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.103 |
| Power | 0.903 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Input Method | Difference of Means | Difference of Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 6 | 5 |
| Std. Deviation (σ) | 17 | 20 |
| Test Statistic | Z | Z |
| Boundary Parameters | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | Gm (-4) | Gm (-2) |
| Sample Size | | |
| Maximum | 282 | 296 |
| Expected Under H0 | 280.581 | 290.033 |
| Expected Under H1 | 216.36 | 228.311 |

We see that the secondary trial requires a total sample size of 296 subjects, over three equally spaced looks with $n_1^{(2)} = 99$, $n_2^{(2)} = 197$ and $n_3^{(2)} = 296$. (Note that this is over and above the 94 subjects already enrolled prior to the adaptive change.)

To monitor the secondary trial, while cursor is on Des2 node, click on the  icon.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 89.67% RCI for δ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|-------------------------|-------|------------------|----|------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | | | | | | | | | | | |
| 2 | | | | | | | | | | | |
| 3 | | | | | | | | | | | |

Suppose the following data are observed and the first and second interim looks, leading to termination of the trial at the second look.

| Look | SampSize | $\hat{\delta}$ | $\hat{\sigma}$ | $se(\hat{\delta})$ | $Z = \hat{\delta}/se(\hat{\delta})$ |
|------|----------|----------------|----------------|--------------------|-------------------------------------|
| 1 | 100 | 5.8 | 20.5 | 4.1 | 1.4146 |
| 2 | 200 | 6.1 | 19.5 | 2.7577 | 2.212 |

56 Muller and Schafer Method

After these values are entered into the interim monitoring worksheet, it looks as shown below

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 89.67% RCI for δ | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|-------------------------|--------|------------------|--------|------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | 0.338 | 100 | 1.415 | 5.8 | 4.1 | 2.154 | Infinity | -3.033 | 0.52 | 0.9174 | 0.7937 |
| 2 | 0.676 | 200 | 2.212 | 6.1 | 2.758 | 1.768 | Infinity | 1.223 | 0.038 | NA | NA |

Boundary Crossed

Since the value of Test Statistic is \geq the critical point for efficacy, H_0 is rejected.

Although boundary has been crossed, East gives you choice either to stop the study or to continue entering further looks. Please make your decision.

stop the study and bar further looks input

allow the study to continue

| Final Inference | |
|---|-------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.025 |
| Adj. Pt. Est. for δ | 5.824 |
| Adj. 79.34% CI for δ | |
| Upper Confidence Bound | 9.396 |
| Lower Confidence Bound | 2.147 |
| Post-Hoc Power | |

The stopping boundary at the second look was crossed and statistical significance has been achieved. The point and interval estimates of δ and the p-value displayed at the bottom right corner of the interim monitoring worksheet are, however, only valid for the secondary trial and not for the overall trial that combines the data from the first and second stages.

56.3.2 Evaluating the BWCI and RCI Methods by Simulation

In the previous section we designed and monitored a clinical trial comparing deep brain stimulation and conventional therapy for Parkinson's disease. EastAdapt provides a simulation tool for evaluating the properties of the two methods of estimation - RCI and BWCI. This tool can be invoked for any one-sided design. We shall demonstrate its utility by applying it to the Parkinson's disease example.

Return to the Des1 design that was created in the previous section for the Parkinson's

disease trial.

| PK:Des 1 | |
|---|---------------------|
| Mnemonic | MN-25-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.903 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Input Method | Difference of Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 6 |
| Std. Deviation (σ) | 17 |
| Test Statistic | Z |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | Gm (-4) |
| Sample Size | |
| Maximum | 282 |
| Expected Under H0 | 280.581 |
| Expected Under H1 | 216.36 |

With the cursor Des1 node, click on Simulate icon. In the resulting simulation input dialog box, click 'Include Options' and select 'Sample size re-estimation'. This will add an additional tab with the same name. In this tab select Müller and Schäfer option. Now the tab dialog box will look as shown below.

56 Muller and Schafer Method

Change the max.sample size multiplier as 2 and promising zone max.CP as 0.9. In the Estimation Method, select RCI. Accept the default value for Confidence Coefficient as 0.95. Now click 'Specify Stage-II Design'. In the resulting dialog box, select number of looks as 3 and specify power as 0.8. Accept other default choices. The dialog box will appear as shown below.

Now click on the tab **Boundary Info**, and specify spending function as Gamma with parameter -2 . Click 'OK'. Now the simulation dialog box will appear as shown below.

| | Stage I | Stage II |
|---|---------------------|------------------|
| Mnemonic | MN-2S-DI | MN-2S-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 3 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | Computed |
| Power | 0.9029 | 0.8 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Input Method | Difference of Means | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 6 | Computed |
| Mean Control (μ_c) | | Computed |
| Mean Treatment (μ_t) | | Computed |
| Std. Deviation (σ) | 17 | Computed |
| Test Statistic | Z | Z |
| Boundary Parameters | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | Gm (-4) | Gm (-2) |
| Sample Size | | |
| Maximum | 282 | Computed |

With the choices made, the design parameters of the secondary trial in each simulation will be estimated from the data generated from the primary trial at the time of the adaptation. The significance level (α) for this trial will be determined from the data of the primary trial, in keeping with the Müller and Schäfer principle. The sample size will be determined by the values of δ and σ that are estimated from the data of the primary trial. Click on 'Simulate' button. After 10,000 simulations are carried out, click 'Close' to get results in the Output Preview and add it a library node. With the

cursor on this node, click 'Details' icon. In the resulting output at the bottom you will see the results RCI Estimation method as shown below.

Estimation Results

Estimation Method: RCI
 Confidence Coefficient: 0.95
 95th Percentile of Lower Confidence Bound 3.901
 Coverage Probability 0.996

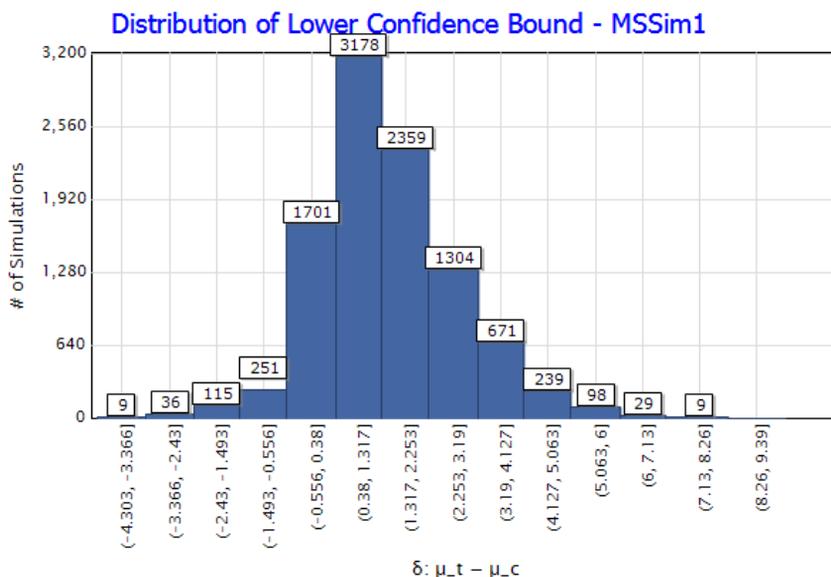
[-] Distribution of Lower Confidence Bounds

| Lower Confidence Bound | # of Sims | % Coverage | % Cum. Coverage |
|------------------------|-----------|------------|-----------------|
| (-4.303,-3.366] | 9 | 0.09 | 0.09 |
| (-3.366,-2.43] | 36 | 0.36 | 0.45 |
| (-2.43,-1.493] | 115 | 1.15 | 1.6 |
| (-1.493,-0.556] | 251 | 2.51 | 4.11 |
| (-0.556,0.38] | 1701 | 17.01 | 21.12 |
| (0.38,1.317] | 3178 | 31.78 | 52.9 |
| (1.317,2.253] | 2359 | 23.59 | 76.49 |
| (2.253,3.19] | 1304 | 13.04 | 89.53 |
| (3.19,4.127] | 671 | 6.71 | 96.24 |
| (4.127,5.063] | 239 | 2.39 | 98.63 |
| (5.063,6] | 98 | 0.98 | 99.61 |
| (6,7.13] | 29 | 0.29 | 99.9 |
| (7.13,8.26] | 9 | 0.09 | 99.99 |
| (8.26,9.39] | 1 | 0.01 | 100 |

You can choose from the plot icon menu, the item 'Distribution of Confidence Bounds'

56 Muller and Schafer Method

to get the following histogram.



With the cursor on MSSim1 node, click Edit button. In the resulting dialog box, choose BWCI as Estimation method and click 'Compute MUE' box.

Use Adaptation Method

CHW
 CDL
 Müller and Schäfer

Adapt at: Look # 2

Max. Sample Size if Adapt (multiplier; total #): 1 282

Promising Zone Scale: Cond. Power CP

Promising Zone:
 Min. CP: 0.3
 Max. CP: 0.9029

CP Computation Based on: Estimated δ/σ

Estimation Method: BWCI

Confidence Coefficient: 0.95

Compute MUE

Now click 'Simulate' button. After 10,000 simulations are done, carry out the required steps to save the simulation in a library node. With the cursor on this node, get the

56 Muller and Schafer Method

detailed output and see the BWCI Estimation results as shown below.

Estimation Results

| | |
|---|---------|
| Estimation Method: | BWCI |
| Confidence Coefficient: | 0.95 |
| 2.5th Percentile of Lower Confidence Bound | -1.853 |
| 97.5th Percentile of Upper Confidence Bound | 19.7059 |
| Median of Point Estimates: | 5.979 |
| Coverage Probability | 0.949 |

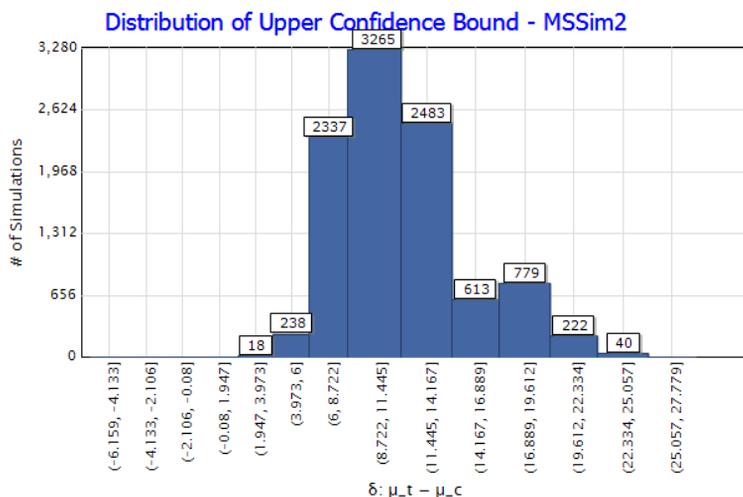
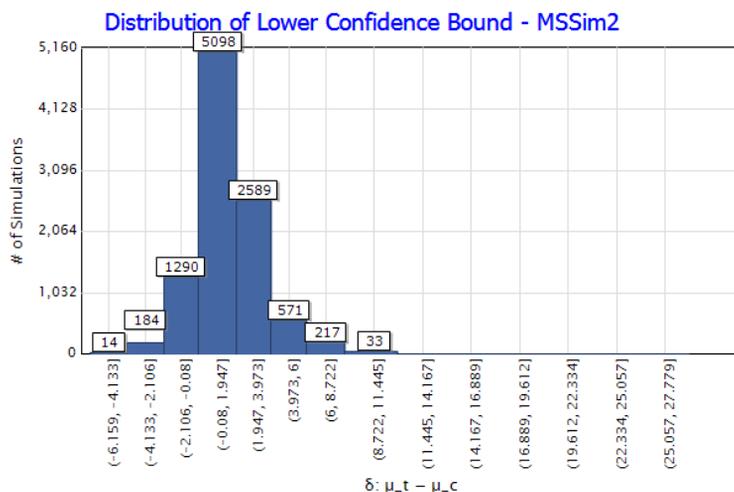
⊖ Distribution of Lower Confidence Bounds

| Lower Confidence Bound | # of Sims | % Coverage | % Cum. Coverage |
|------------------------|-----------|------------|-----------------|
| (-6.159,-4.133] | 14 | 0.14 | 0.14 |
| (-4.133,-2.106] | 184 | 1.84 | 1.98 |
| (-2.106,-0.08] | 1290 | 12.9 | 14.88 |
| (-0.08,1.947] | 5098 | 50.98 | 65.86 |
| (1.947,3.973] | 2589 | 25.89 | 91.75 |
| (3.973,6] | 571 | 5.71 | 97.46 |
| (6,8.722] | 217 | 2.17 | 99.63 |
| (8.722,11.445] | 33 | 0.33 | 99.96 |
| (11.445,14.167] | 4 | 0.04 | 100 |
| (14.167,16.889] | 0 | 0 | 100 |
| (16.889,19.612] | 0 | 0 | 100 |
| (19.612,22.334] | 0 | 0 | 100 |
| (22.334,25.057] | 0 | 0 | 100 |
| (25.057,27.779] | 0 | 0 | 100 |

⊖ Distribution of Upper Confidence Bounds

| Upper Confidence Bound | # of Sims | % Coverage | % Cum. Coverage |
|------------------------|-----------|------------|-----------------|
| [-6.159,-4.133) | 0 | 0 | 100 |
| [-4.133,-2.106) | 0 | 0 | 100 |
| [-2.106,-0.08) | 0 | 0 | 100 |
| [-0.08,1.947) | 1 | 0.01 | 100 |
| [1.947,3.973) | 18 | 0.18 | 99.99 |
| [3.973,6) | 238 | 2.38 | 99.81 |
| [6,8.722) | 2337 | 23.37 | 97.43 |
| [8.722,11.445) | 3265 | 32.65 | 74.06 |
| [11.445,14.167) | 2483 | 24.83 | 41.41 |
| [14.167,16.889) | 613 | 6.13 | 16.58 |
| [16.889,19.612) | 779 | 7.79 | 10.45 |
| [19.612,22.334) | 222 | 2.22 | 2.66 |
| [22.334,25.057) | 40 | 0.4 | 0.44 |
| [25.057,27.779) | 4 | 0.04 | 0.04 |

You can choose from the plot icon menu, the item 'Distribution of Confidence Bounds' to get the following histogram.



The results obtained so far help to evaluate the properties of the BWCI and RCI methods with respect to coverage. Similarly we can carry out simulations to evaluate bias, by specifying in the simulation parameters confidence coefficient as 0.5 for both

56 Muller and Schafer Method

the methods. Under BWCI method, the option 'Compute MUE' also can be chosen. These simulation results are summarized in Table 56.1.

Table 56.1: Comparison of BWCI and RCI methods for Parkinson's Disease example with 3-look $\gamma(-4)$ boundary for primary trial, adaptation at first look, and 3-look $\gamma(-2)$ boundary for secondary trial.

| True δ | Actual Coverage of 95% CI | | Median of $\hat{\delta}_{0.5}$ | |
|---------------|---------------------------|-------|--------------------------------|--------|
| | BWCI | RCI | BWCI | RCI |
| 6 | 0.949 | 0.995 | 5.939 | 1.929 |
| 3 | 0.95 | 0.985 | 3.028 | 0.438 |
| 0 | 0.948 | 0.95 | 0.021 | -3.336 |

The results in the above table shows that while the coverage property of the two methods are similar, the bias in estimation is markedly more in RCI method compared to BWCI method.

We know from the design surv-01 that a hazard ratio of 0.7 will yield 90% power. But what if the true hazard ratio was 0.77? The resultant deterioration in power can be evaluated by simulation. Accordingly we shall alter the Treatment cell, containing the hazard 0.0607, by replacing it with $0.77 * 0.0866 = 0.0667$.

Survival Information

of Hazard Pieces Input Method:

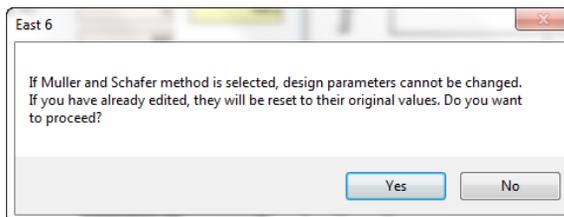
Hazard Ratio

| Piece | Starting at Time | Hazard Rates | | Hazard Ratio |
|-------|------------------|--------------|-----------|--------------|
| | | Control | Treatment | |
| 1 | 0.000 | 0.087 | 0.067 | 0.770 |

The "Sample Size Re-Estimation" Tab

The impact of an adaptive increase in the number of events and sample size on power and study duration can be evaluated by simulation. Click the **Sample Size Re-estimation** tab. This tab contains the input parameters for performing the adaptive simulations and sample size re-estimation in the on-going trial. Select **Muller and**

Schafer button in the dialog box. You will see the following message on the screen:



Click on **Yes**.

Now you see the dialog box shown below.

Use Adaptation Method
 CHW CDL Müller and Schäfer

Specify Stage II Design

Adapt at:

Max. # of Events if Adapt (multiplier; total #):

Max. Sample Size if Adapt (multiplier; total #):

Upper Limit on Study Duration:

Promising Zone Scale:
 Promising: Min. CP:
 Zone: Max. CP:

CP Computation Based on:

Accrual Rate After Adaptation:

Estimation Method:

| | Stage I |
|---|--------------------|
| Mnemonic | SU-25-LRSD |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.7 |
| Var (Log HR) | Null |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Futility Boundary | Gm (-5) (NB) |
| Accrual & Dropout Parameters | |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 483 |
| Events | |
| Maximum | 334 |
| Study Duration | |
| Maximum | 30 |
| Expected Under H0 | 22.92 |
| Expected Under H1 | 22.92 |

The **Sample Size Re-estimation** tab is the main location from which you will be using East to design adaptive time-to-event trials.

Input Parameters for Sample Size Re-estimation

This window consists of 10 input fields into which one may enter various design parameters.

56 Muller and Schafer Method

| | | | |
|---|----------|-------------------------|-----|
| Use Adaptation Method | | Specify Stage II Design | |
| <input type="radio"/> CHW <input type="radio"/> CDL <input checked="" type="radio"/> Müller and Schäfer | | | |
| Adapt at: | Look # | 1 | |
| Max. # of Events if Adapt (multiplier; total #): | | 1 | 334 |
| Max. Sample Size if Adapt (multiplier; total #): | | 1 | 483 |
| Upper Limit on Study Duration: | | 90 | |
| Promising Zone Scale: | | Cond. Power | |
| Promising Zone: | Min. CP: | 0.3 | |
| | Max. CP: | 0.9 | |
| CP Computation Based on: | | Estimated HR | |
| Accrual Rate After Adaptation: | | No Change | |
| Estimation Method: | | None | |

For a given set of design parameters, East will run a number of simulated trials as specified in the **Simulation Control Info** tab:

| Number of Simulations: 100000 Refresh Frequency: 1000 Random Number Seed <input checked="" type="radio"/> Clock <input type="radio"/> Fixed 100 <input type="checkbox"/> Suppress All Intermediate Output <input type="checkbox"/> Pause after Refresh <input checked="" type="checkbox"/> Stop At End | Output Options Output Type: Case Data <input type="checkbox"/> Save summary statistics for every simulation run <input type="checkbox"/> Save subject-level data for 1 simulation runs Note: Max. 100,000 records will be saved. | | | | | | |
|---|--|----------------|---|----|----|----|----|
| | Output for All Trials <table border="1"> <thead> <tr> <th>Percentile (%)</th> </tr> </thead> <tbody> <tr><td>5</td></tr> <tr><td>25</td></tr> <tr><td>50</td></tr> <tr><td>75</td></tr> <tr><td>95</td></tr> </tbody> </table> | Percentile (%) | 5 | 25 | 50 | 75 | 95 |
| Percentile (%) | | | | | | | |
| 5 | | | | | | | |
| 25 | | | | | | | |
| 50 | | | | | | | |
| 75 | | | | | | | |
| 95 | | | | | | | |

On running the simulations, an entry for Simulation output gets added in the **Output Preview** pane and the detailed output can be seen in the Output Summary of Simulations.

The input quantities in the **Sample Size Re-estimation** tab are described below in detail.

1. **Adaptation at:** For a K -look group sequential design, one can decide the time at which conditions for adaptations are to be checked and actual adaptation is to be carried out. This can be done either at some intermediate look or after accumulating data on specified number of events or after some specified information fraction. The value of this parameter depends upon the choice of the user. If it is **Look no.** then this parameter can be any integer number from 1 to $K - 1$. If the adaptation is to be carried out after observing specified events then this parameter can be some integer between [4, No. of events at design stage] and so on. The default choice in East is **look number** to decide the time of adaptation.

Adapt at:

| | |
|----------------|---|
| Look # | ▼ |
| Look # | |
| Events | |
| Info. Fraction | |

2. **Max Number of Events if Adapt :** This quantity is a multiplier with value ≥ 1 for specifying the upper limit (or cap) on the increase in the number of events, should an adaptive increase be called for based on the target conditional power. Notice that, in keeping with the FDA Guidance on Adaptive Clinical Trials (2010), East does not permit an adaptive decrease in the number of events. Therefore multipliers less than 1 are not accepted in this cell. For example, if you use the multiplier 1.5 and if adaptation takes place, the modified number of events is capped at 501. The 501-event cap becomes effective only if the increased number of events (as calculated by the criteria of cells 4, 5 and 6) exceed 501.

Max. # of Events if Adapt (multiplier; total #):

| | |
|-----|-----|
| 1.5 | 501 |
|-----|-----|

3. **Max Subjects if Adapt :** This quantity is a multiplier with value ≥ 1 for specifying the upper limit (or cap) on the number of subjects to be enrolled in the study. Although the power of the trial is determined by the number of events and not the number of subjects, the number of subjects play a role in determining how long it will take to observe the required number of events, and hence for determining the study duration. The number of subjects may only be increased, never decreased. Therefore multipliers less than 1 are not accepted in this cell. For example, if you use the multiplier 1.5 and if adaptation takes place,

56 Muller and Schafer Method

the modified number of subjects is capped at 724 subjects. The trial will continue to enroll subjects until either the required number of events is reached or the cap on the number of subjects is reached.

Max. Sample Size if Adapt (multiplier; total #):

- Upper Limit on Study Duration :** An event driven trial ordinarily continues until the required number of events arrive. This input parameter is provided merely as a safety factor in order to prevent the trial from being prolonged excessively should the required number of events be very large or their rate of arrival be very slow. Its default value is set at three times the expected study duration obtained from the initial design of the trial. Consequently, if the scenarios being simulated are realistic, the required number of events will almost always be attained much before this upper limit parameter becomes operational. It is recommended to leave this parameter unchanged at least for the initial set of simulation experiments since it would interfere with the operating characteristics of the study if it were to become operational.

Upper Limit on Study Duration:

- Target Conditional Power for Re-estimating Events :** This parameter ranges between 0 and 1 and is the *target conditional power* desired at the end of the study. Suppose, for example that the Target CP is set at 0.9.

Target CP for Re-estimating # of Events:

Let the value of the test statistic obtained in the current simulation be z_L at

look L , where an adaptive increase in the number of events is being considered. Then, by setting the left hand side of equation (54.21) to 0.9 we have:

$$0.9 = 1 - \Phi \left\{ b_K \sqrt{1 + \frac{D_L}{D_K - D_L}} - z_L \sqrt{\frac{D_L}{D_K - D_L}} - \delta \sqrt{r(1-r)} \sqrt{D_K^* - D_L} \right\}. \quad (56.11)$$

Upon solving equation (56.11) for D_K^* we obtain the increased number of events

that are needed to achieve the target conditional power of 0.9 in this simulation. Let us illustrate with Des 1. In Des 1 $K = 2, L = 1, r = 0.5$ and the critical value for declaring statistical significance at the end of the trial is $b_2 = -1.9687$, as can be seen by examining the stopping boundaries displayed in the **Simulation Parameters** tab. The interim analysis is performed when $D_1 = 167$ events are obtained. In the absence of any adaptive change, the trial will terminate when $D_2 = 334$ events are obtained. Suppose the current simulation generates a value $z_1 = 1.5$ for the logrank statistic at look 1. Since the target conditional power is 0.9, equation (56.11) takes the form

$$0.9 = 1 - \Phi \left\{ -1.9687 \sqrt{1 + \frac{167}{334 - 167}} - 1.5 \sqrt{\frac{167}{334 - 167}} - 0.5\delta \sqrt{D_2^* - 167} \right\}. \quad (56.12)$$

In order to evaluate D_2^* , however, it is necessary to specify a value for the log hazard ratio δ in equation (56.12). This parameter is of course unknown. East gives you the option to perform simulations with either the current estimate $\hat{\delta}_1$ or to use the value of δ specified under the alternative hypothesis at the design stage. The choice can be made by selecting **Estimated HR** or **Design HR** from a drop-down list of the quantity **CP Computation Based on** of the **Sample Size Re-estimation** tab.

The default value is **Estimated HR**, (or equivalently $\hat{\delta}_1 = \ln \hat{HR}_1$) and we recommend using this default until you have gained some experience with the simulation output and can judge for yourselves which option provides better operating characteristics for your studies. East uses the formula

$$\hat{\delta}_1 = \frac{z_1}{\sqrt{r(1-r)D_1}}$$

to obtain the current estimate of δ . Upon substituting $z_1 = 1.5, D_1 = 167$ and $r = 0.5$ in the above expression we obtain $\hat{\delta}_1 = 0.232$, or equivalently a hazard ratio estimate of $\exp(0.232) = 1.2611$. Substituting the estimate of $\hat{\delta}_1$ into equation (56.12) and solving for D_2^* yields $D_2^* = 656$. Since the maximum number of events has been capped at 501, this simulation will terminate the trial when the number of events reaches 501 instead of going all the way to 656 events. In this case the desired target conditional power of 0.9 will not be met. Indeed in this case the conditional power (with $\hat{\delta}_1$ being used in place of the unknown true δ) is only

$$1 - \Phi \left\{ 1.9687 \sqrt{1 + \frac{167}{333 - 167}} - 1.5 \sqrt{\frac{333}{333 - 167}} - 0.5\delta \sqrt{500 - 167} \right\} = 0.798$$

For a more detailed discussion of conditional power, including the use of a special conditional power calculator that computes conditional power accurately

56 Muller and Schafer Method

without relying on the approximate assumption that the next look will be the last one, see Chapter 57.

6. **Promising Zone Scale :** Promising Zone is such that the number of events will only be increased if the conditional power at the interim look falls in this zone. East asks you to select the scale on which the promising zone is to be defined. It can be defined based on the conditional power or the test statistic or the estimated effect size and should be specified by entering the minimum and maximum of these quantities.
Let us go ahead with the default option which is Conditional Power.
7. **Promising Zone – Min CP :** In this cell you specify the minimum conditional power (in the absence of any adaptive change) at which you will entertain an increase in the number of events. That is, you specify the lower limit of the promising zone.
8. **Promising Zone – Max CP :** In this cell you specify the maximum conditional power (in the absence of any adaptive change) at which you will entertain an increase in the number of events. That is, you specify the upper limit of the promising zone.

Suppose, for example, that the number of events is only increased in a promising zone specified by the range $0.45 \leq CP < 0.8$, and suppose that in that case, the number of events is re-estimated so as to achieve a target conditional power of 0.99. Then the Input Parameters Table will contain the entries shown below.

| | | | | | |
|--|---|----------|------|----------|-----|
| <u>T</u> arget CP for Re-estimating # of Events: | 0.99 | | | | |
| <u>P</u> romising Zone Scale: | Cond. Power ▾ | | | | |
| Promising Zone: | <table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding-right: 10px;">Min. CP:</td> <td style="border: 1px solid #ccc; padding: 2px 10px; text-align: center;">0.45</td> </tr> <tr> <td style="padding-right: 10px;">Max. CP:</td> <td style="border: 1px solid #ccc; padding: 2px 10px; text-align: center;">0.8</td> </tr> </table> | Min. CP: | 0.45 | Max. CP: | 0.8 |
| Min. CP: | 0.45 | | | | |
| Max. CP: | 0.8 | | | | |

The zone to the left of the *promising zone* ($CP < 0.45$) is known as the *unfavorable zone*. The zone to the right of the *promising zone* ($CP \geq 0.8$) is known as the *favorable zone*. In a group sequential design that includes early stopping boundaries for futility and efficacy, the unfavorable zone contains within it an even more extreme region for early futility stopping and the favorable zone contains within it an even more extreme region for early efficacy

stopping.

9. **HR Used in CP Computations:** In this cell you specify whether the simulations should utilize conditional power based on $\hat{\delta}_L$ estimated at the time of the interim analysis or should utilize the value of δ specified under the alternative hypothesis, in equations (54.21) and (56.11). The adaptive design will have rather different operating characteristics in each case. The default is to use the estimated value $\hat{\delta}_L$.

CP Computation Based on:

Estimated HR ▼

10. **Accrual Rate After Adaptation :** East gives you the option to alter the rate of enrollment after an adaptive increase in the number of events. This feature would be useful, for example, to evaluate the extent to which the follow-up time and hence the total study duration can be shortened if the rate of enrollment is increased after the adaptive change is implemented.
11. **Estimation Method** East gives you the choice of None, RCI, or BWCI methods for parameter estimation.
12. **Specify Stage II Design** Clicking on this button will bring up the following dialog box. Here you specify the desired choices for Stage-II design. The specification of alpha for the Stage-II design is the most important component for Muller and Schafer method. We will keep the default choice. The other choice is **user Specified**.
Keep all other default choices in this dialog box and click **OK**.

56 Muller and Schafer Method

The dialog box will now look as shown below.

Click on **Simulate** button. Store the simulation results in the library and see the details as shown below.

Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Rejecting H1 | | Total Simulations | | Average Sample Size | Average Number of Events | Average Accrual Duration | Average Study Duration |
|-------------|--------------------------|----------|--------------------------|----------|-------------------|----------|---------------------|--------------------------|--------------------------|------------------------|
| | Count | Row % | Count | Row % | Count | Column % | | | | |
| Futility | 0 | 0.000% | 360 | 100.000% | 360 | 3.600% | 362.753 | 167 | 17.983 | 18.036 |
| Unfavorable | 854 | 34.283% | 1637 | 65.717% | 2491 | 24.910% | 483 | 334 | 23.95 | 29.188 |
| Promising | 2984 | 85.014% | 0 | 0.000% | 3510 | 35.100% | 692.842 | 462.85 | 34.332 | 35.256 |
| Favorable | 2383 | 90.643% | 246 | 9.357% | 2629 | 26.290% | 483 | 334 | 23.95 | 29.313 |
| Efficacy | 1010 | 100.000% | 0 | 0.000% | 1010 | 10.100% | 366.781 | 167 | 18.131 | 18.181 |
| All Trials | 7231 | 72.310% | 2243 | 22.430% | 10000 | 100.000% | 540.487 | 356.347 | 26.791 | 29.837 |

Promising Zone defined as 0.3 <= CP < 0.9

Average Sample Size and Look Times (Integrated Trial)

| Look # | Average Sample Size | Average Events | Average Events | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|----------------|-----------|-------------------|-------------------|
| | | | Control | Treatment | | |
| 1 | 364.657 | 167 | 90.587 | 76.413 | 18.103 | 6.002 |
| 2 | 568.347 | 386.406 | 204.764 | 181.642 | 31.694 | 8.964 |
| Average | 540.487 | 356.347 | | | 29.837 | |

Simulation Results for Integrated Trial

| Look # | Events | Stopping For | | Total Simulations | |
|--------|---------|--------------|----------|-------------------|---------|
| | | Efficacy | Futility | Count | % |
| 1 | 167 | 1010 | 360 | 1370 | 13.700% |
| 2 | 386.406 | 6221 | 1883 | 8630 | 86.300% |
| Total | 366.347 | 7231 | 2243 | 10000 | |
| % | | 72.310% | 22.430% | | |

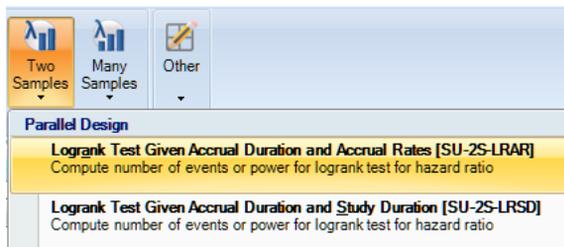
The interpretation of these results is very similar to what was described in CHW chapter section 54.5. Please also see the example for parameter estimation by BWCI and RCI methods given in section 56.3.1.

56.4 Survival Endpoint: Pancreatic Cancer Trial

A multi-center, double-blind, placebo-controlled randomized clinical trial is planned for subjects with advanced pancreatic cancer with the goal of comparing the current standard of care (gemcitabine + nap-paclitaxel) to an experimental regimen containing the two standard of care drugs plus a recombinant human enzyme. The primary endpoint is Overall Survival (OS). The study is required to have one-sided $\alpha = 0.025$, and 90% power to detect an improvement in median survival, from 8.5 months on the control arm to 12.744 months on the experimental arm, which corresponds to a hazard ratio of 0.667. The average enrollment is expected to be 15 subjects/month. We shall first create a two-look group sequential design for this study in East, and shall then show how the design may be improved by permitting an increase in the number of events and sample size at the time of the interim analysis.

56.4.1 Base Design

The base design is a two look group sequential design with a Lan and DeMets O’Brien-Fleming LD(OF) efficacy boundary, and a futility boundary for terminating if the estimated hazard ratio exceeds 1.0. To enter these design parameters into East select the design option for the **Logrank Test Given Accrual Duration and Accrual Rates** from the Tab on the menu bar as shown below.



56 Muller and Schafer Method

Enter the design inputs as shown in the **Test Parameters** tab.

Design Type: Superiority Number of Looks: 2

Test Parameters Boundary Accrual / Dropouts

Test Type: 1-Sided # of Hazard Pieces: 1 Input Method: Median Survival Times

Type I Error (α): 0.025 Hazard Ratio (Optional) Alternative

Power: 0.9 Hazard Ratio (λ_t/λ_c) 0.667

No. of Events: Computed Ratio of Medians (m_t/m_c) 1.49925

Allocation Ratio: 1 (n_t/n_c)

| Med.Surv.Time | |
|---------------|-----------------|
| Control | Treatment: Alt. |
| 8.5 | 12.744 |

Variance of Log Hazard Ratio

Null Alternative

Specify the efficacy and futility boundaries in the **Boundary** tab.

Design Type: Superiority Number of Looks: 2

Test Parameters Boundary Accrual / Dropouts

Efficacy

Boundary Family: Spending Functions Futility

Spending Function: Lan-DeMets Boundary Family: HR

Parameter: OF Non-Binding

Type I Error (α): 0.025 Binding

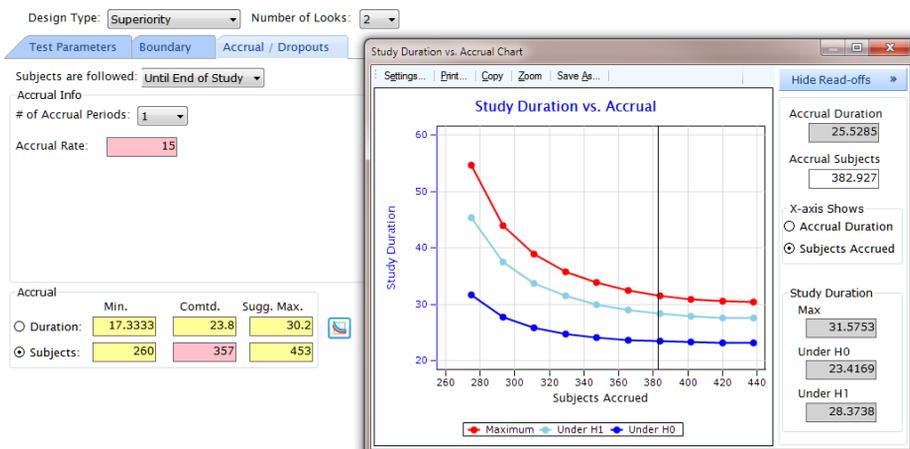
Type II Error (β): 0.1

Info. Fraction at Interim Look: 0.5 Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Futility HR |
|--------|----------------|-------------------------------------|-------------------------------------|---------------------|-------------------|-------------|
| 1 | 0.5000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.0015 | -2.9626 | 1.0000 |
| 2 | 1.0000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.0250 | -1.9686 | 0.7831 |

Specify the accrual rate 15/month in the tt Accrual/Dropouts tab, and display the

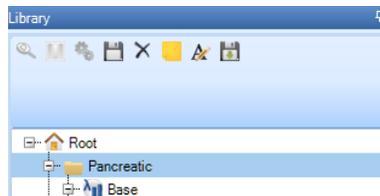
Study Duration vs. Accrual chart by clicking on its icon 



After examining this chart it is decided to enroll 360 subjects over 24 months resulting in a total study duration of about 34 months.

| Accrual | | | |
|--|---------|--------|------------|
| | Min. | Comtd. | Sugg. Max. |
| <input type="radio"/> Duration: | 17.3333 | 24 | 30.2 |
| <input checked="" type="radio"/> Subjects: | 260 | 360 | 453 |

Click the  button to compute and store this design temporarily in the **Output Preview** window, and then save the design permanently in the **Library** by clicking the  button. Rename the saved design by the name **Base**. Also rename the workbook, currently named as **Wbk1**, by the name **Pancreatic** and save it on your computer. The library should now look as shown.



56 Muller and Schafer Method

You may view a summary of this design by clicking on the  icon.

| Pancreatic:Base | |
|---|--------------------|
| Mnemonic | SU-25-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 2 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Attained α | 0.0247 |
| Power | 0.9008 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.667 |
| Var (Log HR) | Null |
| Boundary Parameters | |
| Info. Fraction at Interim Look | 0.5 |
| Efficacy Boundary | LD (OF) |
| Futility Boundary | HR (NB) |
| Accrual & Dropout Parameters | |
| Accrual Rate | 15 |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 360 |
| Expected Under H0 | 315.877 |
| Expected Under H1 | 342.292 |
| Events | |
| Maximum | 260 |
| Expected Under H0 | 194.802 |
| Expected Under H1 | 225.282 |
| Study Duration | |
| Maximum | 32.925 |
| Expected Under H0 | 23.853 |
| Expected Under H1 | 29.361 |
| Accrual Duration | |
| Maximum | 24 |
| Expected Under H0 | 21.058 |
| Expected Under H1 | 22.819 |

Alternatively, you may view this design in greater detail by clicking on the  icon.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

| Test Parameters | |
|--------------------------------|-------------|
| Design ID | Base |
| Design Type | Superiority |
| Number of Looks | 2 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Attained α | 0.025 |
| Power | 0.901 |
| Model Parameters | |
| HR = λ_1/λ_0 | |
| Under H0 | 1 |
| Under H1 | 0.667 |
| Ratio of % Surv. at Period #1: | 1.028 |
| Cum. % Surv. by Time = | 1 |
| Control (S_c) | 92.169 |
| Treatment (S_t) | 94.706 |
| Var (Log HR) | Null |
| Allocation Ratio (n_1/n_0) | 1 |
| Boundary Parameters | |
| Efficacy Boundary | LD (OF) |
| Futility Boundary | HR (NB) |
| Accrual / Dropouts Parameters | |
| Accrual Rate | 15 |
| Dropout | No |

Variable Follow-Up Design: All subjects are followed until failure, drop out or end of study.

Sample sizes and events have been rounded.

☞ **Sample Size Information**

| | Control Arm | Treatment Arm | Total |
|-------------------------|-------------|---------------|---------|
| Sample Size (n) | | | |
| Maximum | 179 | 181 | 360 |
| Expected H1 | 171.146 | 171.146 | 342.292 |
| Expected H0 | 157.938 | 157.938 | 315.877 |
| Events (s) | | | |
| Maximum | 141 | 119 | 260 |
| Expected H1 | 128.981 | 104.911 | 233.892 |
| Expected H0 | 100.075 | 100.075 | 199.902 |
| Maximum Information (I) | | | 65 |

☞ **Accrual and Study Duration**

| | Accrual Duration | Study Duration |
|-------------|------------------|----------------|
| Maximum | 24 | 32.925 |
| Expected H1 | 22.819 | 29.361 |
| Expected H0 | 21.058 | 23.853 |

☞ **Stopping Boundaries: Look by Look**

| Look # | Info. Fraction (s/s_max) | Events (s) | Cumulative α Spent | Cumulative β Spent | Boundaries | |
|--------|--------------------------|------------|---------------------------|--------------------------|------------|-------------|
| | | | | | Efficacy Z | Futility HR |
| 1 | 0.5 | 130 | 0.002 | 0.01 | -2.963 | 1 |
| 2 | 1 | 260 | 0.025 | 0.099 | -1.969 | 0.783 |

☞ **Events, Sample Size, Pipeline and Analysis Times: Look by Look (Under H0)**

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (s) | Pipeline (n-s) | Analysis Time | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|------------|----------------|---------------|---|----------|
| | | | | | | Efficacy | Futility |
| 1 | 0.5 | 273 | 130 | 143 | 18.135 | 0.002 | 0.5 |
| 2 | 1 | 360 | 260 | 100 | 29.606 | 0.023 | 0.475 |

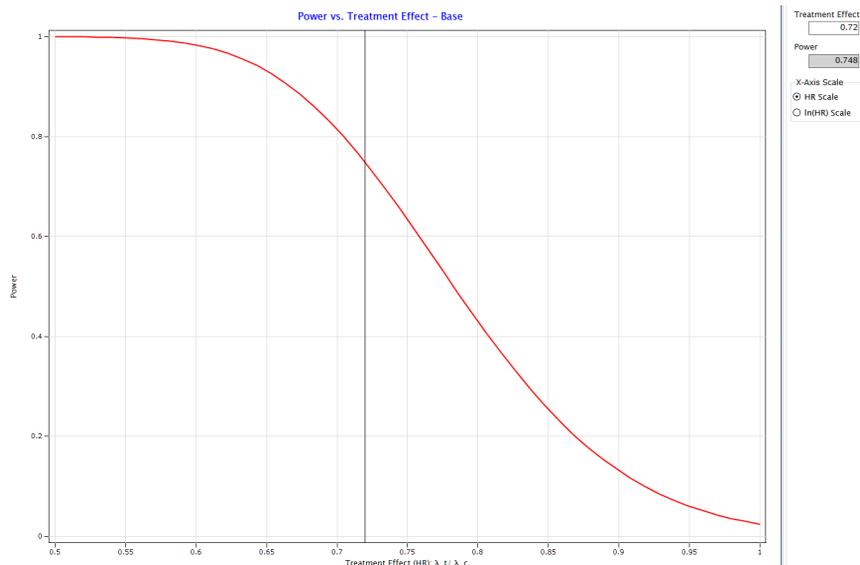
☞ **Events, Sample Size, Pipeline and Analysis Times: Look by Look (Under H1)**

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (s) | Pipeline (n-s) | Analysis Time | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|------------|----------------|---------------|---|----------|
| | | | | | | Efficacy | Futility |
| 1 | 0.5 | 294 | 130 | 164 | 19.58 | 0.257 | 0.01 |
| 2 | 1 | 360 | 260 | 100 | 32.925 | 0.644 | 0.089 |

You may also examine the various charts associated with this design by activating them from the  icon. For example it is interesting to examine **Power versus Treatment Effect** chart on the HR scale. Notice that if the actual hazard ratio is

56 Muller and Schafer Method

0.72 instead of 0.67, then the power deteriorates from 90% to 74.8%.



56.4.2 Simulate without Adaptation

Click on the simulation icon . You will be taken to the simulation window which contains four tabs:

- The **Test Parameters** tab

Test Parameters
Response Generation
Accrual / Dropouts
Simulation Controls

Trial Type:

Test Type:

Fix at Each Look:

Total No. of Events:

Test Statistic:

| Look # | Info. Fraction | Cum. α Spent | Efficacy Z | Futility HR |
|--------|----------------|---------------------|------------|-------------|
| 1 | 0.500 | 0.002 | -2.963 | 1.000 |
| 2 | 1.000 | 0.025 | -1.969 | 0.783 |



Do not make any changes to the entries in this tab.

■ The **Response Generation** tab

| | | | |
|----------------------|---------------------|--------------------|---------------------|
| Test Parameters | Response Generation | Accrual / Dropouts | Simulation Controls |
| Survival Information | | | |
| # of Hazard Pieces | 1 | Input Method: | Cum. % Survival |
| By Time | Cum. % Survival | | Hazard Ratio |
| | Control | Treatment | |
| 1 | 92.169 | 94.706 | 0.667 |

Note: Same hazard rates apply before and after time 1.

In order to study the operating characteristics of the adaptive design we will simulate the design with a hazard ratio of 0.72. Therefore please change the value for the hazard ratio from 0.67 to 0.72.

| | | | |
|----------------------|---------------------|--------------------|-----------------------|
| Test Parameters | Response Generation | Accrual / Dropouts | Simulation Controls |
| Survival Information | | | |
| # of Hazard Pieces | 1 | Input Method: | Median Survival Times |
| | Med.Surv.Time | | Hazard Ratio |
| | Control | Treatment | |
| | 8.5 | 11.8056 | 0.72 |

■ The **Accrual/Dropouts** tab

| | | | |
|---------------------------------------|---------------------|--------------------|---------------------|
| Test Parameters | Response Generation | Accrual / Dropouts | Simulation Controls |
| Sample Size: | 360 | | |
| Total No. of Events: | 260 | | |
| Subjects are followed: | Until End of Study | | |
| Distribution of Accrual Time: Uniform | | | |
| Accrual Info | | | |
| # of Accrual Periods: | 1 | Input Method: | Accrual Rates |
| Accrual Rate: | 15 | | |
| Piecewise Dropout Information | | | |
| # of Pieces: | 0 | | |

Do not make any changes to these entries.

56 Muller and Schafer Method

■ The **Simulation Controls** tab

Test Parameters | Response Generation | Accrual / Dropouts | **Simulation Controls**

Number of Simulations:

Refresh Frequency:

Random Number Seed

Clock

Fixed

Suppress All Intermediate Output

Pause after Refresh

Pause at End

Output Options

Output Type:

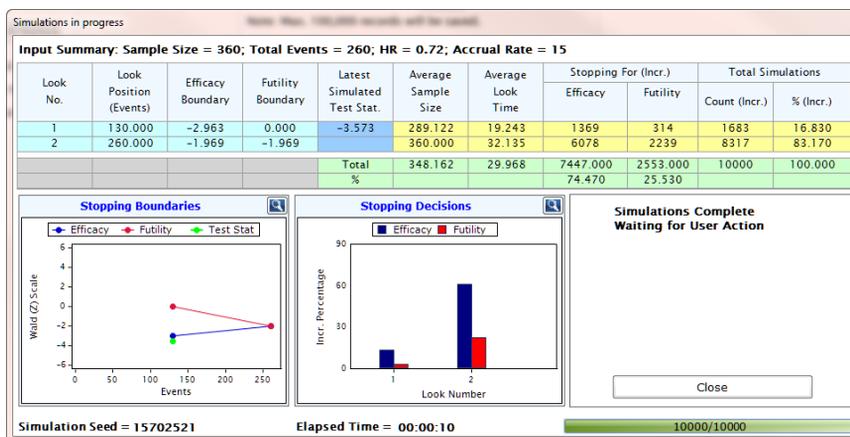
Save summary statistics for every simulation run

Save subject level data for simulation runs

Note: Max. 100,000 records will be saved.

Change the number of simulations to 100000 for greater Monte Carlo accuracy.

Thus far the only change that we have made the original design is to increase the hazard ratio from 0.67 to 0.72 for the simulations. We can simulate the design with the increased hazard ratio by clicking on the **Simulate** button. Notice that the power, based on 100000 simulated trials with HR=0.72, is only 74.47%.



Press the **Close** button, then move the simulated design from the **Output Preview** window to the **Library** and save the simulated design in the library.

Rename it as **Sim-0.72**.

| Pancreatic:Base:Sim-0.72 | |
|---|--------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| Test Type | 1-Sided |
| Power | 0.745 |
| No. of Looks | 2 |
| Test Statistic | Logrank |
| Model Parameters | |
| No. of Hazard Pieces | 1 |
| Hazard Ratio | 0.72 |
| Boundary Parameters | |
| Spacing of Looks | User Specified |
| Efficacy Boundary | User Specified |
| Futility Boundary | User Specified |
| Accrual & Dropout Parameters | |
| Subjects Followed-up | Until End of Study |
| Accrual Rate | 15 |
| Sample Size | |
| Maximum | 360 |
| Events | |
| Maximum | 260 |
| Simulation Results (Overall) | |
| Average Study Duration | 29.968 |
| Average Sample Size | 348.162 |

You may examine various other operating characteristics by examining the more detailed simulation output that is available by clicking on 

56.4.3 Adaptive Simulation

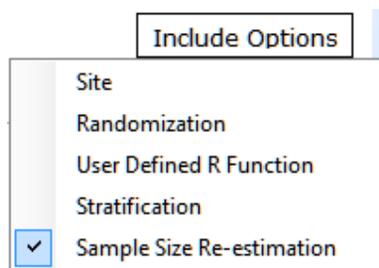
There is some uncertainty about the hazard ratio at which to power this study. A hazard ratio of 0.72 is still clinically meaningful. But as we just showed, the power at that hazard ratio is only about 75%. We can recover the lost power by implementing an adaptive increase to the number of events and sample size at the interim analysis time point. We shall do this by simulation using the Müller and Schäfer method. Return to the simulation input window. The easiest way to do this is to click on the Input icon

 located on the task bar at the bottom of the current window. This action will always open the input window that was most recently used. Alternatively, you can open the input window by selecting **Sim-0.72** in the library and clicking on the **Edit Simulation** icon . Either way you will be taken back to the

56 Muller and Schafer Method

Simulation Inputs window with the four tabs.

At the far right corner of this window is the **Include Options** button. Click on this button and select **Sample Size Re-estimation** from the drop-down list.



An additional tab labelled **Sample Size Re-Estimation** is created. Select that tab and choose the **Muller and Schafer** radio button. You'll be taken to the **Sample Size Re-estimation** window.

| Stage I | |
|---|--------------------|
| Mnemonic | SU-25-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 2 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.901 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.667 |
| Var (Log HR) | Null |
| Boundary Parameters | |
| Info. Fraction at Interim Look | 0.5 |
| Efficacy Boundary | LD (OF) |
| Futility Boundary | HR (NB) |
| Accrual & Dropout Parameters | |
| Accrual Rate | 15 |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 360 |
| Events | |
| Maximum | 260 |
| Study Duration | |
| Maximum | 32.925 |
| Expected Under H0 | 23.853 |
| Expected Under H1 | 29.361 |
| Accrual Duration | |
| Maximum | 24 |
| Expected Under H0 | 21.058 |
| Expected Under H1 | 22.819 |

Let us examine this window carefully, for it conveys a large amount of information. It

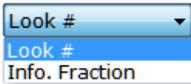
is convenient separate this window into two panels; a *Left Panel* and a *Right Panel*. The Left Panel, displayed below, is primarily for specifying the criteria that will be used to determine whether or not the current design should be adapted.

The Right Panel, displayed below, is for specifying how the original design will be adapted if indeed the criteria for adaptation that have been entered into the left panel are satisfied. At present only the original design (i.e., the **Base** design saved in the library) is shown in the Right Panel. For simulation purposes it is referred to as the **Stage I** design. If, in any simulation round, the adaptation criteria specified in the Left Panel are not met, only the **Stage I** design that will be simulated.

Now we have stipulated on the top line of the Left Panel that the Stage I design would be adapted at look 1.

Therefore, if the the adaptation criteria specified in the Left Panel are met, then the remainder of the **Stage I** design beyond look 1 will be adapted in accordance with specifications that will be provided through the creation of a Stage II design. We shall explain how the **Stage II** design is created shortly.

Left Panel: Specification of Criteria for Adaptation We first we enter the inputs into the Left Panel. At the top of the Left Panel we specify when the adaptation will take place. This may either be specified in terms of the **Look #** or **Information Fraction**

Adapt at: 

For this example, we will make the adaptation after completing Look 1.

Adapt at:  

We next specify the maximum allowable number of events and the maximum allowable sample size should the trial be adapted. This is achieved by specifying an appropriate multiplier in the **Max. # of Events** and **Max. Sample Size** edit boxes. We will use a multiplier of 1.5 for both the events and the sample size.

| | | |
|--|-----|-----|
| Max. # of Events if Adapt (multiplier; total #): | 1.5 | 390 |
| Max. Sample Size if Adapt (multiplier; total #): | 1.5 | 540 |

56 Muller and Schafer Method

By this specification we have placed a cap on the magnitude of the increase in events and sample size. For example, if in any simulation round the decision rule used to re-estimate events produces the value 400, the re-estimated number of events will be nevertheless be truncated to 390.

The next specification is the **Upper Limit on Study Duration**. By default it is three times the maximum study duration of the **Stage I** design. It is provided as a precaution against excessive prolongation of a simulated trial in case of very slow arrival of events and its default value is typically not altered. The **Stage I** design displayed on the Right Panel shows a maximum study duration of 32.925 months.

| Study Duration | |
|-------------------|--------|
| Maximum | 32.925 |
| Expected Under H0 | 23.853 |
| Expected Under H1 | 29.361 |

Therefore the **Upper Limit on Study Duration** is $3 \times 32.925 = 98.776$.

Upper Limit on Study Duration: 98.776

The next three entries describe the criteria for trial adaptation.

Promising Zone Scale: Cond. Power CP

| | | |
|-----------|----------|-----|
| Promising | Min. CP: | 0.3 |
| Zone: | Max. CP: | 0.9 |

The adaptation criteria in East are based on the promising zone design proposed by Mehta and Pocock (2011). The interim analysis results are partitioned into three zones; **Unfavorable**, **Promising** and **Favorable**. If the interim results fall in the unfavorable or favorable zones, there is no adaptation. But if they fall in the promising zone, the trial is adapted. In the Müller and Schäfer method permits the permissible adaptations go beyond mere sample size re-estimation. One may, in addition increase the number and spacing of the future looks and also alter the spending function. The partitioning of the interim sample space into zones can be based on three different scales – **Conditional Power**, **Test Statistic** or **Estimated HR**. One

can pick the desired scale from a drop-down list as shown below.

Promising Zone Scale: Cond. Power ▾
Cond. Power
Test Statistic
Estimated HR

The three scales are in one-to-one correspondence, so that the selection of the scale is simply a matter of choosing the one that is easiest to interpret in a given situation. In the current example the **Conditional Power** scale has been chosen. Accordingly the promising zone is defined as the region of the interim analysis sample space in which the conditional power is between 0.3 and 0.9.

To see this same zone on the hazard ratio scale, choose **Estimated Hazard** form the drop-down choice of scales for the promising zone.

Promising Zone Scale: Estimated HR ▾

| | |
|-----------------|---|
| Promising Zone: | Min. HR: 0.7001 |
| | Max. HR: 0.8202 |

It is seen that on the estimated HR scale the same promising zone corresponds to the interim estimate of the hazard ratio lying between 0.8202 and 0.7001. On the test statistic (or Wald statistic, or Z-statistic) scale the promising zone corresponds to $-2.0328 \leq Z \leq -1.1298$.

Promising Zone Scale: Test Statistic ▾

| | |
|-----------------|--|
| Promising Zone: | Min. Test Stat.: -2.0328 |
| | Max. Test Stat.: -1.1298 |

If the promising zone is defined in terms of conditional power, one needs to specify what hazard ratio will be assumed for computing conditional power. The conditional power calculations may be performed either with the interim estimate of hazard ratio or with the value of hazard ratio that was used to create the base design (i.e., 0.667). The default choice is **Estimated HR**. We will use the default specification.

CP Computation Based on: Estimated HR ▾

The next entry is used to specify the rate of accrual after the adaptation. We shall

56 Muller and Schafer Method

assume that there will be no change in the accrual rate.

Accrual Rate After Adaptation:

The final entry in the Left Panel is a specification of the method for estimating the hazard ratio at the final analysis that adjusts for the fact that an adaptive group sequential design was used. There are three choices; none, RCI method and BWCI method.

Estimation Method:
None
RCI
BWCI

The purpose of this input is to verify by simulation the properties of the RCI method (Mehta, Bauer, Posh, Brannath, *Statistics in Medicine*, 2007) and the BWCI method (Gao, Liu, Mehta, *Statistics in Medicine*, 2013) for computing point estimates and confidence intervals that adjust for having used an adaptive group sequential design. We shall use the **None** option for the present since the RCI and BWCI options are intended as tools for methodological research rather than for the actual design of a trial.

Right Panel: Specification of the Stage II Design Next we consider the Right Panel. At present it displays the **Stage I** or **Base** design in summary form. The Stage I design has 2 looks and we have indicated that we will be altered if the adaptation criterion of being in the promising zone is met. We must, however, specify to East precisely how the remainder of the trial beyond look 1 will be adapted if the adaptation criterion is satisfied. As we have explained in Section 56.1.1, although we are dealing with a single trial that is adapted at an interim analysis, it is more convenient to specify the portion of this trial that is implemented after the adaptation as a separate **Stage II** trial having a type-1 error equal to the conditional type-1 error of the Stage I trial obtained at the time of the adaptation. This is the essence of the Müller and Schäfer method of adapting an on-going study while preserving the overall type-1 error. Accordingly, click on the

button. You are now taken to the following dialog box where you must specify the

design parameters of the Stage II design.

You must specify the Stage II design in this dialog box. The complete design specification consists of type-1 error, power, number of looks, hazard ratio and efficacy/futility boundaries. East will then compute the number of events that are needed to attain the specified power. Because this dialog box is used for simulation only power and number of looks are specified explicitly. All other quantities depend on the data that are obtained from the Stage I trial at the time of the adaptive look, and therefore vary from simulation to simulation. Let us illustrate by simulating an adaptive trial in which we will adapt at look 1 of the Stage I design. The adaptation will consist of an increase in the number of events and sample size, and one additional interim look, resulting in a two-look Stage II design with a Pocock error spending function for the efficacy boundary and a hazard ratio of 1 for the futility boundary. We enter the appropriate inputs as follows:

- Specify that the Stage II design has two looks.
This specification causes a **Boundary** tab to appear.
- Enter the Pocock efficacy boundary and the HR=1 futility boundary in the

56 Muller and Schafer Method

Boundary tab as shown.

| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Futility HR |
|--------|----------------|-------------------------------------|-------------------------------------|---------------------|-------------------|-------------|
| 1 | 0.500 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | Computed | Computed | 1.000 |
| 2 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | Computed | Computed | Computed |

The actual amount of α available for spending and the actual efficacy boundary cannot be displayed. These design parameters depend on the conditional type-1 error of the Stage I trial at the time of the adaptation and will therefore vary from simulation to simulation in accordance with the Müller and Schäfer method.

- Return to the **Test Parameters** tab. This input dialog box requires the following specifications:
 1. Specification of alpha: There are two choices.

Specification of alpha:

For a Müller and Schäfer design the correct choice is **Cond. Type 1 Error From Stage-I**. Only this choice will ensure that the overall type-1 error of the adaptive design is preserved. The alternative choice, **User Specified**, has been included simply for illustrative purposes, to demonstrate that if a fixed type-1 error is specified in an adaptive trial, it would not be preserved. Thus select **Cond. Type 1 Error from Stage -I** and note that it will vary from simulation to simulation depending on the value of the test statistic obtained in the **Stage I** trial at look 1.

2. Specification of HR: Here too there are two choices

Specification of HR

In this case the choice depends on the user's preference. If **Estimated from Stage-I** is selected the sample size will be computed by

assuming that the hazard ratio that was obtained at Stage I at the time of the adaptive interim look is the true hazard ratio. Therefore it will vary from simulation to simulation. Alternatively one might desire to simulate the adaptive design with a fixed hazard ratio, say the HR that was specified for the original design. Here we will select the **Estimated from Stage-I** option thereby letting the data from the first stage determine the value of HR from simulation to simulation.

3. Power. This is the desired power for the Stage II trial. However this power may not be attainable in every simulated trial. Depending on the type-1 error and hazard ratio that have been estimated from Stage I, East will compute the number of events, say D_r , that are *required* to attain the desired power in the **Stage II** trial. Now you have already specified in the Left Panel of the **Sample Size Re-estimation** tab the maximum number of events if the trial is adapted – in this case 390 events: Therefore if, in any simulation, $D_r > 390$, East will only generate 390 events, and the desired power will not be attained. More generally let D_{max} denote the *maximum* allowable number of events specified in the **Sample Size Re-estimation** tab. Then D_a , the *actual* number of events that East will generate in any simulation, is given by

$$D_a = \min(D_r, D_{max})$$

Let N_{max} denote the maximum sample size if the trial is adapted, in this case 540.

East will generate patient arrivals until either the D_a events have arrived or N_{max} subjects have arrived. In the latter case East will follow the N_{max} subjects until D_a events have arrived.

For the current simulation experiment enter the value 0.999 into the edit box for power.

Power:

0.999

By selecting such a high value for power we are assured that in every simulated trial the required number of events, D_r , will hit the cap $D_{max} = 390$; that is, $D_a = D_{max} = 390$ in every simulation. Thus this choice for power is an implicit way of specifying that there will be a one-time 50% increase in the number of events if the Stage I results fall in the promising zone at the time of the interim analysis.

- The inputs in the **Test Parameters** tab for the specification of the Stage II design now look as follows.

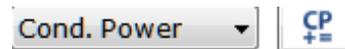
56 Muller and Schafer Method

To complete the specification, press the **OK** button. East will return you to the **Sample Size Re-estimation** tab and will display both the Stage I and Stage II designs side by side in the Right Panel.

| | Stage I | Stage II |
|---|--------------------|-------------|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 2 | 2 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | Computed |
| Power | 0.901 | 0.999 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Hazard Ratio (Alt.) | 0.667 | Computed |
| Var (Log HR) | Null | Null |
| Boundary Parameters | | |
| Info. Fraction at Interim Look | 0.5 | 0.5 |
| Efficacy Boundary | LD (OF) | LD (PK) |
| Futility Boundary | HR (NB) | HR (NB) |
| Accrual & Dropout Parameters | | |
| Accrual Rate | 15 | NA |
| Subjects are Followed | Until End of Study | NA |
| No. of Accrual Periods | 1 | NA |
| No. of Dropout Pieces | 0 | NA |
| Sample Size | | |
| Maximum | 360 | Computed |
| Events | | |
| Maximum | 260 | Computed |
| Study Duration | | |
| Maximum | 32.925 | NA |
| Expected Under H0 | 23.853 | NA |
| Expected Under H1 | 29.361 | NA |
| Accrual Duration | | |
| Maximum | 24 | NA |
| Expected Under H0 | 21.058 | NA |
| Expected Under H1 | 22.819 | NA |

Displaying Stage I and Stage II as Single Integrated Design Although we are dealing with a single integrated design we have regarded the remainder of the trial after the adaptive look at Stage I as a separate Stage II trial whose type-1 error is equal to the conditional type-1 error of the Stage I design. This is extremely convenient for

design purposes because we can use all the functionality that already exists in East to design a separate Stage II trial without considering how it will be integrated with the existing Stage I design. On the other hand, this artificial separation of a single design into two separate designs makes it difficult to visualize the stopping boundaries of the integrated design or the trajectory traced by the test statistic during the interim monitoring phase of the study. Thus, although it is technically correct to monitor the Stage II design independently, with the test statistic starting out at the value zero, it is not very intuitive to do so. To gain a better understanding of the how the Stage I and Stage II designs are integrated in to a single design we have provided the **CP+–** button to the right of **Cond. Power**.



Clicking on this button will open up a conditional power calculator. By default, the calculator will open with the radio button for the Stage I design selected

and the radio button for specifying that the **HR to be Used in Conditional Power Computation** will be estimated from the data rather than specified separately by the user.

With this choice the observed value of the test statistic Z and the estimate of HR, say \hat{HR} , at the time of the adaptive look in the Stage I trial are in one to one correspondence through the relationship due to Schoenfeld (Biometrika, 1981)

$$Z = \ln(HR) \sqrt{Dr(1-r)}$$

where D is the number of events at the time of the adaptive look (here $D = 130$) and r is the randomization fraction for allocating subjects to the two treatment arm (here $r = 0.5$). We can specify either \hat{HR} or Z in the appropriate edit box and the calculator will output the corresponding value of conditional power. If the conditional power falls in the promising zone, the trial will be adapted through the creation of a Stage II design. Otherwise the trial will continue as planned. Below we provide three examples to show how different values of Z result in different Stage II designs and how the two Stages may be viewed as a single integrated adaptive design.

Example 1: $Z = -1.5187$ In the view below, if $Z = -1.5187$, then $HR = 0.7661$ (by Schoenfeld's formula above) and the conditional power is computed as 0.6.

56 Muller and Schafer Method

Conditional Power Calculator

MS Simulation Calculator for

Stage I Design
 Stage II Design

Input

Current Look:
 Current # of Events:

Input/Output

HR to be Used in Conditional Power Computation

Arbitrary (HR) Estimated (HR, z)

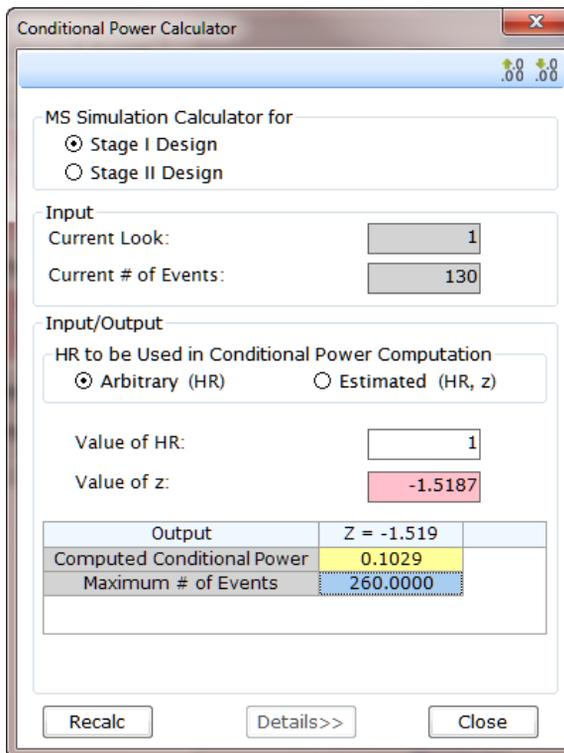
Value of HR:
 Value of z:

| Output | Z = -1.519 |
|----------------------------|------------|
| Computed Conditional Power | 0.600 |
| Maximum # of Events | 260.000 |

Recalc Details>> Close

This would imply that the interim result has fallen in the promising zone (CP between 0.3 and 0.9) and hence the remainder of the Stage I trial should be adapted. To obtain the conditional type-1 error, (also referred to as the Conditional Error Rate (CER) or the conditional Rejection Probability (CRP)) of the remainder of the Stage I trial, select the **Arbitrary HR** radio button. Now HR and Z are no longer in one to one correspondence so that we can set **Value of HR** to 1 and separately set **Value of**

z to -1.5187 as shown below.



The calculator reveals that the conditional type-1 error at $HR = 1$ and $Z = -1.5187$ is 0.1029. Since $HR = 1$ corresponds to the null hypothesis, 0.1029 is the conditional type-1 error of the Stage I trial if $Z = -1.5187$ is observed at look1. Thus the amount of α we would use for the Stage II design is 0.1029. Now set the radio button for HR back to **Estimated HR**, z and set $Z = -1.5187$. Once again, Z and HR are in one to one correspondence via Schoenfeld's formula so that $HR = 0.7661$. Now

56 Muller and Schafer Method

select the **Stage II Design** radio button.

Conditional Power Calculator

MS Simulation Calculator for

Stage I Design
 Stage II Design

Input

Current Look in Stage I:
 Current # of Events in Stage I:

Input/Output

HR to be Used in Conditional Power Computation

Arbitrary (HR) Estimated (HR, z)

Value of HR:
 Value of z:

| Output | Z = -1.519 |
|--------------------------|------------|
| Conditional Type I Error | 0.1029 |
| CP (Stage I) | 0.6000 |
| CP (Stage II) | 0.7639 |
| Events (Stage II) | 260.0000 |
| Events (Integrated) | 390.0000 |

* In promising zone and trial adapted

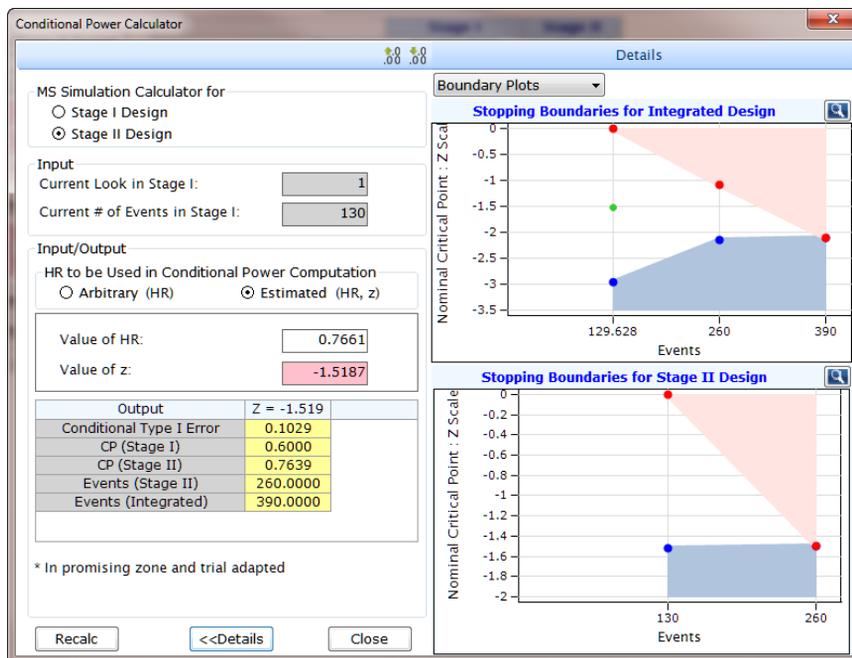
Recalc Details>> Close

The following output is obtained:

- Conditional Type I Error = 0.1029. This is the amount of α or conditional error rate that will be available for the Stage II design
- CP(Stage I) = 0.6. This is the conditional power of the Stage I design if the observed value at look 1 is $Z = -1.5187$ and $HR = 0.7661$. This puts the look 1 result in the promising zone so that the trial may be adapted
- CP(Stage II) = 0.7639; Events(Stage II) = 260; Events(Integrated) = 390. These outputs show that the interim analysis of the Stage I trial is in the promising zone and therefore the Stage II design is invoked. Although the Stage II design is intended to achieve 99% power at the estimated value of $HR = 0.7661$, it

cannot do so because of the cap of 390 events on the integrated design (or 260 events on the Stage II design). Because the number of events cannot be increased further, the power of the Stage II design is 0.7639 and not 0.999.

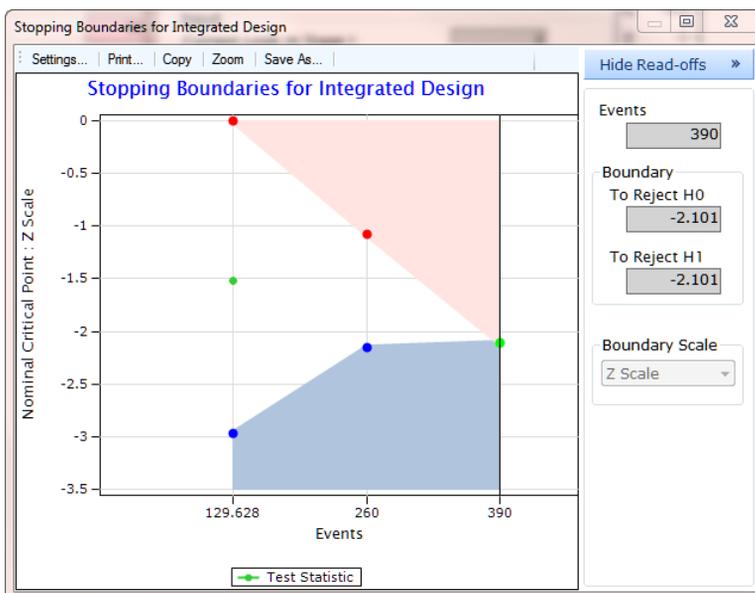
If we click on the **Details** button we get more insight into the Stage II and integrated designs.



Each of the charts in the right panel can be magnified by clicking in its icon. The top panel shows the integrated design in which the Stage I design was adapted after 130

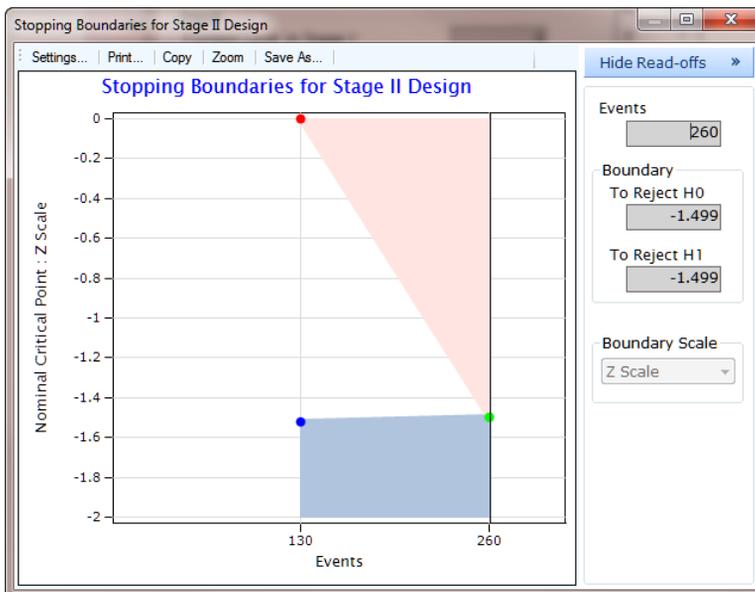
56 Muller and Schafer Method

events were observed.



The green dot is the observed value of the test statistic at look 1, $Z = -1.519$. It falls in the promising zone. The lower panel shows the two-look Stage II design with the Pocock efficacy boundary and the HR=1 futility boundary. The type-1 error of the Stage II design is 0.1029, which corresponds to the conditional type-1 error from Stage

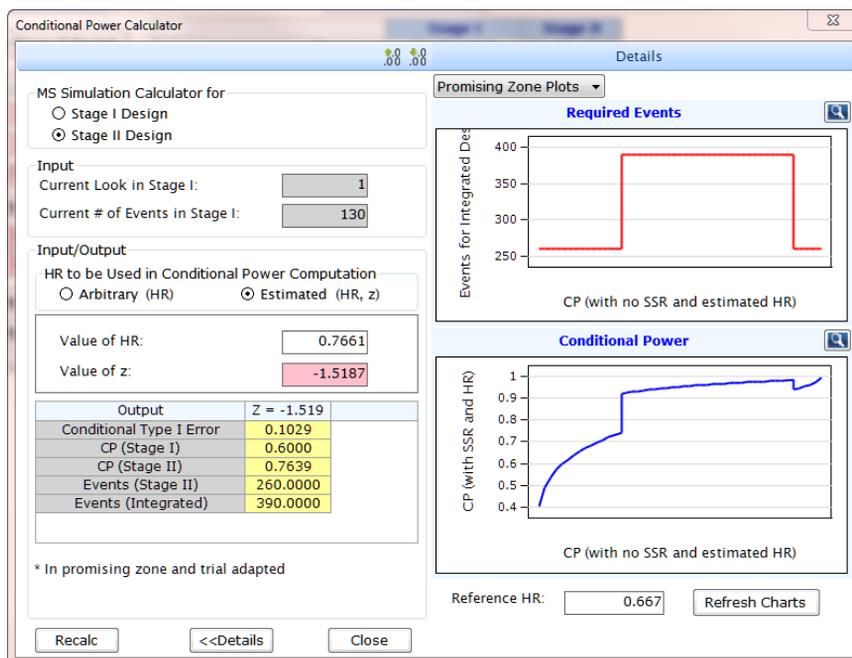
I.



It is instructive to view the adaptation rule and its impact on conditional power graphically. This can be achieved by switching from **Boundary Plots** to

56 Muller and Schafer Method

Promising Zone Plots from the drop-down list as shown below.



These are the familiar **Promising Zone Plots** that have been well documented in Chapter 54. The top plot displays the promising zone (CP between 0.3 and 0.9) on the X-axis and the number of events for the integrated design on the Y-axis. Outside the

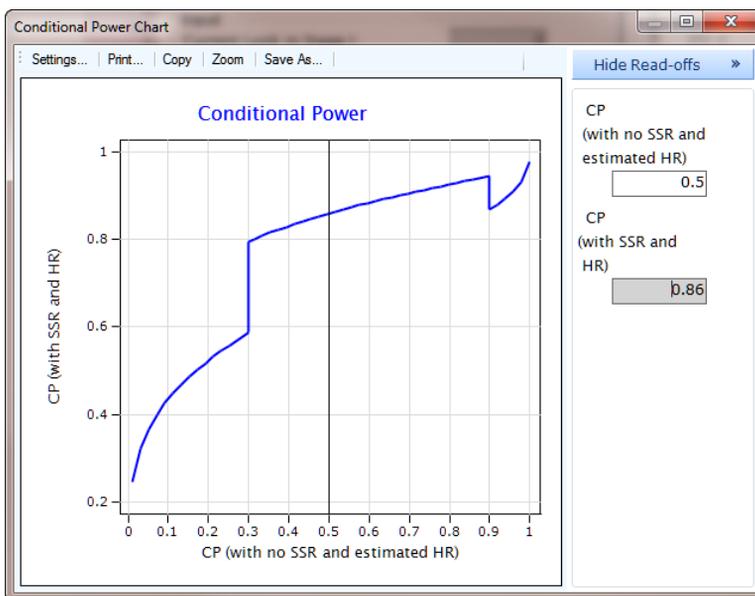
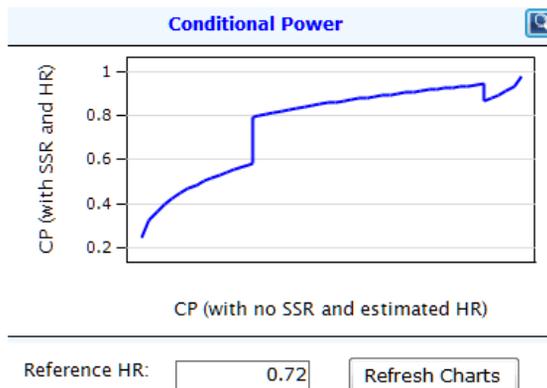
promising zone the number of events is 260, rising to 390 in the promising zone.



The X-axis of the bottom plot is the same as for the top plot and shows, the conditional power based on the current value of the test statistic and the current estimate of the hazard ratio. The Y-axis shows the conditional power if the number of events are increased in accordance with the rule implied by the top plot, and under the hazard

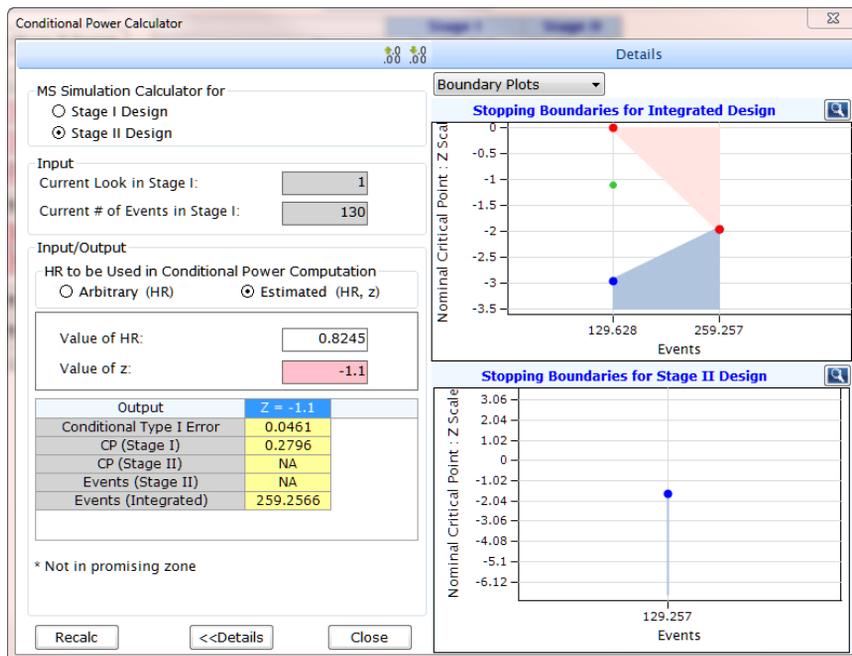
56 Muller and Schafer Method

ratio specified in the **Reference HR** edit box.



Example 2: $Z = -1.1$ For additional insight enter the value $Z = -1.1$ into the CP calculator and press the **Recalc** button. This time the result from the Stage I design is not in the promising zone. Therefore the integrated design and the Stage I design are identical with a total of 260 events. The Stage II design is simply the continuation of the Stage I design for an additional 130 events and has the same final critical value of

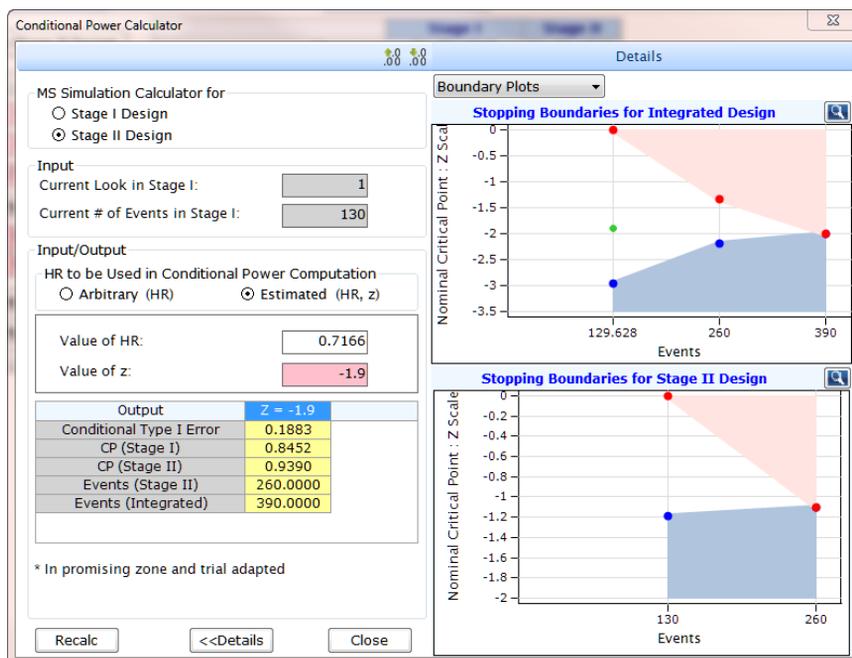
-1.9686.



Example 3: $Z = -1.9$ Finally, enter the value $Z = -1.9$ (corresponding to $HR = 0.7166$) into the calculator. Now $CP = 0.8452$ which is in the promising zone.

56 Muller and Schafer Method

Thus the trial is adapted.



The conditional type-1 error from stage I that is utilized in the Stage II design is 0.1883. Unlike Example 1, the Stage II design achieves the full 90% power with 260 events. The pre-specified cap of 360 events for the integrated design (or 260 events for the stage II design) was not exceeded in the computation of events required to obtain 90% power for the Stage II design.

To further your understanding of the adaptive design you might find it helpful to enter additional values of Z or HR into the conditional power calculator and view the resulting numerical and graphical outputs. When you are done with exploring the properties of the integrated and Stage II design in this manner, press the **Close** button to return to the **Sample Size Re-estimation** tab of the simulation inputs window.

Simulation Results

Test Parameters

Use Adaptation Method
 CHW CDL Müller and Schäfer

Adapt at:

Max. # of Events if Adapt (multiplier; total #):

Max. Sample Size if Adapt (multiplier; total #):

Upper Limit on Study Duration:

Promising Zone Scale:

Promising Zone:

CP Computation Based on:

Accrual Rate After Adaptation:

Estimation Method:

Specify Stage II Design

| | Stage I | Stage II |
|---|--------------------|-------------|
| Mnemonic | SU-25-LRAR | SU-25-LRAR |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 2 | 2 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | Computed |
| Power | 0.901 | 0.999 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Hazard Ratio (Alt.) | 0.667 | Computed |
| Var (Log HR) | Null | Null |
| Boundary Parameters | | |
| Info. Fraction at Interim Look | 0.5 | 0.5 |
| Efficacy Boundary | LD (OF) | LD (PK) |
| Futility Boundary | HR (NB) | HR (NB) |
| Accrual & Dropout Parameters | | |
| Accrual Rate | 15 | NA |
| Subjects are Followed | Until End of Study | NA |
| No. of Accrual Periods | 1 | NA |
| No. of Dropout Pieces | 0 | NA |
| Sample Size | | |
| Maximum | 360 | Computed |
| Events | | |
| Maximum | 260 | Computed |
| Study Duration | | |
| Maximum | 32.925 | NA |
| Expected Under H0 | 23.853 | NA |
| Expected Under H1 | 29.361 | NA |
| Accrual Duration | | |
| Maximum | 24 | NA |
| Expected Under H0 | 21.058 | NA |
| Expected Under H1 | 22.819 | NA |

Now that you have completed the specification of the adaptive design through the Stage I and Stage II specification in the **Sample Size Re-estimation** tab, let us evaluate its operating characteristics by simulation. The design is shown below.

This two-stage design will be simulated 100,000 times. In each simulation, look 1 of the Stage I design will be taken after 130 events and if the resulting conditional power based on the estimated hazard ratio lies in the promising zone (conditional power of the Stage I design between 30

Test Parameters

Response Generation

Accrual / Dropouts

Survival Information

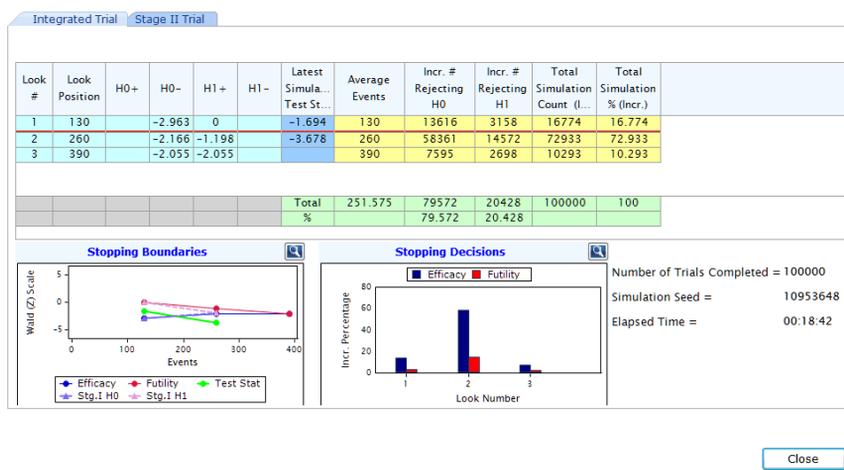
of Hazard Pieces Input Method:

| Med.Surv.Time | | Hazard Ratio |
|----------------------------------|-------------------------------------|-----------------------------------|
| Control | Treatment | |
| <input type="text" value="8.5"/> | <input type="text" value="11.806"/> | <input type="text" value="0.72"/> |

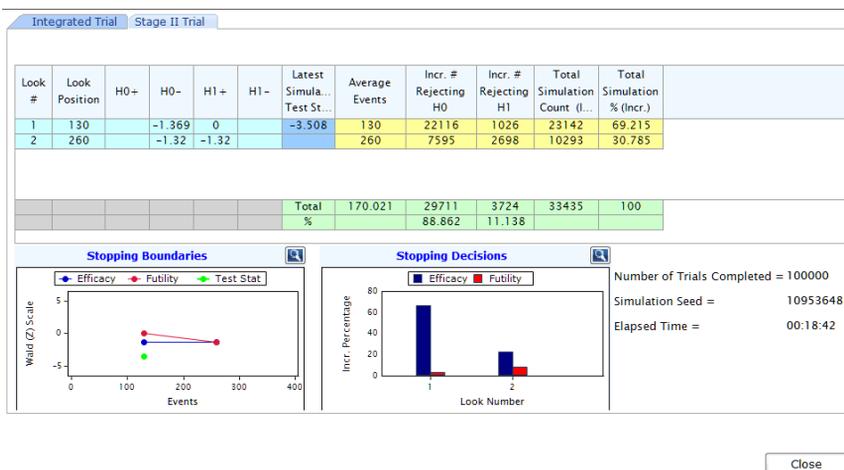
To simulate this design press the **Simulate** button at the bottom right of the screen.

56 Muller and Schafer Method

East will generate 100000 simulated trials with a hazard ratio of 0.72, the value that we entered in the **Response Generation** tab. The simulation results may be viewed in the temporary tables shown below for the Integrated Trial

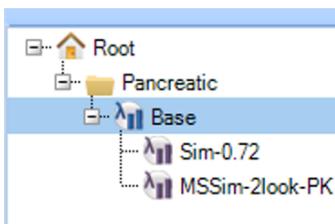


and the Stage II Trial



Press the **Close** to move the simulation results to the Output Preview.

Then press the  button to move the simulation results to the library where you can view the detailed output. Name the saved library node as **MSSim-2look-PK**.



Open MSSim-2look-PK with the tool and examine the output. Notice that the overall power is 79.57% while in the promising zone is 88.86%. The cost in terms of average study duration is 30.304 months for all trials and 30.292 months in the promising zone. The average sample size in the promising zone, however, is 478 compared to 388 for all trials.

Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Rejecting H1 | | Total Simulations | | Average Sample Size | Average Number of Events | Average Accrual Duration | Average Study Duration |
|-------------|--------------------------|----------|--------------------------|----------|-------------------|----------|---------------------|--------------------------|--------------------------|------------------------|
| | Count | Row % | Count | Row % | Count | Column % | | | | |
| Futility | 0 | 0.000% | 3158 | 100.000% | 3158 | 3.158% | 288.01 | 130 | 19.106 | 19.172 |
| Unfavorable | 8515 | 42.499% | 11521 | 57.501% | 20036 | 20.036% | 360 | 260 | 23.934 | 32.059 |
| Promising | 29711 | 88.862% | 3724 | 11.138% | 33435 | 33.435% | 477.711 | 300.021 | 31.711 | 33.031 |
| Favorable | 27730 | 93.194% | 2025 | 6.806% | 29755 | 29.755% | 360 | 260 | 23.934 | 32.217 |
| Efficacy | 13616 | 100.000% | 0 | 0.000% | 13616 | 13.616% | 230.46 | 130 | 19.274 | 19.34 |
| All Trials | 79572 | 79.572% | 20428 | 20.428% | 100000 | 100.000% | 387.615 | 251.575 | 25.747 | 30.292 |

Promising Zone defined as 0.3 <= CP < 0.9

It would be interesting to compare this performance with that of the Base design when the true hazard ratio is 0.72. To make this comparison, change the multipliers for

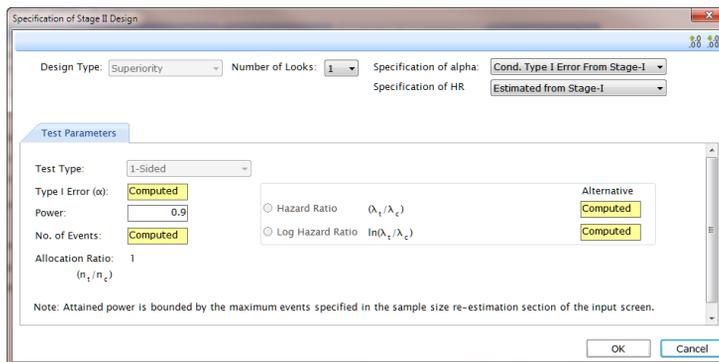
56 Muller and Schafer Method

sample size and events to 1 in the **Sample Size Re-estimation** tab

| | | | |
|---|----------|-------------------------|-----------------------------------|
| Use Adaptation Method | | Specify Stage II Design | |
| <input type="radio"/> CHW <input type="radio"/> CDL <input checked="" type="radio"/> Müller and Schäfer | | | |
| Adapt at: | Look # | 1 | |
| Max. # of Events if Adapt (multiplier; total #): | | 1 | 260 |
| Max. Sample Size if Adapt (multiplier; total #): | | 1 | 360 |
| Upper Limit on Study Duration: | | 98.776 | |
| Promising Zone Scale: | | Cond. Power | <input type="button" value="CP"/> |
| Promising Zone: | Min. CP: | 0.3 | |
| | Max. CP: | 0.9 | |
| CP Computation Based on: | | Estimated HR | |
| Accrual Rate After Adaptation: | | No Change | |
| Estimation Method: | | None | |

Then specify that the Stage II design will be a single-look design with 90% power subject to the cap on events and sample size in the **Sample Size**

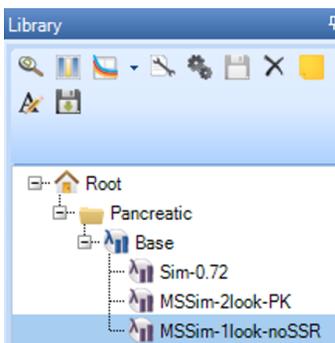
Re-estimation tab.



| | Stage I | Stage II |
|---|--------------------|-------------|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 2 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | Computed |
| Power | 0.901 | 0.9 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Hazard Ratio (Alt.) | 0.667 | Computed |
| Var (Log HR) | Null | Null |
| Boundary Parameters | | |
| Info. Fraction at Interim Look | 0.5 | |
| Efficacy Boundary | LD (OF) | |
| Futility Boundary | HR (NB) | |
| Accrual & Dropout Parameters | | |
| Accrual Rate | 15 | NA |
| Subjects are Followed | Until End of Study | NA |
| No. of Accrual Periods | 1 | NA |
| No. of Dropout Pieces | 0 | NA |
| Sample Size | | |
| Maximum | 360 | Computed |
| Events | | |
| Maximum | 260 | Computed |
| Study Duration | | |
| Maximum | 32.925 | NA |
| Expected Under H0 | 23.853 | NA |
| Expected Under H1 | 29.361 | NA |
| Accrual Duration | | |
| Maximum | 24 | NA |
| Expected Under H0 | 21.058 | NA |
| Expected Under H1 | 22.819 | NA |

56 Muller and Schafer Method

Now press the **Simulate** button. Save the new design in the library with the name **MSSim-1look-noSSR**



Examine the simulation details of MSSim-1look-noSSR

Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Rejecting H1 | | Total Simulations | | Average Sample Size | Average Number of Events | Average Accrual Duration | Average Study Duration |
|-------------|--------------------------|----------|--------------------------|----------|-------------------|----------|---------------------|--------------------------|--------------------------|------------------------|
| | Count | Row % | Count | Row % | Count | Column % | | | | |
| Futility | 0 | 0.000% | 3119 | 100.000% | 3119 | 3.119% | 288.217 | 130 | 19.1 | 19.167 |
| Unfavorable | 8206 | 41.390% | 11620 | 58.610% | 19826 | 19.826% | 360 | 260 | 23.934 | 32.059 |
| Promising | 24945 | 74.576% | 0 | 0.000% | 33449 | 33.449% | 360 | 260 | 23.934 | 32.139 |
| Favorable | 27968 | 93.165% | 2052 | 6.835% | 30020 | 30.020% | 360 | 260 | 23.933 | 32.213 |
| Efficacy | 13586 | 100.000% | 0 | 0.000% | 13586 | 13.586% | 290.441 | 130 | 19.269 | 19.335 |
| All Trials | 74705 | 74.705% | 16791 | 16.791% | 100000 | 100.000% | 348.311 | 238.284 | 23.149 | 30.001 |

Promising Zone defined as $0.3 \leq CP < 0.9$

Notice that the power in the Promising Zone is only 74.71%. And the average cost in terms of study duration 32.139 in the promising zone and 30.001 months for all trials, about the same as for MSSim-2look-PK. On the other hand the average sample size in the promising zone is only 360 subjects for MSSim-1look-noSSR, compared to 478 subjects for MSSim-2look-PK. This is the cost associated with increasing the power from 74.58% to 88.86% and it is only incurred if the interim results are promising.

It would be interesting to compare the adaptive design obtained by the Müller and Schäfer method with the adaptive design obtained by the CHW method. The main limitation of the CHW method is that the only adaptation permitted is an increase in events and sample size. There is no flexibility to alter the number or spacing of the future looks after the adaptation. To run the CHW design make the changes shown below in the **Sample Size Re-estimation** tab. (Notice that that **Target CP for Re-estimating # of Events** is set to 0.999 so as to ensure a one-time

56 Muller and Schafer Method

therefore expect that if we were to construct a Müller and Schäfer design with only one look for the Stage II portion, it would have similar operating characteristics to the CHW design. To verify this conjecture create such a design by using the following inputs.

Use Adaptation Method
 CHW CDL Müller and Schäfer

Adapt at:

Max. # of Events if Adapt (multiplier, total #):

Max. Sample Size if Adapt (multiplier, total #):

Upper Limit on Study Duration:

Promising Zone Scale:
 Min. CP:
 Max. CP:

CP Computation Based on:

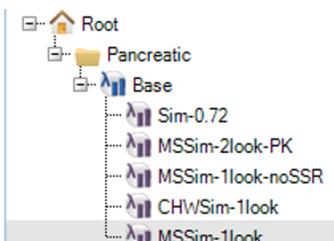
Accrual Rate After Adaptation:

Estimation Method:

Specify Stage II Design

| | Stage I | Stage II |
|---|--------------------|-------------|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 2 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | Computed |
| Power | 0.901 | 0.999 |
| Model Parameters | | |
| Allocation Ratio (nt, nc) | 1 | 1 |
| Hazard Ratio (Alt.) | 0.667 | Computed |
| Var (Log HR) | Null | Null |
| Boundary Parameters | | |
| Info. Fraction at Interim Look | 0.5 | |
| Efficacy Boundary | LD (PF) | |
| Futility Boundary | HR (NB) | |
| Accrual & Dropout Parameters | | |
| Accrual Rate | 15 | NA |
| Subjects are Followed | Until End of Study | NA |
| No. of Accrual Periods | 1 | NA |

Simulate the design and save it in the library with the name MSSim-1look.



Zone-wise Averages

| Zone | Simulations Rejecting H0 | | Simulations Rejecting H1 | | Total Simulations | | Average Sample Size | Average Number of Events | Average Accrual Duration | Average Study Duration |
|--------------------------|--------------------------|---------|--------------------------|---------|-------------------|----------|---------------------|--------------------------|--------------------------|------------------------|
| | Count | Row % | Count | Row % | Count | Column % | | | | |
| ⊕ Futility + Unfavorable | 8376 | 36.285% | 14708 | 63.715% | 23084 | 23.084% | 349,911 | 241,765 | 23.257 | 30.255 |
| ⊖ Promising | 30790 | 91.964% | 0 | 0.000% | 33484 | 33.484% | 540 | 390 | 35.867 | 40.073 |
| ⊕ Efficacy + Favorable | 41425 | 95.379% | 2007 | 4.621% | 43432 | 43.432% | 338,161 | 219,218 | 22.468 | 28.178 |
| All Trials | 80591 | 80.691% | 16715 | 16.715% | 100000 | 100.000% | 408,457 | 281,607 | 27.137 | 32.64 |

Promising Zone defined as 0.3 <= CP < 0.9

As we anticipated, the CHWSim-1look, the MSSim-1look designs have similar operating characteristics; about the same power and same study duration in all zones. This confirms the claim by Mehta and Liu (2016) that for the special case of a single future look following an interim analysis the CHW and Müller and Schäfer methods are equivalent. These examples show that the Müller and Schäfer method has greater flexibility for trading off power versus study duration in an adaptive setting than the CHW method; it permits more complex adaptations for the Stage II design, without sacrificing power or study duration in the special case of a single-look adaptation.

56.4.4 Interim Monitoring

We will now use East to monitor the Base design. With cursor on **Base** in the library window, click on the interim monitoring icon **IM**. You will be taken to the interim monitoring worksheet.

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | Futility | Rep. Conf. Bounds for HR | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|-------------------|----------------|------------------|--------------------------------|----------|----------|--------------------------|--------------|------------------|----|------------------|
| | | | | | | | | Upper (1...) | Lower (2...) | | | |
| 1 | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | |

The First Interim Look In order to populate this worksheet with interim data you must click on the **Enter Interim Data** button on the tool bar at the top of the worksheet. Thereby a form, titled **Test Statistic Calculator**, appears and you are requested to enter the interim number of events, the interim estimate of δ and its standard error into this form.

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Read from Analysis Node

Select Workbook: [Dropdown]

Select Analysis Node: [Dropdown]

Cumulative Events: [Input: 130]

Input for Survival end point

Estimate of δ : [Input: -0.405]

$\delta = \ln(\lambda_t / \lambda_c)$

Standard Error of Estimate of δ : [Input: 0.1754]

Output

Test Statistic: [Output: -2.3087]

Recalc OK Cancel

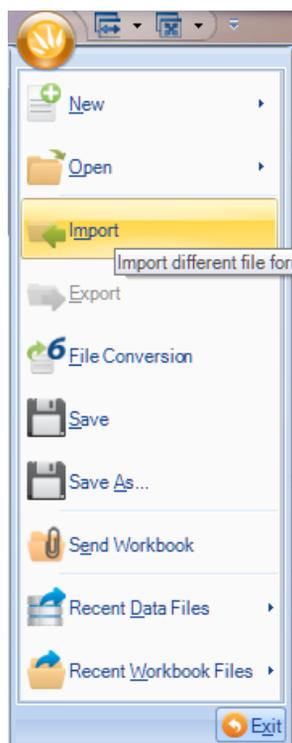
You now have two options.

56 Muller and Schafer Method

Option 1 Enter the requested quantities directly into the **Test Statistic Calculator** for the current look and click the **OK** button at the bottom. East will then perform the necessary calculations and post the interim results in the worksheet as shown below.

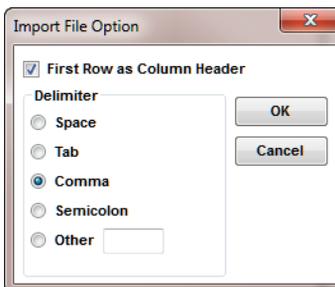
| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of... | Est. of δ | Std. Error of Est. of δ | Efficacy | Futility | Rep. Conf. Bounds for HR | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|-------------------|----------------|------------|------------------|--------------------------------|----------|----------|--------------------------|--------------|------------------|--------|------------------|
| | | | | | | | | | Upper (1...) | Lower (2...) | | | |
| 1 2 | 0.5 | 130 | -2.3087 | 0.667 | -0.405 | 0.1754 | -2.9626 | 0 | 1.1215 | 0.4449 | 0.0703 | 0.9666 | 0.9026 |

Option 2 If the actual patient level data are saved as a file in one of the acceptable file formats, you can import the file into East through the **File > Import**.



In this example there is a file titled Pancreatic-Look1.csv in the sub-folder **Samples** in your East installation folder containing the data required for the interim analysis at look 1. When you select this file with **File > Import > ~Pancreatic-Look1.csv** you will be asked to select the appropriate

Delimiter from the **Import File Format Option** dialog box.



Choose the default **Comma** delimiter and attach the file to the **Pancreatic** workbook. East will then display the data in its **Data Editor**.

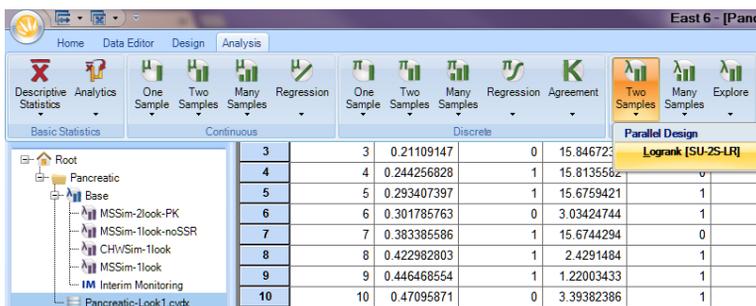
| | SubjectID | ArrivalTime | TreatmentID | TimeOnStud | CensorInd | Status |
|----|-----------|-------------|-------------|-------------|-----------|--------|
| 1 | 1 | 0.091101201 | 0 | 15.9667138 | 0 | 0 |
| 2 | 2 | 0.136753868 | 0 | 0.736892474 | 1 | 1 |
| 3 | 3 | 0.21109147 | 0 | 15.8467236 | 0 | 0 |
| 4 | 4 | 0.244256828 | 1 | 15.8135582 | 0 | 0 |
| 5 | 5 | 0.293407397 | 1 | 15.6759421 | 1 | 1 |
| 6 | 6 | 0.301785763 | 0 | 3.03424744 | 1 | 1 |
| 7 | 7 | 0.383385586 | 1 | 15.6744294 | 0 | 0 |
| 8 | 8 | 0.422982803 | 1 | 2.4291484 | 1 | 1 |
| 9 | 9 | 0.446468554 | 1 | 1.22003433 | 1 | 1 |
| 10 | 10 | 0.47095871 | 0 | 3.39382386 | 1 | 1 |
| 11 | 11 | 0.632641096 | 0 | 11.9518103 | 1 | 1 |

This file must contain, at a minimum, the above six variables, **Subject ID**, **ArrivalTime**, **TreatmentID**, **TimeOnStudy**, **CensorInd**, **Status**, although it may contain many other variables as well. The names of these variables may differ in your data set from the ones given in this example, but they must carry the same meaning above. The variable names used in this example are mostly self-explanatory. However there is a distinction between **CensorInd** and **Status**. **CensorInd** assumes the value 1 if the event (in this case death) has occurred, and assumes the value 0 if the observation is administratively censored while the subject is still in follow-up. **Status** assumes the value 1 if the event (in this case death) has occurred, assumes the value 0 if the observation is administratively censored, and assumes the value -1 if the subject has dropped out of the study. In this example **CensorInd** and **Status** are the same, because there are no drop-outs.

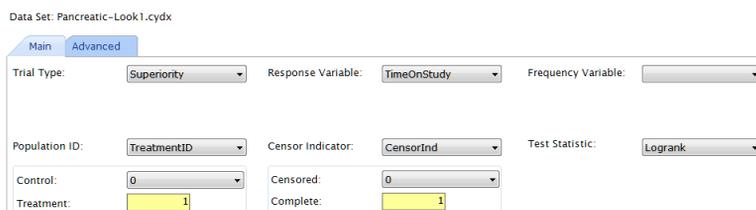
Before East can populate the interim analysis worksheet it is necessary to create an **Analysis Node** from this data set. Accordingly select the

56 Muller and Schafer Method

Analysis>Two Sample>Logrank options from the top-level menu



and complete the entries in the ensuing form as shown.



Upon clicking the OK button East will create the following **Analysis of Time to Event Response** node.

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | | Average Follow-up Time |
|--------------|-----------------|------------|----------------|------------|----------------|------------------------|
| | | Count | % | Count | % | |
| 0 | 133 | 69 | 51.880% | 64 | 48.120% | 5.1962 |
| 1 | 130 | 61 | 46.923% | 69 | 53.077% | 6.2139 |
| Total | 263 | 130 | 49.430% | 133 | 50.570% | 5.6992 |

Test of Hypothesis:

| Log Rank Score | Std. Error | Standardized Test Statistic | (1-Sided) | | (2-Sided) |
|----------------|------------|-----------------------------|-----------|---------|-----------|
| | | | Tail | p-value | p-value |
| -8.9569 | 5.672 | -1.5792 | L.E. | 0.0572 | 0.1143 |

Parameter Estimates from Cox Model:

| Hazard Ratio (HR) | ln(Hazard Ratio) (δ) | Std. Error of Estimate of δ | Wald Statistic | 95% Confidence Interval of HR(2-Sided) | |
|-------------------|-------------------------------|------------------------------------|----------------|--|-------------|
| | | | | Lower Limit | Upper Limit |
| 0.7579 | -0.2772 | 0.1761 | -1.5741 | 0.5366 | 1.0703 |

Estimated Hazard Rates:

| | |
|--------------------------------|--------|
| Control (λ_c) | 0.0998 |
| Treatment ($\lambda_c * HR$) | 0.0757 |

Four tables are created. The **Summary of Observed Data** table shows

that 130 events have been observed from 263 enrolled subjects. The **Parameter Estimates from Cox Model** table displays the current estimate of hazard ratio (HR=0.7579), and other related output from the Cox model including the estimate of δ (-0.2772), its standard error (0.1761), and the corresponding Wald statistic (-1.5741). This information can now be utilized to populate the interim analysis worksheet.

Select the interim monitoring node from the library and click the edit button from the library menu to retrieve the worksheet.

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | Futility | Rep. Conf. Bounds for HR | | Repeated p-value | CP | Predictive Power | |
|--------|----------------------|-------------------|----------------|------------------|--------------------------------|----------|----------|--------------------------|--------------|------------------|--------|------------------|--------|
| | | | | | | | | Upper (1...) | Lower (2...) | | | | |
| 1 | 0.5 | 130 | -2.3087 | 0.667 | -0.405 | 0.1754 | -2.9626 | 0 | 1.1215 | 0.4449 | 0.0703 | 0.9666 | 0.9028 |
| 2 | | | | | | | | | | | | | |

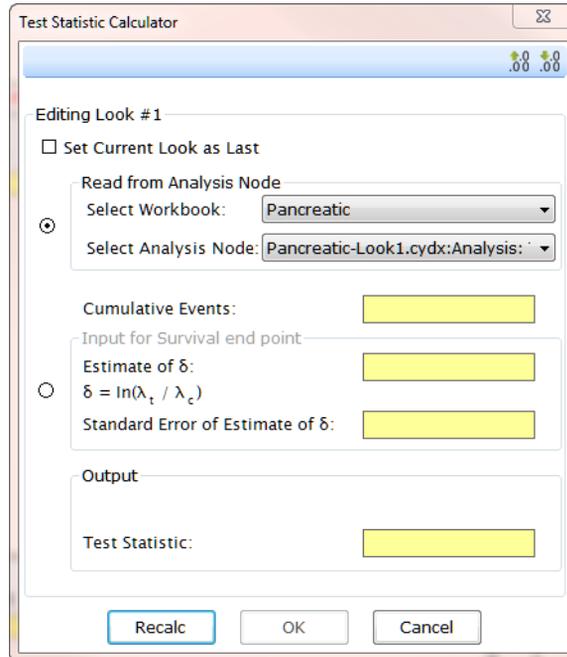
If any row of the worksheet is already populated clear away the entries by selecting that row and clicking on the **Delete Look** button

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | Futility | Rep. Conf. Bounds for HR | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|-------------------|----------------|------------------|--------------------------------|----------|----------|--------------------------|--------------|------------------|----|------------------|
| | | | | | | | | Upper (1...) | Lower (2...) | | | |
| 1 | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | |

Now click on the **Enter Interim Data** button and select the **Read from**

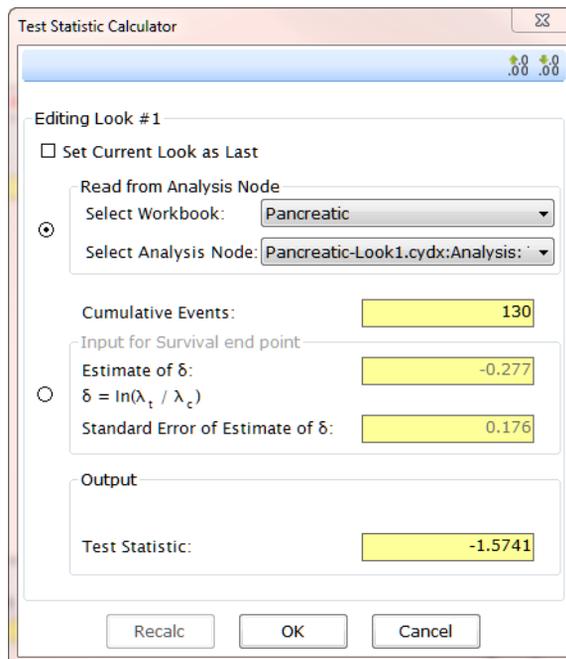
56 Muller and Schafer Method

Analysis Node radio button.



Make sure that the appropriate **Workbook** and **Analysis Node** are selected

in this dialog box. Then click the **Recalc** button.



Upon clicking **OK Look 1** of the **Interim Analysis Worksheet** gets populated with the results of the first interim analysis.

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of... | Est. of δ | Std. Error of Est. of δ | Efficacy | Futility | Rep. Conf. Bounds for HR | | Repeated p-value | CP | Predictive Power |
|--------|----------------------|-------------------|----------------|------------|------------------|--------------------------------|----------|----------|--------------------------|-------------|------------------|--------|------------------|
| | | | | | | | | | Upper (1... | Lower (2... | | | |
| 1 | 0.5 | 130 | -1.5741 | 0.758 | -0.2772 | 0.1761 | -2.9626 | 0 | 1.277 | 0.5055 | 0.1796 | 0.6421 | 0.6016 |
| 2 | | | | | | | | | | | | | |

We observe that the test statistic, -1.5741 has not crossed the efficacy boundary, -2.9626. However, the conditional power, 0.6421, is in the promising zone. Thus an adaptive increase in number of events and sample size is indicated. This can be confirmed by simulation as we show next.

The Predictive Interval Plot (Note: The Predictive Interval Plots (PIPs) are introduced and fully described with examples in Chapter 65.) Clicking on the **PIP** button on the menu bar will enable you to simulate the future course of the trial conditional on the current data. You will be requested in fill in the names of required

56 Muller and Schafer Method

variables from the **Pancreatic-Look1** data set.

Input for Predicted Intervals Plot

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

PIP for Look: Final Look Required No. of Events: 260 Total Sample Size: 360

Specify Subject Info

Select Workbook: Pancreatic

Select Subject Data: Pancreatic-Look1.cydx

Choose Variables

Population ID: Status Indicator: 1=Complete 0=Censored -1=Dropout

Control: Treatment:

Arrival Time: Response Variable:

Optional: Estimate Parameters from Data

Parameters of Current Dataset (Editable)

Sample Size: No. of Events: Accrual Rate: 15 Hazard Ratio (λ_1/λ_2): 0.667

Hazard Rate (λ): Control: 0.082 Treatment: 0.054

Number of Simulations: 1000 Refresh Frequency: 100

Random Number Seed: Clock Fixed: 100

Suppress Intermed. Output Pause after Refresh

Simulate Cancel

Select the appropriate variables from the respective drop-down menus as shown.

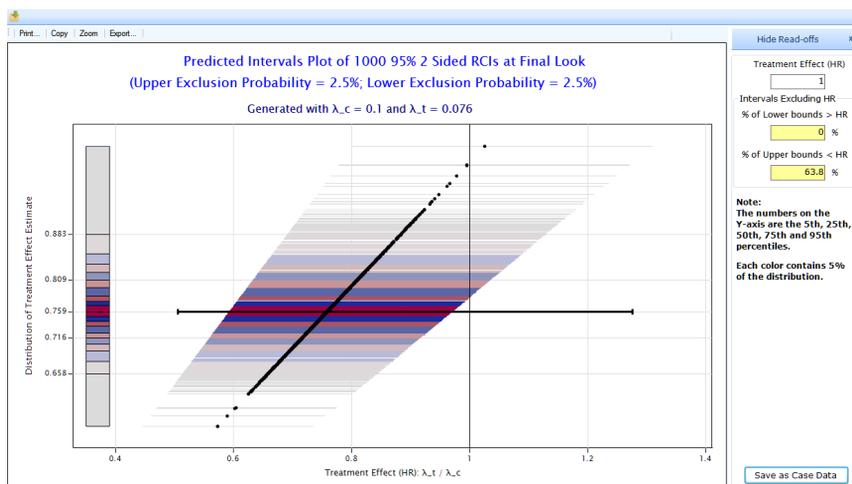
Click on the **Optional: Estimate Parameters from Data** button so that East can estimate the individual hazards and hazard ratio and can input the sample size and number of events

from the **Pancreatic-Look1** data set.

If you now click on the **Simulate** button you will simulate the future course of the trial from the current data in **Pancreatic-Look1** and obtain 1000 repeated confidence intervals (RCIs) each representing a possible final analysis for the trial. These RCIs are sorted and stacked on top of one another to provide an intuitive plot called a Predicted Interval Plot (see, for example, Li, Evans, Uno and Wei, Statistics in

56 Muller and Schafer Method

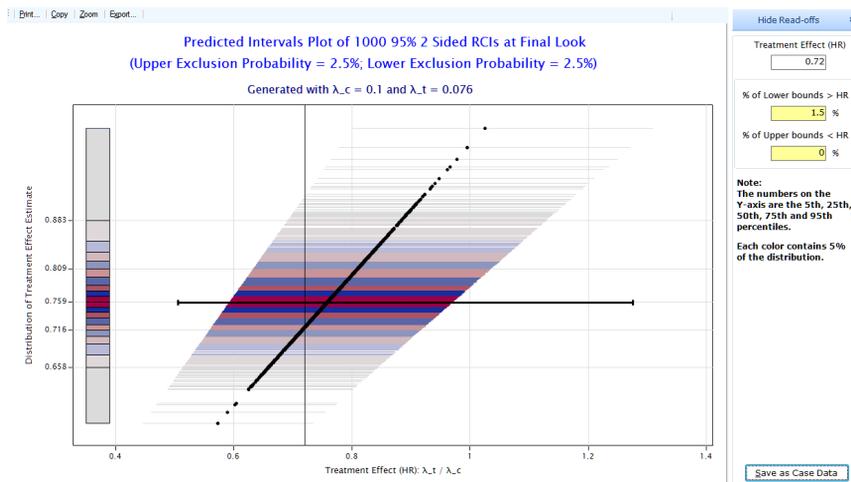
Biopharmaceutical Research, 2009).



The black dot at the center of each RCI is the estimate of hazard ratio for that simulation. The X-axis displays a range of possible hazard ratios with a vertical cursor positioned by default at HR=1. The vertical cursor can be dragged to the left or right or be moved to a specific location by entering a value in the **Treatment Effect** edit box at the top right of the window.

It is seen that 63.8% of these RCI's have their upper bounds to the left of HR=1, thereby demonstrating the estimated conditional power of 64.2%. The color coded vertical bar on the right of the graph is a heat plot representing the distribution of the 1000 hazard ratios. Each color represents 5% of the observed hazard ratios. For example, the lowest 5% of hazard ratios have values less than or equal to 0.658, the lowest 25% of hazard ratios have values less than 0.716, and so on. The PIP plot is more informative than a conditional power calculation. To see this let us suppose that only hazard ratios that are smaller than 0.72 are considered to be clinically meaningful. If we move the vertical cursor to 0.72, we find that only 1.5% of the 1000 simulated future

trial have a clinically meaningful hazard ratio.



To save this PIP plot in the library for future use click on the **Save in Workbook** icon at the top of the window

A snapshot of the current entries in the interim monitoring worksheet is saved in the library along with the PIP plot.

One can examine the contents of these newly created nodes by double-clicking or by selecting and the clicking on the **Details** tool in the library tool bar.

Adaptive Increase in Events and Number of Looks Since the conditional power (64%) is in the promising zone we decide to make an adaptive change. In keeping with the simulations that were performed at the design stage, let us alter the future course of the trial in two ways:

1. Increase the total number of events by 50%. Thus the total number of events will be increased from the 260 to 390.
2. Increase the number for future looks from 1 to 2 and alter the efficacy spending function from LD(OF) to LD(PK).

In order to make these changes click on the **Adapt** button on the tool bar at the top of

56 Muller and Schafer Method

the interim monitoring worksheet.

Specification of Stage II Design

Design Type: Superiority Number of Looks: 1

Test Parameters Accrual / Dropouts

Test Type: 1-Sided # of Hazard Pieces: 1 Input Method: Median Survival Times

Conditional Error Rate: 0.1132 Hazard Ratio (Optional)

Power: 0.9 Hazard Ratio (λ_1/λ_2) Alternative: 0.75788

Incremental Ratio of Medians (m_1/m_2) Alternative: 1.31947

No. of Events: Computed

Cumulative Med. Surv. Time

No. of Events: Computed

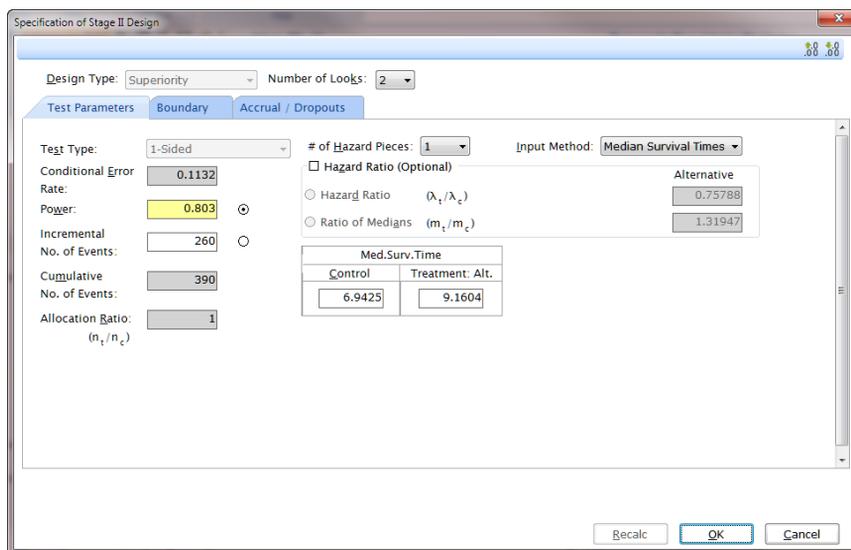
| | Control | Treatment: Alt. |
|-----------------|---------|-----------------|
| Med. Surv. Time | 6.9425 | 9.1604 |

Allocation Ratio: 1 (n_1/n_2)

Recalc OK Cancel

Change the **Number of Looks** to 2 and the **Incremental Number of Events** to 260 as shown. (Note: to change the events rather than the power, you will

have to switch the selection of the radio button.). Press the **Recalc** button when done.

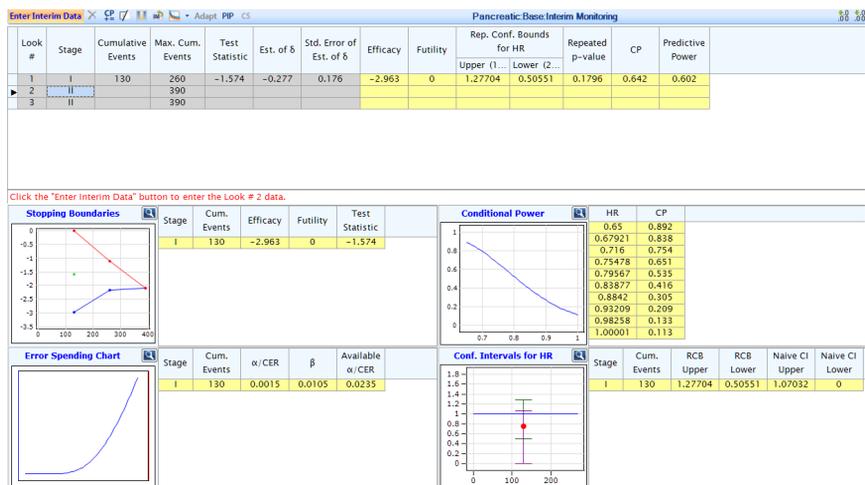


Next go to the **Boundary** tab and change the error spending function for efficacy from LD(OF) to LD(PK). Also change the HR for early stopping for futility to 1.0. Press **Recalc** when done.

Finally examine the **Accrual/Dropouts** tab and leave these entries unchanged and

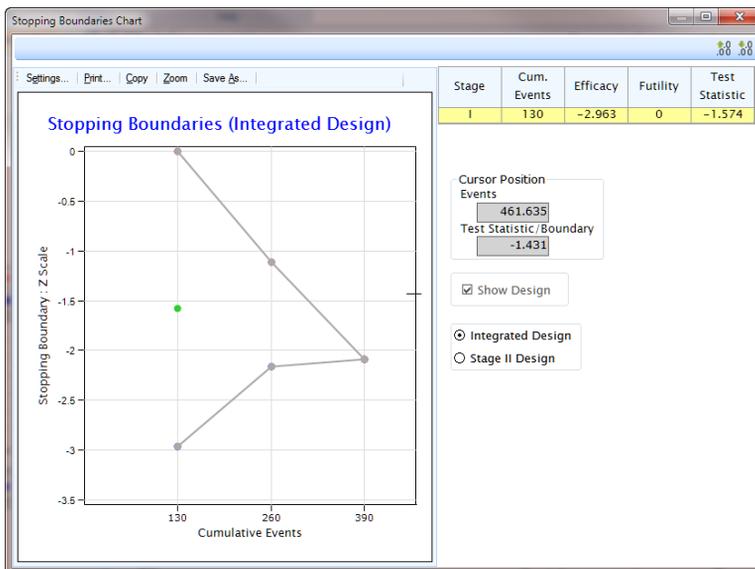
56 Muller and Schafer Method

click **OK**.



With these adaptations, the study will enroll an additional 209 subjects for a total of 472 subjects, and will follow them until 390 total events are obtained. Two additional interim looks are planned, one at 260 events and one at 390, with a Pocock spending function efficacy boundary and a HR=1 futility boundary. The interim monitoring worksheet has been modified to reflect these changes. Below is a screenshot of the

integrated 3-look adaptive design.



The Second Interim Look Import the data set for the second interim analysis into East with the **File>Import>~Pancreatic-Look2.csv** commands.

| SubjectID: 1 Value: 1 | | SubjectID | ArrivalTime | TreatmentID | TimeOnStud | CensorInd | Status |
|-----------------------|----|-----------|-------------|-------------|-------------|-----------|--------|
| 1 | 1 | 1 | 0.091101201 | 0 | 24.182712 | 1 | 1 |
| 2 | 2 | 2 | 0.136753868 | 0 | 0.736892474 | 1 | 1 |
| 3 | 3 | 3 | 0.21109147 | 0 | 27.4831151 | 0 | 0 |
| 4 | 4 | 4 | 0.244256828 | 1 | 27.4499497 | 0 | 0 |
| 5 | 5 | 5 | 0.293407397 | 1 | 15.6759421 | 1 | 1 |
| 6 | 6 | 6 | 0.301785763 | 0 | 3.03424744 | 1 | 1 |
| 7 | 7 | 7 | 0.383385586 | 1 | 24.9278675 | 1 | 1 |
| 8 | 8 | 8 | 0.422982803 | 1 | 2.4291484 | 1 | 1 |
| 9 | 9 | 9 | 0.446468554 | 1 | 1.22003433 | 1 | 1 |
| 10 | 10 | 10 | 0.47095871 | 0 | 3.39382386 | 1 | 1 |
| 11 | 11 | 11 | 0.632641096 | 0 | 11.9518103 | 1 | 1 |
| 12 | 12 | 12 | 0.716651754 | 1 | 3.6106417 | 1 | 1 |
| 13 | 13 | 13 | 0.726867964 | 1 | 4.41061828 | 1 | 1 |
| 14 | 14 | 14 | 0.733233724 | 0 | 4.83555969 | 1 | 1 |

Create an analysis node for the **Pancreatic-Look2** data set in the same manner as

56 Muller and Schafer Method

was done for **Pancreatic-Look1**.

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | | Average Follow-up Time |
|--------------|-----------------|--------|----------|----------|----------|------------------------|
| | | Count | % | Count | % | |
| 0 | 204 | 140 | 68.6275% | 64 | 31.3725% | 6.92128 |
| 1 | 210 | 120 | 57.1429% | 90 | 42.8571% | 8.22528 |
| Total | 414 | 260 | 62.8019% | 154 | 37.1981% | 7.58273 |

Test of Hypothesis:

| Log Rank Score | Std. Error | Standardized Test Statistic | (1-Sided) | | (2-Sided) | |
|----------------|------------|-----------------------------|-----------|---------|-----------|--|
| | | | Tail | p-value | p-value | |
| -21.859 | 7.989 | -2.736 | L.E. | 0.0031 | 0.0062 | |

Parameter Estimates from Cox Model:

| Hazard Ratio (HR) | ln(Hazard Ratio) (δ) | Std. Error of Estimate of δ | Wald Statistic | 95% Confidence Interval of HR(2-Sided) | |
|-------------------|-------------------------------|------------------------------------|----------------|--|-------------|
| | | | | Lower Limit | Upper Limit |
| 0.71133 | -0.341 | 0.125 | -2.723 | 0.55669 | 0.90893 |

Estimated Hazard Rates:

| | |
|-------------------------------|---------|
| Control (λ_0) | 0.09915 |
| Treatment (λ_1 * HR) | 0.07053 |

Return to the interim monitoring worksheet by clicking on followed by in the library. Select Row 1 of the interim monitoring worksheet.

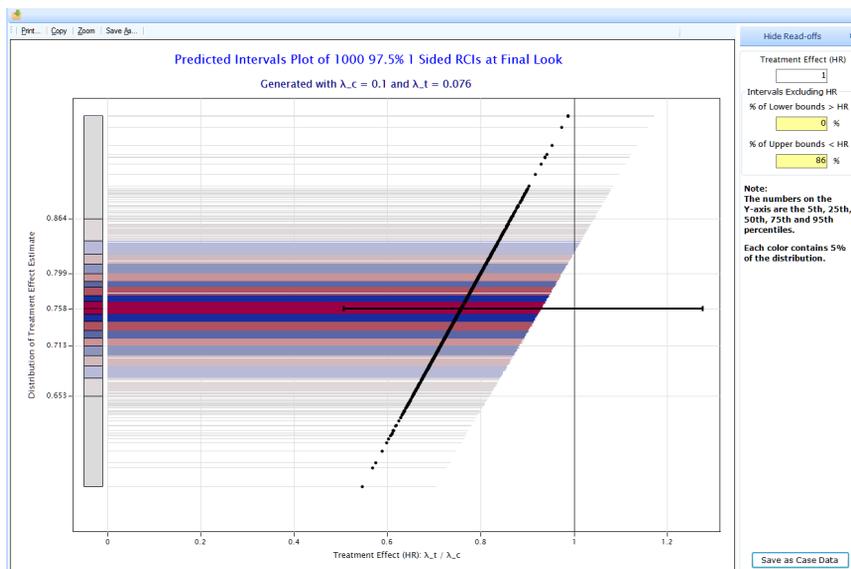
Click on the **PIP** button and complete the entries as shown below to simulate the future course of this adaptive trial. As before, you will select the **Pancreatic-Look1.cydx** data set from the **Select Subject Data** drop

down box.

Click **Simulate** and obtain 1000 **one-sided** repeated confidence intervals adjusted for the adaptive design by the published method of Mehta, Bauer, Posch and Brannath

56 Muller and Schafer Method

(2008).



With the adaptive change of total number of events, the conditional power has improved considerably and is now 86.0%.

Return to the interim monitoring worksheet once more and select Row 2.

| Enter Interim Data | | | | | | | | | | Pancreatic-Basis Interim Monitoring | | | |
|--------------------|-------|-------------------|------------------|----------------|------------------|--------------------------------|----------|----------|--------------------------|-------------------------------------|------------------|-------|------------------|
| Look # | Stage | Cumulative Events | Max. Cum. Events | Test Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | Futility | Rep. Conf. Bounds for HR | | Repeated p-value | CP | Predictive Power |
| | | | | | | | | | Upper (1... | Lower (2... | | | |
| 1 | I | 130 | 260 | -1.574 | -0.277 | 0.176 | -2.963 | 0 | 1.27704 | 0.50551 | 0.1796 | 0.642 | 0.602 |
| 2 | II | | 390 | | | | | | | | | | |
| 3 | II | | 390 | | | | | | | | | | |

Read the Pancreatic-Look2 data into East by clicking on

and then selecting **Pancreatic-Look2.cydx** for the **Select Analysis**

Node drop down box.

Test Statistic Calculator

Editing Look #2 of Integrated Trial

Set Current Look as Last

Read from Analysis Node

Select Workbook: Pancreatic

Select Analysis Node: Pancreatic-Look2.cydx:Analysis:

Cumulative Events: []

Input for Survival end point

Estimate of δ : []

$\delta = \ln(\lambda_t / \lambda_c)$

Standard Error of Estimate of δ : []

Output

Test Statistic: []

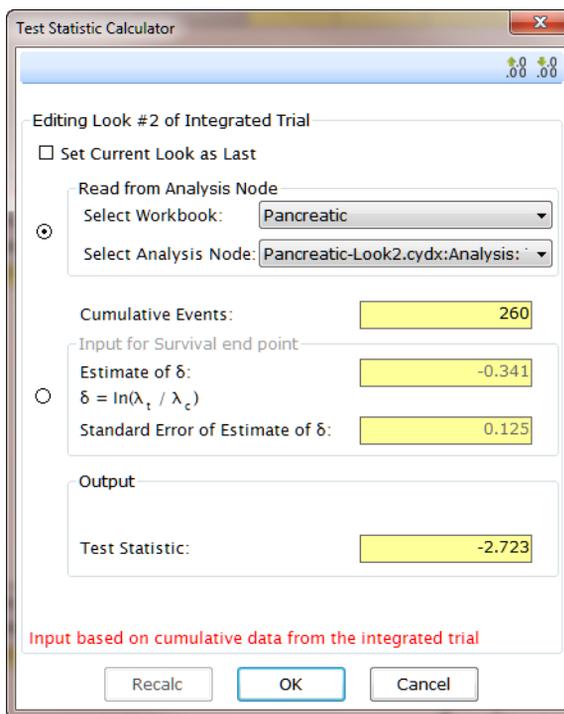
Input based on cumulative data from the integrated trial

Recalc OK Cancel

Click on the **Recalc** button to complete the entries in the **Test Statistic**

56 Muller and Schafer Method

Calculator.



Finally, click on the **OK** button. East tells us that the efficacy boundary has been crossed.

| Look # | Stage | Cumulative Events | Max. Cum. Events | Test Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | Futility | Rep. Conf. Bounds for HR | | Repeated p-value | CP | Predictive Power |
|--------|-------|-------------------|------------------|----------------|------------------|--------------------------------|----------|----------|--------------------------|--------------|------------------|-------|------------------|
| | | | | | | | | | Upper (1...) | Lower (2...) | | | |
| 1 | I | 130 | 260 | -1.574 | -0.277 | 0.176 | -2.963 | 0 | 1.27704 | 0.50551 | 0.1796 | 0.642 | 0.602 |
| 2 | II | 260 | 390 | -2.723 | -0.341 | 0.125 | -2.156 | -1.113 | 0.92843 | 0 | 0.0048 | 0.985 | 0.962 |

Boundary Crossed

Since the value of Test Statistic is \leq the critical point for efficacy, H_0 is rejected.

Although boundary has been crossed, East gives you choice either to stop the study or to continue entering further looks. Please make your decision.

Stop stop the study and bar further looks input

Continue allow the study to continue

| Final Inference | |
|---------------------------|---------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.0043 |
| Adj. Ft. Est. for HR | 0.71526 |
| 95% of BWCI for HR | |

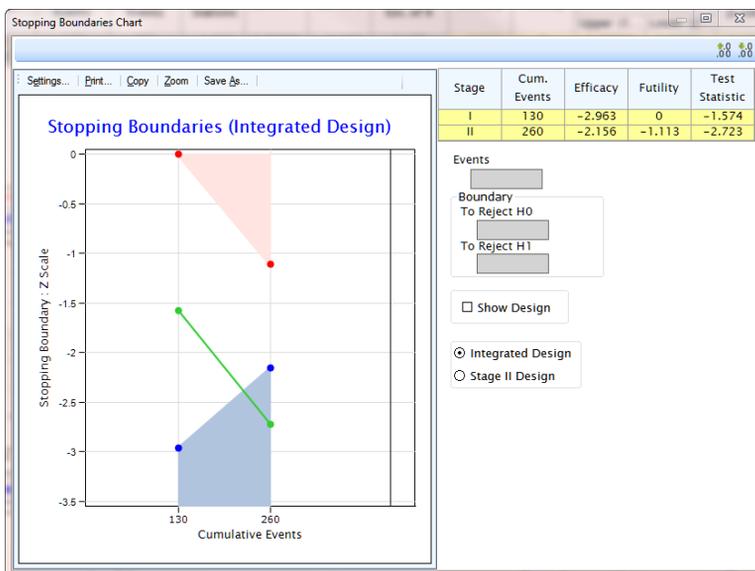
Click on the **Stop** button to complete the trial. The final inference is displayed in the

following table.

| Final Inference | |
|--|---------|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.0043 |
| Adj. Pt. Est. for HR | 0.71526 |
| 95% of BWCI for HR | |
| Upper Confidence Bound | 0.9168 |
| Lower Confidence Bound | 0.5598 |
| Adj. Pt. Est. for δ | -0.3351 |
| 95% of BWCI for δ | |
| Upper Confidence Bound | -0.0869 |
| Lower Confidence Bound | -0.5802 |
| Post-Hoc Power | |

Statistical significance has been achieved. The stage wise adjust p-value after accounting for the adaptation is 0.0043. The 95% confidence interval for hazard ratio is (0.5598, 0.9168) and the point estimate is 0.7153. Examine the various charts on the interim monitoring worksheet.

Chart: Stopping Boundaries (Integrated Design)



56 Muller and Schafer Method

Chart: Confidence Intervals for HR

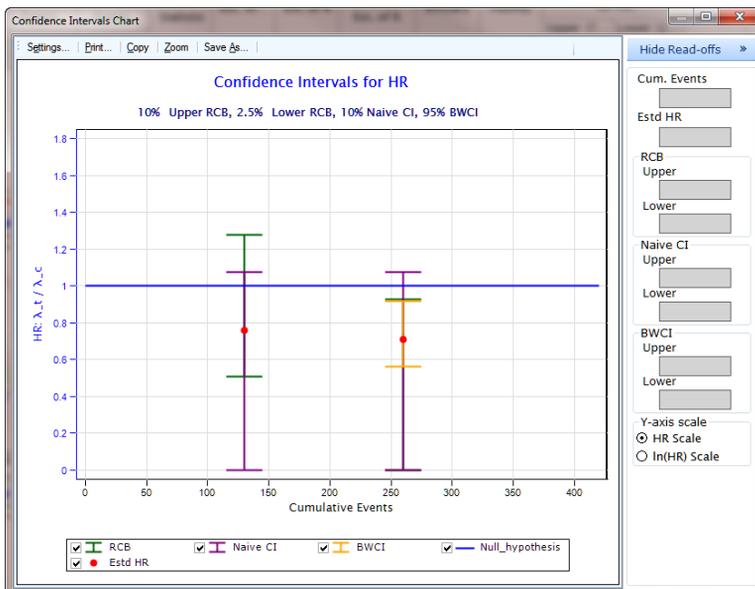


Chart: Error Spending Function (Stage II Design)



57 *Conditional Power for Decision Making*

In the course of conducting an adaptive clinical trial, many decisions that are to be made on matters such as determining sample size, stopping the trial for futility, and whether or not and when and how to adapt the trial design, depend primarily on the values of ‘power estimates’. East provides facilities to compute or use different types of ‘power estimates’ while designing, simulating, or monitoring a trial.

This chapter describes the special conditional power calculators that EastAdapt and EastSurvAdapt have provided for computing conditional power either in the interim monitoring worksheets or in the simulation worksheets. Informally, conditional power is the probability, given the current data, that the trial will ultimately achieve statistical significance. For a more formal definition refer to Chapter 54, Section 54.1.3. Conditional power calculations depend on assumptions that you make about the unknown parameters δ and σ . The conditional power calculators in East accept as inputs either user-specified values of δ and σ , or estimates of δ and σ obtained at the interim analysis. This will be illustrated through several examples in this chapter.

This chapter is arranged into the following sections:

- CP Calculator-CHW:Interim Monitoring
 - Normal Endpoint
 - Binomial Endpoint
 - Time to Event Endpoint
- CP Calculator-CHW:Simulation
 - Normal Endpoint
 - Binomial Endpoint
 - Time to Event Endpoint

57.1 CP Calculator - CHW: Interim Monitoring

57.1.1 Normal Endpoint

57.1.2 Binomial Endpoint

57.1.3 Time to Event Endpoint

This section explains the use of conditional power calculator while performing Interim Monitoring.

57.1.1 Normal Endpoint

Consider a two-arm trial to determine if there is an efficacy gain for an experimental drug relative to the industry standard treatment for negative symptoms schizophrenia. The primary endpoint is the improvement from baseline to week 26 in the Negative Symptoms Assessment (NSA), a 16-item clinician-rated instrument for measuring the negative symptomatology of schizophrenia. The trial is designed for one-sided

alternative hypothesis that $\delta > 0$. It is expected, from limited data on related studies, that the difference of mean is expected to be 10 with a standard deviation of 50.

Create a design worksheet as shown below.

| CPower1:Des1 | |
|---|---------------------|
| Mnemonic | MN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 5 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Input Method | Difference of Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 10 |
| Std. Deviation (σ) | 50 |
| Test Statistic | Z |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Sample Size | |
| Maximum | 1075 |
| Expected Under H0 | 1071.471 |
| Expected Under H1 | 797.168 |

We will now monitor the trial. Invoke the CHW interim monitoring by clicking  icon which will appear as displayed below.

| CHW Interim Monitoring: Des1 | | | | | | | | | | | |
|------------------------------|-------------------------|-----------------------|------------------------|----------------------|----------------------|----------|----------------|----------|------------------------|-------|--------------------|
| Look # | Incremental Sample Size | Incremental Statistic | Cumulative Sample Size | Prespecified Weights | Weight ... Statistic | δ | Standard Error | Efficacy | 97.5% RCI for δ | | Repeat ... p-value |
| | | | | | | | | | Upper | Lower | |
| 1 | | | | 0.2 | | | | 4.877 | | | |
| 2 | | | | 0.2 | | | | 3.357 | | | |
| 3 | | | | 0.2 | | | | 2.68 | | | |
| 4 | | | | 0.2 | | | | 2.29 | | | |
| 5 | | | | 0.2 | | | | 2.031 | | | |

Click on the  icon and in the ensuing Test Statistic Calculator, enter sample size as 215, δ as 8, and SE as 7.2291. Click OK. The incremental test

57 Conditional Power for Decision Making

statistic, will be computed as 1.1066 ($\hat{\delta}/\hat{SE} = 8/7.2291 = 1.1066$).

Similarly, for the second look, click on the **Enter Interim Data** icon and then enter the estimates of the incremental accrual, δ and SE as 220, 9.2 and 7.4162 respectively in the Test Statistic Calculator. Click OK. The computed values will be posted in the interim monitoring sheet as shown below.

| Look # | Incremental Sample Size | Incremental Statistic | Cumulative Sample Size | Prespecified Weights | Weight ... Statistic | δ | Standard Error | Efficacy | 97.5% RCI for δ | | Repeat ... p-value |
|--------|-------------------------|-----------------------|------------------------|----------------------|----------------------|----------|----------------|----------|------------------------|---------|--------------------|
| | | | | | | | | | Upper | Lower | |
| 1 | 215 | 1.107 | 215 | 0.2 | 1.107 | 8 | 7.229 | 4.877 | Infinity | -27.255 | 0.503 |
| 2 | 220 | 1.241 | 435 | 0.2 | 1.66 | 9.2 | 7.416 | 3.357 | Infinity | -8.787 | 0.216 |
| 3 | | | | 0.2 | | | | 2.68 | | | |
| 4 | | | | 0.2 | | | | 2.29 | | | |
| 5 | | | | 0.2 | | | | 2.031 | | | |

After any interim look, based on the observed values of δ and σ , you will be able to use conditional power calculator to estimate either conditional power or sample size using appropriate inputs as shown in Table 57.1.

Let us examine the use of conditional power calculator with a few examples. From the interim monitoring sheet, click on the **Conditional Power Calculator** icon

Table 57.1: Conditional Power Calculator Use

| Estimate | Input |
|-------------------|--|
| conditional power | observed δ/σ design sample size |
| conditional power | observed δ/σ user specified sample size |
| conditional power | user specified δ/σ design sample size |
| conditional power | user specified δ/σ user specified sample size |
| sample size | observed δ/σ desired conditional power |
| sample size | user specified δ/σ desired conditional power |

 to invoke the conditional power calculator as displayed below.

Conditional Power Calculator

Input

Look #:

Cumulative Sample Size:

Weighted Test Statistic:

Input/Output

Value of δ :

Value of σ :

Value of δ/σ :

Computed Conditional Power:

Sample Size (Overall):

*Use radio button to select the quantity to be computed.

Recalc Plot Close

57 Conditional Power for Decision Making

The calculator is divided into two parts. The first part is the input part. The values for the cells in this part are automatically filled using the interim monitoring sheet values. The calculator indicates that the second interim look has been taken, the cumulative sample size is 435 and the weighted z statistic after the second interim look is 1.66.

The second part is the input/output part. Here you may decide to estimate either conditional power or sample size by clicking on the appropriate radio button, and then specifying the required input as detailed in Table 57.1. By default, the calculator is showing the value of δ/σ as 0.2, which is the estimate obtained from the incremental data of the second look. The interpretation is that if the hypothesized value of δ/σ is 0.2, then the conditional power to reach significance at any future look with a maximum sample size of 1075 is 0.905.

Computing Conditional Power for Specified Sample Size Now, suppose, you estimate, using cumulated data, that δ/σ is likely to be 0.1593, then enter this value in the calculator and click on **Recalc** button. The calculator will display the new estimate for conditional power, which is 0.789.

Conditional Power Calculator

Input

Look #:

Cumulative Sample Size:

Weighted Test Statistic:

Input/Output

Value of δ :

Value of σ :

Value of δ/σ :

Computed Conditional Power:

Sample Size (Overall):

*Use radio button to select the quantity to be computed.

If you want to enter another set of estimates, say, $\delta = 7.5$ and $\sigma = 45.0$, and want to enter each value separately, you can do that first by clicking on the top check box and then entering the values. Then click on **Recalc** button to get the new estimate of

conditional power as 0.815.

Conditional Power Calculator

Input

Look #:

Cumulative Sample Size:

Weighted Test Statistic:

Input/Output

Value of δ :

Value of σ :

Value of δ/σ :

Computed Conditional Power:

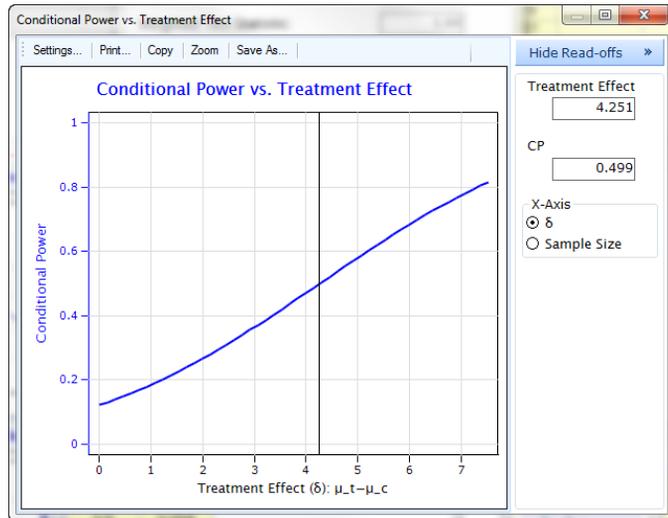
Sample Size (Overall):

*Use radio button to select the quantity to be computed.

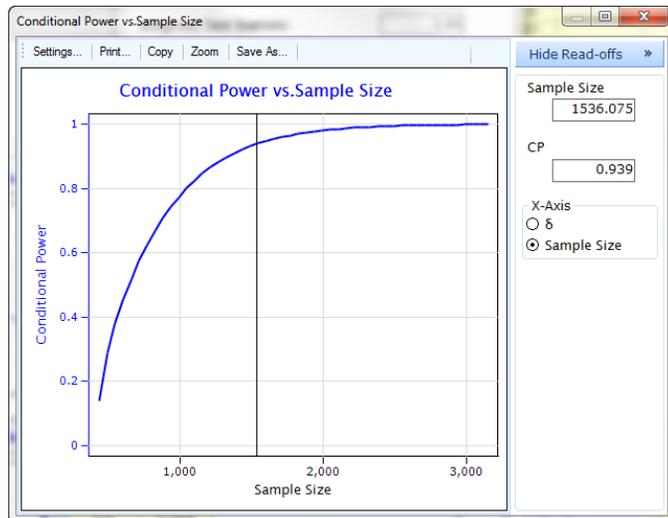
To view a plot of **conditional power vs. delta**, click on the **Plot** button

57 Conditional Power for Decision Making

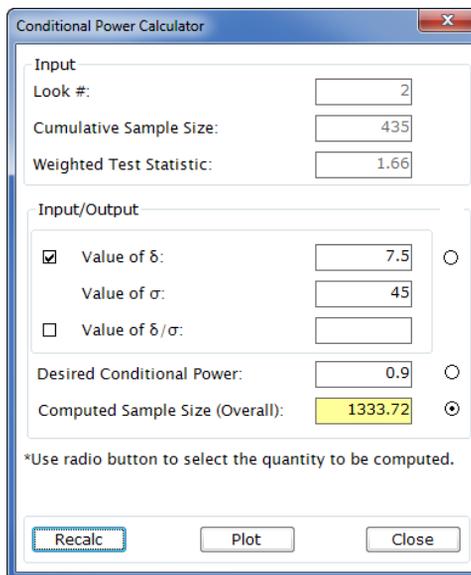
and a plot will appear as shown below.



In the above plot, if you click on the radio button against **sample size**, you will get the **conditional power vs. sample size** plot displayed below.



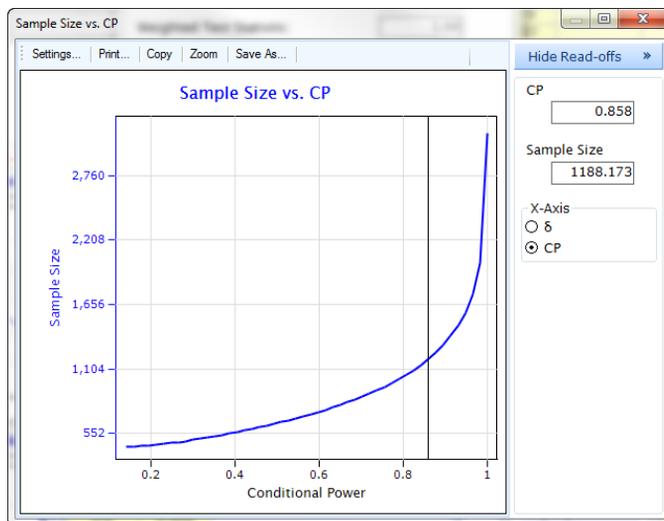
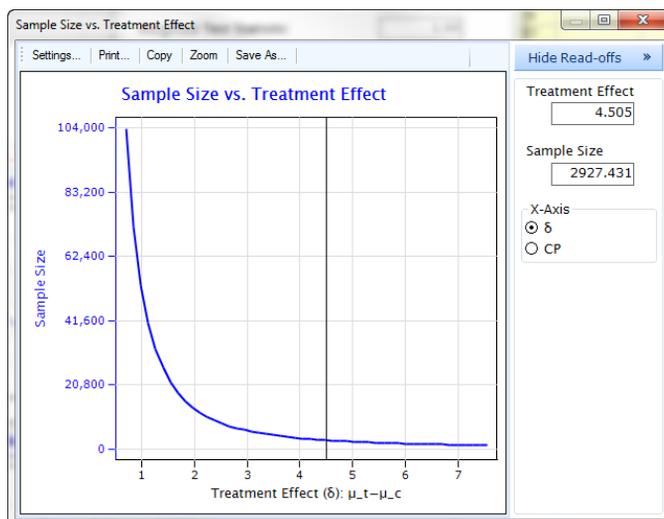
Computing Sample Size for Desired Conditional Power With the values of $\delta = 7.5$, $\sigma = 45.0$, and $sample\ size = 1075$, we obtained the conditional power estimate as 0.815. Now, keeping the same values for δ and σ , if you want to estimate the sample size for a desired conditional power of 0.90, you can proceed like this: Click on the radio button against **sample size** input box, enter the value of 0.90 for conditional power and then click on **Recalculate** button. You will get the estimate of sample size to be 1334 as displayed in the screen shot below.



With the above setting in the conditional power calculator, you can click on the **Plot** button to get the **sample size vs. delta** and **sample size vs.**

57 Conditional Power for Decision Making

conditional power plots as displayed below.



57.1.2 Binomial Endpoint

Consider a two-arm, placebo controlled randomized clinical trial for subjects with acute cardiovascular disease undergoing percutaneous coronary intervention (PCI).

The primary endpoint is a composite of death, myocardial infarction or ischemia-driven revascularization during the first 48 hours after randomization. We assume on the basis of prior knowledge that the event rate for the placebo arm is 8.7%. The investigational drug is expected to reduce the event rate by at least 20%. The investigators are planning to randomize a total of 8000 subjects in equal proportions to the two arms of the study. Let us design with help of East that a group sequential 3 looks design to detect a 20% risk reduction with a one-sided level-0.025 test of significance (with 0.087 on the control arm and $0.8 \times 0.087 = 0.0696$ on the treatment arm). It is also decided that two interim looks, one after 4000 subjects are enrolled (50% of total information) and the second after 5600 subjects are enrolled (70% of total information) will be taken. Early stopping efficacy boundaries are derived from the Lan and DeMets (1983) O'Brien-Fleming type error spending function.

With the above specifications, create a plan in East as shown below.

| | Bin:Des 1 |
|--|-------------------|
| Mnemonic | PN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.82 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.087 |
| Proportion under Treatment (π_t) | 0.07 |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.017 |
| Variance | Unpooled Estimate |
| Boundary Parameters | |
| Spacing of Looks | Unequal |
| Efficacy Boundary | LD (OF) |
| Sample Size | |
| Maximum | 8000 |
| Expected Under H0 | 7979.837 |
| Expected Under H1 | 6535.034 |

We will now monitor the trial. Select  icon to invoke the CHW interim monitoring sheet. You will then be taken to the interim monitoring worksheet

57 Conditional Power for Decision Making

displayed below.

| Look # | Incremental Sample Size | Incremental Statistic | Cumulative Sample Size | Prespecified Weights | Weight ... Statistic | δ | Standard Error | Efficacy | 97.5% RCI for δ | | Repeat ... p-value |
|--------|-------------------------|-----------------------|------------------------|----------------------|----------------------|----------|----------------|----------|------------------------|-------|--------------------|
| | | | | | | | | | Upper | Lower | |
| 1 | | | | 0.5 | | | | -2.963 | | | |
| 2 | | | | 0.2 | | | | -2.462 | | | |
| 3 | | | | 0.3 | | | | -2.002 | | | |

The first interim look was taken after accruing 4000 patients, 2000 per treatment arm. We input this number in the Incremental accrual number in the row corresponding to the first look. To calculate the incremental statistic, we utilize the test statistic calculator. There are 174 events in the control arm and 147 events in the treatment arm. Based on these data the estimate of δ is $(147/2000) - (174/2000) = -0.0135$ and the estimate of $SE = 0.0086$. So the value of the test statistic is $SE/\text{estimate of } \delta = -1.5718$. These values are entered in the test statistic calculator as shown below.

Test Statistic Calculator

Editing Look # 1 For Incremental Sample Size

Sample Size and Responses

| | Control | Treatment |
|--------------------------|-----------------------------------|-----------------------------------|
| Incremental Sample Size: | <input type="text" value="2000"/> | <input type="text" value="2000"/> |
| Incremental Response: | <input type="text" value="174"/> | <input type="text" value="147"/> |

Sample Size:

Input for Binomial end point

Estimate of δ :

$\delta = (\pi_t - \pi_c)$

Standard Error of Estimate of δ :

Output

Weighted Statistic:

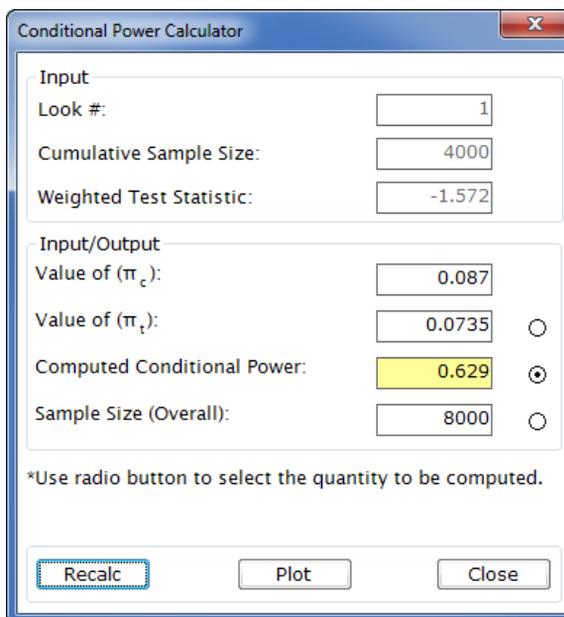
Incremental Statistic:

Click on **OK** and the values of incremental accrual and incremental test statistic will appear in the IM sheet.

| Look # | Incremental Sample Size | Incremental Statistic | Cumulative Sample Size | Prespecified Weights | Weight ... Statistic | δ | Standard Error | Efficacy | 97.5% RCI for δ | | Repeat ... p-value |
|--------|-------------------------|-----------------------|------------------------|----------------------|----------------------|----------|----------------|----------|------------------------|-------|--------------------|
| | | | | | | | | | Upper | Lower | |
| 1 | 4000 | -1.572 | 4000 | 0.5 | -1.572 | -0.014 | 0.009 | -2.963 | 0.012 | -1 | 0.18 |
| 2 | | | | 0.2 | | | | -2.462 | | | |
| 3 | | | | 0.3 | | | | -2.002 | | | |

Conditional Power Calculator After each look during interim monitoring the decision to alter the sample size can be made using the conditional power calculator. After any interim look, based on the observed data, you will be able to use conditional power calculator to estimate any one of the three quantities - conditional power or sample size or π_t -given the estimates of other two and any specified value of π_c .

Computing Power for a pre specified sample size Click on the icon  from the IM toolbar to invoke the conditional power calculator as shown below.



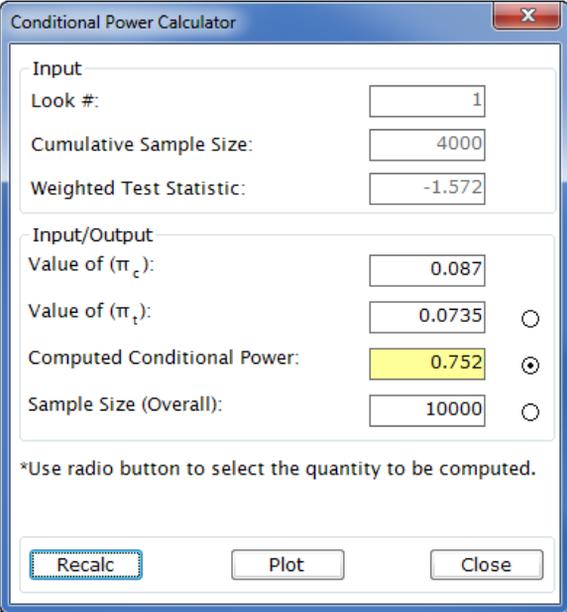
The calculator is divided into 2 parts. The first part displays the inputs that are used in the interim monitoring sheet till the current look. The Cumulative accrual is 4000 and the current weighted test statistic is -1.5718.

57 Conditional Power for Decision Making

The second part helps the user to estimate the value of a desired parameter by selecting the radio button against the parameter and then entering the values for other parameters and clicking on **Recalc** button.

In the current scenario, for a final overall size of 8000, and the hypothesized values of $\pi_c = 0.087$ and $\pi_t = 0.0735$, the conditional power is estimated to be 0.629.

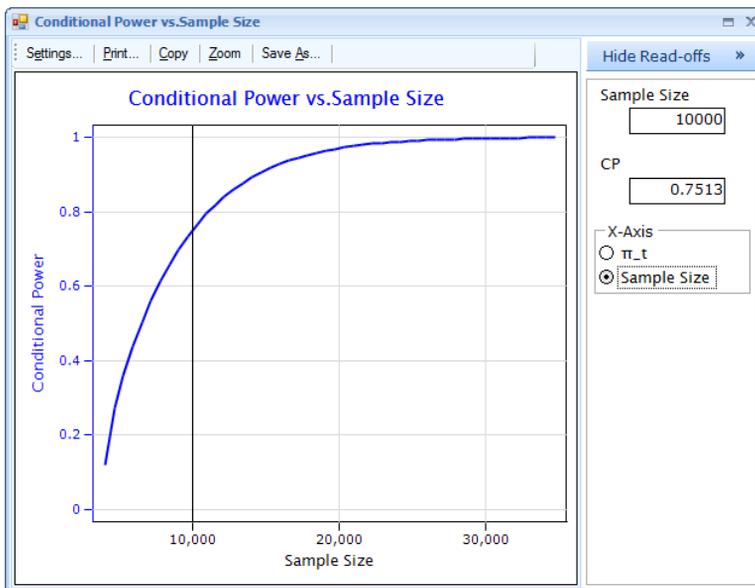
Now keeping the values of π_c , and π_t same, if you want to estimate the conditional power for an increased sample size of 10,000, then enter this value and click on **Recalc** button to see the conditional power estimate to be 0.752.



The image shows a software dialog box titled "Conditional Power Calculator". It contains several input fields and radio buttons. The "Input" section has fields for "Look #:" (value 1), "Cumulative Sample Size:" (value 4000), and "Weighted Test Statistic:" (value -1.572). The "Input/Output" section has fields for "Value of (π_c):" (value 0.087), "Value of (π_t):" (value 0.0735), "Computed Conditional Power:" (value 0.752, highlighted in yellow), and "Sample Size (Overall):" (value 10000). There are radio buttons next to the π_t and Sample Size fields. A note at the bottom says "*Use radio button to select the quantity to be computed." At the bottom of the dialog are three buttons: "Recalc", "Plot", and "Close".

Now you may click on the **Plot** button and choose x-axis to represent sample size, to see the graph of conditional power vs. sample size, assuming $\pi_t = 0.0735$,

$$\pi_c = 0.087.$$



Re-estimating Sample Size for a desired power

If a final overall sample size is to be estimated for a desired value of conditional power, the user can do so by selecting the sample size radio button in the calculator. Suppose, the user wants to estimate the increase required in the final sample size for a desired conditional power of 80%. The user can select the radio button next to the **Final Sample Size** input box and enter the value of 0.8 for conditional power and then click on **Recalc** button. The result in the conditional power calculator will appear as

57 Conditional Power for Decision Making

shown below.

The image shows a software dialog box titled "Conditional Power Calculator". It is divided into two main sections: "Input" and "Input/Output".

Input Section:

- Look #: 1
- Cumulative Sample Size: 4000
- Weighted Test Statistic: -1.572

Input/Output Section:

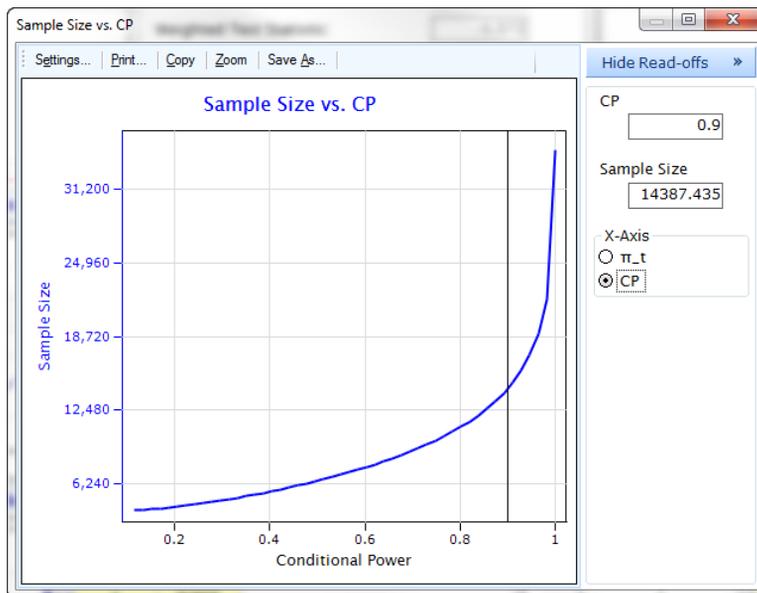
- Value of (π_c): 0.087
- Value of (π_t): 0.0735 (radio button is unselected)
- Desired Conditional Power: 0.8 (radio button is unselected)
- Computed Sample Size (Overall): 11057.84 (radio button is selected, and the value is highlighted in yellow)

Below the input/output section, there is a note: "*Use radio button to select the quantity to be computed."

At the bottom of the dialog box, there are three buttons: "Recalc", "Plot", and "Close".

The calculator estimates a final sample size of 11058 for a desired conditional power of 0.8 for the values of $\pi_t = 0.0735$ and $\pi_c = 0.087$. After clicking on the **Plot**

button, the user can view the plot of sample size vs. conditional power as shown below.



57.1.3 Time to Event Endpoint

A two-arm multi-center randomized clinical trial is planned for subjects with advanced metastatic non-small cell lung cancer with the goal of comparing the current standard second line therapy (docetaxel+cisplatin) to a new docetaxel containing combination regimen. The primary endpoint is overall survival (OS). The study is required to have one-sided $\alpha = 0.025$, and 90% power to detect an improvement in median survival, from 8 months on the control arm to 11.4 months on the experimental arm, which corresponds to a hazard ratio of 0.7. Accrual duration is 24 months and the study duration 30 months. We shall first create a three look group sequential design for

57 Conditional Power for Decision Making

this study in East as shown below.

| Surv:Des 1 | |
|---|--------------------|
| Mnemonic | SU-2S-LRSD |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9008 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.702 |
| Var (Log HR) | Null |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Accrual & Dropout Parameters | |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 492 |
| Expected Under H0 | 491.5725 |
| Expected Under H1 | 464.4281 |
| Events | |
| Maximum | 340 |
| Expected Under H0 | 339.3008 |
| Expected Under H1 | 272.5305 |
| Study Duration | |
| Maximum | 30 |
| Expected Under H0 | 27.523 |
| Expected Under H1 | 25.25 |
| Accrual Duration | |
| Maximum | 24 |
| Expected Under H0 | 23.979 |
| Expected Under H1 | 22.655 |

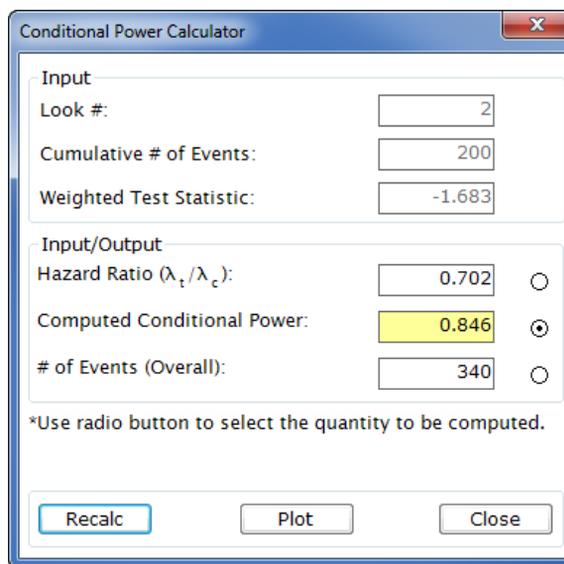
We will now monitor the trial. Invoke the CHW interim monitoring sheet. Enter at the first look, the cumulative events as 110 and the cumulative test statistic, using test statistic calculator, as 1.220 ($\hat{\delta}/\hat{SE} = -0.288/0.236 = -1.220$). At the second look, enter the incremental accrual as 200 and again use the test calculator to enter ($\hat{\delta}/\hat{SE} = -0.324/0.195 = -1.662$).

Now the interim monitoring sheet will appear as displayed below.

| Look # | Cumulative Events | Cumulative Statistic | Incremental Statistic | Prespecified Weights | Weight... Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | 97.5% RCI for HR | | Repeated p-value |
|--------|-------------------|----------------------|-----------------------|----------------------|---------------------|------------------|--------------------------------|----------|------------------|-------|------------------|
| | | | | | | | | | Upper | Lower | |
| 1 | 110 | -1.22 | -1.22 | 0.332 | -1.22 | -0.288 | 0.236 | -3.716 | 1.802 | 0 | 0.358 |
| 2 | 200 | -1.662 | -1.16 | 0.335 | -1.683 | -0.324 | 0.195 | -2.509 | 1.178 | 0 | 0.106 |
| 3 | | | | 0.332 | | | | -1.993 | | | |

After any interim look, based on the observed values of δ and its SE, you will be able to use conditional power calculator to estimate either conditional power or number of events using appropriate inputs.

Let us examine the use of conditional power calculator with a few examples. From the interim monitoring sheet, click on the icon  to invoke the conditional power calculator as displayed below.



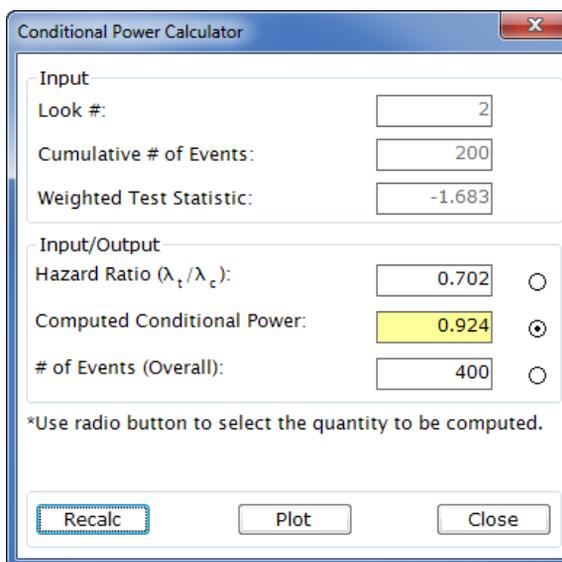
The calculator is divided into two parts. The first part is the input part. The values for the cells in this part are automatically filled using the interim monitoring sheet values. The calculator indicates that the second interim look has been taken, the cumulative number of events is 200 and the weighted z statistic after the second interim look is -1.683 .

The second part is the input/output part. Here you may decide to estimate any of the three quantities - required HR, conditional power, number of events by clicking on the appropriate radio button, and then specifying the input for the other two quantities.

Computing Conditional Power for Specified Number of Events Now, suppose, you estimate that with the available budget you can extend the study to cover 400 events. In that scenario you may want to know the effect on the conditional power. With the radio

57 Conditional Power for Decision Making

button selected to compute conditional power, enter the value of 400 as the number of events and click on **Recalc** button. The calculator will display the new estimate for conditional power, which is 0.9239.



The image shows a software dialog box titled "Conditional Power Calculator". It contains two main sections: "Input" and "Input/Output".

Input Section:

- Look #:
- Cumulative # of Events:
- Weighted Test Statistic:

Input/Output Section:

- Hazard Ratio (λ_t / λ_c):
- Computed Conditional Power:
- # of Events (Overall):

*Use radio button to select the quantity to be computed.

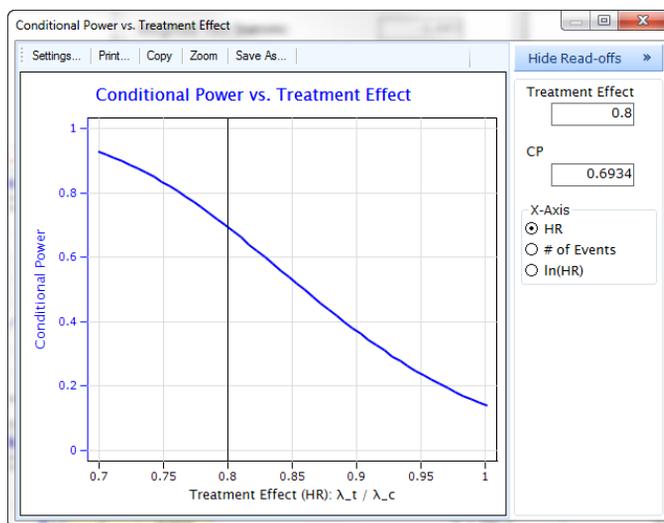
At the bottom, there are three buttons: "Recalc" (highlighted with a dashed border), "Plot", and "Close".

To view a plot of **conditional power vs. number of events**, click on the **Plot** button, select number of events as the x-axis variable and a plot will appear

as shown below.

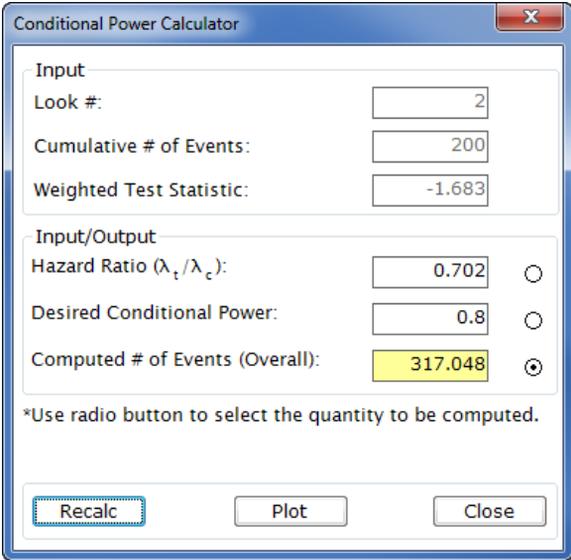


In the above plot, if you click on the radio button against **HR**, you will get the **conditional power vs. HR** plot displayed below.



57 Conditional Power for Decision Making

Computing Sample Size for Desired Conditional Power If you would like to estimate the number of events required for a specified conditional power, say 0.80, you can click on the radio button against # of events, enter 0.80 as conditional power, and then click on **Recalc** button. The calculator will display the estimate for number of events, which is 317.



The image shows a software dialog box titled "Conditional Power Calculator". It contains two main sections: "Input" and "Input/Output".

Input Section:

- Look #: 2
- Cumulative # of Events: 200
- Weighted Test Statistic: -1.683

Input/Output Section:

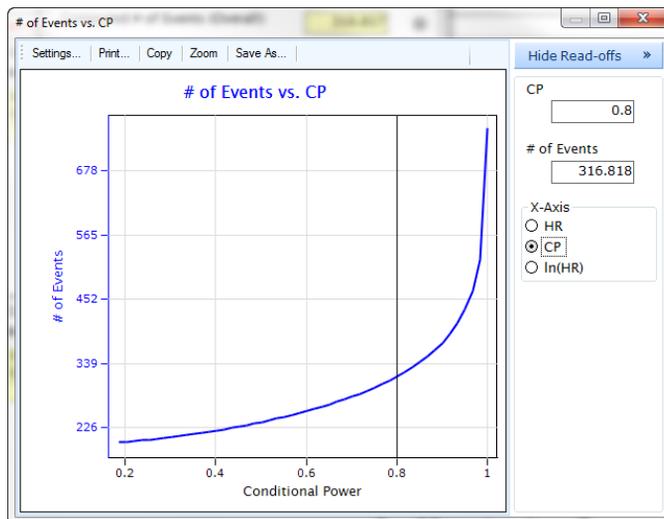
- Hazard Ratio (λ_t / λ_c): 0.702
- Desired Conditional Power: 0.8
- Computed # of Events (Overall): 317.048

Below the radio buttons, there is a note: "*Use radio button to select the quantity to be computed."

At the bottom of the dialog, there are three buttons: "Recalc", "Plot", and "Close".

To view a plot of **number of events vs. conditional power**, click on the **Plot** button, select conditional power as the x-axis variable and a plot will appear

as shown below.



57.2 CP Calculator - CHW: Simualtion

This section explains the use of conditional power calculator while performing adaptive simulations. Simulation capabilities can be useful in verifying the operating characteristics of the design.

57.2.1 Normal Endpoint

Let us use the design for normal endpoint that we discussed in section 57.1.1 which is

57 Conditional Power for Decision Making

shown below.

| | Wbk1:Des1 |
|---|---------------------|
| Mnemonic | MN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 5 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Input Method | Difference of Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 10 |
| Std. Deviation (σ) | 50 |
| Test Statistic | Z |
| Allocation Ratio (nt/nc) | 1 |
| Boundary Parameters | |
| Efficacy Boundary | LD (OF) |
| Spacing of Looks | Equal |
| Sample Size | |
| Maximum | 1075 |
| Expected Under H0 | 1071.4709 |
| Expected Under H1 | 797.1678 |

Save this design in the library and then click on the icon  to get the simulation worksheet. In this sheet, in the **Include Options** button, choose **Sample Size Re-estimation**. You will get the a simulation worksheet. Click on the tab **Sample Size**

Re-estimation to see the simulation sheet with this tab opened.

Keep the simulation parameters displayed in other tabs without any change. The default values for simulation suggest that the difference of means to be 10.0 and the standard deviation to be 50. The maximum sample size to be used till look $L=4$, is $N_{max} = 1075$. Let us change this max value to 2150 by modifying the multiplier value from 1 to 2. Also the criterion for when to adapt the sample size is specified by a range of conditional power value from 0.3 to 0.9. Thus after the second look, if the conditional power computed lies between 0.3 and 0.9, the simulation will increase the sample size to a maximum of 2150, so that the conditional power can rise to the desired level of 0.9. In order to assess and observe the effect of varying the values of these simulation parameters, we use the conditional power calculator. Based on the results we get from the conditional power calculator, we decide on a set of values for the simulation parameters and then carry out the simulation.

Open the conditional power calculator by clicking on the icon  and the

57 Conditional Power for Decision Making

calculator will appear as shown below.

The image shows a software dialog box titled "Conditional Power Calculator". It is divided into two main sections. The first section, "Input", contains two text input fields: "Current Look:" with the value "4" and "Current Sample Size:" with the value "860". The second section, "Input/Output", is titled " δ/σ to be Used in Conditional Power Computation" and contains two radio buttons: "Arbitrary (δ/σ)" (unselected) and "Estimated ($\delta/\sigma, Z$)" (selected). Below these are two groups of input fields. The first group, under the "Arbitrary" option, has a checkbox (unchecked) and three input fields: "Value of δ :" (empty), "Value of σ :" (empty), and "Value of δ/σ :" (containing "0.131"). The second group, under the "Estimated" option, has a checkbox (checked) and two input fields: "Value of δ/σ :" (containing "0.131") and "Value of z :" (containing "1.918"). At the bottom of the dialog, there are two more input fields: "Computed Conditional Power:" (containing "0.6") and "Sample Size (Overall):" (containing "1075"). A note at the bottom reads: "* Use radio button to select the quantity to be computed." At the very bottom are two buttons: "Recalc" and "Close".

The conditional power calculator is divided into two parts. The first part lists the inputs and gives the current look position and the sample size at the current look.

The second part is used to compute either conditional power or sample size given the other quantity and appropriate values among δ/σ , z , δ , and σ depending on the choice made between the options **Arbitrary** and **Estimated**. For example if you want to estimate the conditional power for the estimated values of $\delta = 8$ and $\sigma = 70$ and for a maximum sample size of 2150, enter these values in the calculator and click on **Recalc** button. The calculator will compute the values of z and the conditional

power as shown below.

Conditional Power Calculator

Input
 Current Look:
 Current Sample Size:

Input/Output
 δ/σ to be Used in Conditional Power Computation
 Arbitrary (δ/σ) Estimated ($\delta/\sigma, Z$)

Value of δ :
 Value of σ :
 Value of δ/σ :
 Value of z :

Computed Conditional Power:
 Sample Size (Overall):

* Use radio button to select the quantity to be computed.

The estimated conditional power of 0.8058 indicates that even with a maximum sample size of 2150, the conditional power cannot reach the desired level of 0.90, if the estimated values of $\delta = 8$ and $\sigma = 70$ represent the true values of the population.

The quantity δ/σ can either be estimated or design values, using the drop down box against **CP Computation Based on:**. For this simulation, we will use the estimated value of δ/σ and Z . Select the option **Estimated $\delta/\sigma, Z$** .

Computing overall Sample Size

Suppose we wish to compute the overall sample size required for a conditional power of 0.9. Select the radio button next to the overall sample size and enter the value of 0.9

57 Conditional Power for Decision Making

for conditional power. Press **Recalc** to obtain the result as shown below

Conditional Power Calculator

Input
 Current Look:
 Current Sample Size:

Input/Output
 δ/σ to be Used in Conditional Power Computation
 Arbitrary (δ/σ) Estimated ($\delta/\sigma, Z$)

Value of δ :
 Value of σ :
 Value of δ/σ :
 Value of z :

Conditional Power:
 Computed Sample Size (Overall):

* Use radio button to select the quantity to be computed.

The calculator shows that the overall sample size for the desired conditional power of 0.90 is 2730.8. You may enter this sample size as the **Max. Usable sample size** by specifying the multiplier as $2730.8/1075 = 2.5403$ in the simulation sheet along with appropriate values for other simulation parameters and then carry out the simulation.

57.2.2 Binomial Endpoint

This section looks at the use of conditional power calculator during the adaptive simulation of trials with binomial endpoints.

Let us use the design for binomial endpoint that we discussed in section 57.1.2 which

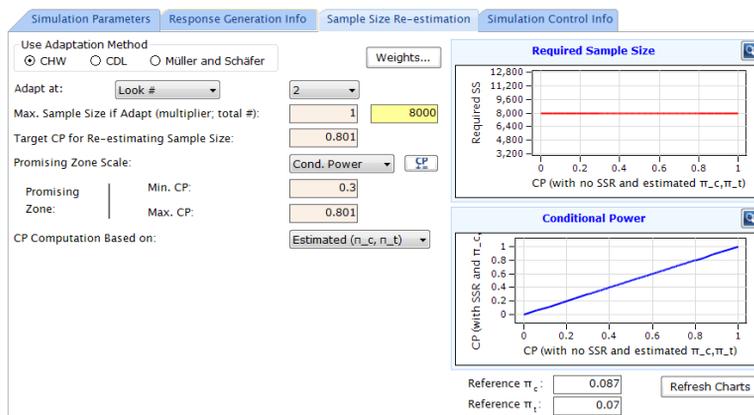
is shown below.

| | Bin:Des 1 |
|--|-------------------|
| Mnemonic | PN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.82 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.087 |
| Proportion under Treatment (π_t) | 0.07 |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.017 |
| Variance | Unpooled Estimate |
| Boundary Parameters | |
| Spacing of Looks | Unequal |
| Efficacy Boundary | LD (OF) |
| Sample Size | |
| Maximum | 8000 |
| Expected Under H0 | 7979.8367 |
| Expected Under H1 | 6535.0339 |

We will now simulate this plan using adaptive simulation in East. click on the icon  to get the simulation worksheet. In this sheet, in the **Include Options** button, choose **Sample Size Re-estimation**. You will get the a simulation worksheet. Click on

57 Conditional Power for Decision Making

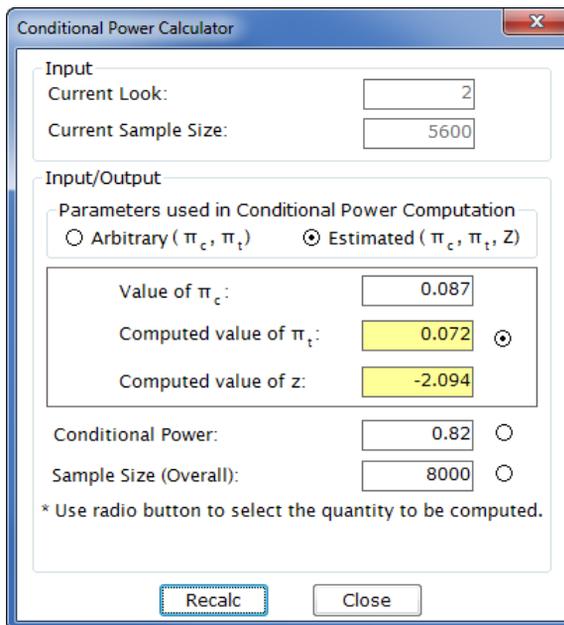
the tab **Sample Size Re-estimation** to see the simulation sheet with this tab opened.



Keep the simulation parameters displayed in other tabs without any change. The default values indicate that the proportion of response for control arm to be 0.087 and for the treatment arm to be 0.0696. The criterion for when to adapt the sample size is specified by a range of conditional power value from 0.3 to 0.82. Change the multiplier value from 1 to 2, so as to get maximum sample size if adapt to become 16000. Thus, after the second look, if the estimated conditional power lies between 0.3 and 0.82, then the simulation process will increase the sample size to a maximum of 16000, so as to raise the conditional power to the desired level of 0.82. In order to assess and observe the effect of varying the values of these simulation parameters, you may use the conditional power calculator. Based on the results you get from the conditional power calculator, you may decide on a set of values for the simulation parameters and then carry out the simulation.

Open the conditional power calculator by clicking on the icon  and the

calculator will appear as shown below.



The conditional power calculator is divided into two parts. The first part lists the inputs and gives the current look position and the current sample size.

The second part is used to compute either conditional power or sample size given the other quantity and appropriate values among π_c , π_t , and z , depending on the choice made between the options **Arbitrary** and **Estimated**.

Computing conditional power for a specified sample size For example if you want to estimate the conditional power for the estimated values of $\pi_c = 0.085$ and $\pi_t = 0.074$ and for a maximum sample size of 16,000, enter these values in the calculator and click on **Recalc** button. The calculator will compute the values of z

57 Conditional Power for Decision Making

and the conditional power as shown below.

The image shows a software dialog box titled "Conditional Power Calculator". It contains several input fields and radio buttons. Under "Input", "Current Look" is set to 2 and "Current Sample Size" is 5600. Under "Input/Output", the "Parameters used in Conditional Power Computation" section has two radio buttons: "Arbitrary (π_c, π_t)" (unselected) and "Estimated (π_c, π_t, Z)" (selected). Below this, "Value of π_c " is 0.085, "Value of π_t " is 0.074 (with an unselected radio button), and "Value of z" is -1.522. The "Computed Conditional Power" is 0.771 (with a selected radio button), and "Sample Size (Overall)" is 16000 (with an unselected radio button). A note at the bottom states: "* Use radio button to select the quantity to be computed." At the bottom of the dialog are "Recalc" and "Close" buttons.

The computed conditional power of 0.7715 indicates that the maximum usable sample size of 16,000 may have to be increased in order to attain the desired conditional power of 0.82.

Computing overall Sample Size for a desired conditional power Now suppose we wish to compute the overall sample size required for a conditional power of 0.82. Select the radio button next to the overall sample size and enter the value of 0.82 for Computed conditional power. Press **Recalc** button to obtain the result as shown

below.

Conditional Power Calculator

Input
Current Look: 2
Current Sample Size: 5600

Input/Output
Parameters used in Conditional Power Computation
 Arbitrary (π_c, π_t) Estimated (π_c, π_t, Z)

Value of π_c : 0.085
Value of π_t : 0.074
Value of z: -1.522

Conditional Power: 0.82
Computed Sample Size (Overall): 17793.723

* Use radio button to select the quantity to be computed.

Recalc Close

East computes that the required over all sample size for the desired conditional power is 17793.7. Now you may enter this value for **Max. Sample Size** in the simulation sheet and then carry out simulation.

57.2.3 Time to Event Endpoint

Let us use the design for survival endpoint that we discussed in section 57.1.3 which is

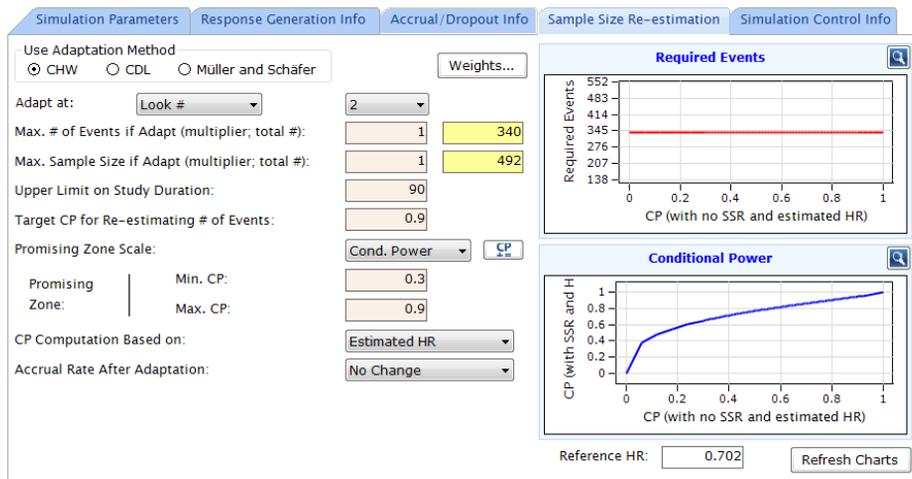
57 Conditional Power for Decision Making

shown below.

| | Surv:Des 1 |
|---|--------------------|
| Mnemonic | SU-25-LRSD |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9008 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.702 |
| Var (Log HR) | Null |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Accrual & Dropout Parameters | |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 492 |
| Expected Under H0 | 491.5725 |
| Expected Under H1 | 464.4281 |
| Events | |
| Maximum | 340 |
| Expected Under H0 | 339.3008 |
| Expected Under H1 | 272.5305 |
| Study Duration | |
| Maximum | 30 |
| Expected Under H0 | 27.523 |
| Expected Under H1 | 25.25 |
| Accrual Duration | |
| Maximum | 24 |
| Expected Under H0 | 23.979 |
| Expected Under H1 | 22.655 |

We will now simulate this plan using adaptive simulation in East. click on the icon  to get the simulation worksheet. In this sheet, in the **Include Options** button, choose **Sample Size Re-estimation**. You will get the a simulation worksheet. Click on

the tab **Sample Size Re-estimation** to see the simulation sheet with this tab opened.



Keep the simulation parameters displayed in other tabs without any change. The default values for simulation suggest that the hazard rates for control and treatment arms as 0.0866 and 0.0607 respectively with the resulting Hazard Ratio of 0.70. The maximum number of events to be used till look $L=2$, is $MaxEvents = 340$ with the multiplier at the default value of 1.0. Also the criterion for when to adapt the number of events is specified by a range of conditional power value from 0.3 to 0.9. The target CP is at the default value of 0.90. In order to assess and observe the effect of varying the values of these simulation parameters, we use the conditional power calculator. Based on the results we get from the conditional power calculator, we decide on a set of values for the simulation parameters and then carry out the simulation.

Open the conditional power calculator by clicking on the icon  and the

57 Conditional Power for Decision Making

calculator will appear as shown below.

The screenshot shows a dialog box titled "Conditional Power Calculator" with a close button (X) in the top right corner. The dialog is divided into two main sections: "Input" and "Input/Output".

Input Section:

- Current Look:
- Current # of Events:

Input/Output Section:

HR to be Used in Conditional Power Computation

Arbitrary (HR) Estimated (HR, z)

Value of HR:

Value of z:

Computed Conditional Power:

of Events (Overall):

* Use radio button to select the quantity to be computed.

Buttons: Recalc, Close

The conditional power calculator is divided into two parts. The first part lists the inputs and gives the current look position and the number of events at the current look.

The second part is used to compute either conditional power or number of events given the other quantity and appropriate values of HR, and z , depending on the choice made between the options **Arbitrary** and **Estimated**. For example if you want to estimate the conditional power for the estimated value of HR = 0.8 and for a maximum no.of events of 500, enter these values in the calculator and click on **Recalc** button.

The calculator will compute the values of z and the conditional power as shown below.

The estimated conditional power of 0.779 indicates that even with the number of events at a maximum of 500, the conditional power cannot reach the desired level of 0.90, if the estimated Hazard Ratio of 0.80 represents the true value of the population.

The quantity of Hazard Ratio can either be defined by the user or estimated or design values. For this simulation, we will use the estimated value of Hazard Ratio. Select the radio button next to **Estimated (HR, Z)**.

Computing Number of Events (Overall)

Suppose we wish to compute the number of events (overall) required for a conditional power of 0.9. Select the radio button next to the # of Events (Overall) and enter the

57 Conditional Power for Decision Making

value of 0.9 for conditional power. Press **Recalc** to obtain the result as shown below

The image shows a dialog box titled "Conditional Power Calculator". It contains several input fields and radio buttons. The "Input" section has "Current Look:" set to 2 and "Current # of Events:" set to 227. The "Input/Output" section has "HR to be Used in Conditional Power Computation" with "Arbitrary (HR)" unselected and "Estimated (HR, z)" selected. Below this, "Value of HR:" is 0.8 and "Value of z:" is -1.681. The "Conditional Power:" is 0.9 and "Computed # of Events (Overall):" is 673.089. A note at the bottom says "* Use radio button to select the quantity to be computed." There are "Recalc" and "Close" buttons at the bottom.

| Field | Value |
|--------------------------------|---------|
| Current Look | 2 |
| Current # of Events | 227 |
| Value of HR | 0.8 |
| Value of z | -1.681 |
| Conditional Power | 0.9 |
| Computed # of Events (Overall) | 673.089 |

The calculator shows that the overall sample size for the desired conditional power of 0.90 is 673. You may specify this number as the **Max. Events if Adapt (multiplier, total #)** by entering the multiplier as $673/340 = 1.9794$ in the simulation sheet. You may then specify appropriate values for other simulation parameters and then carry out the simulation.

Volume 8 *Special Topics*

- 58 *Introduction to Volume 8* 1388
- 59 *Design and Monitoring of Maximum Information Studies* 1393
- 60 *Design and Interim Monitoring with General Endpoints* 1423
- 61 *Early Stopping for Futility* 1434
- 62 *Flexible Stopping Boundaries in East* 1460
- 63 *Confidence Interval Based Design* 1493
- 64 *Simulation in East* 1552
- 65 *Predictive Interval Plots* 1575
- 66 *Enrollment/Events Prediction - At Design Stage (By Simulation)* 1609
- 67 *Conditional Simulation* 1658
- 68 *Enrollment/Events Prediction - Analysis* 1675
- 69 *Interfacing with East PROCs* 1787

58 *Introduction to Volume 8*

This volume contains Chapters 58 through 69. These chapters describe special design and monitoring tools that, rather than being end-point specific, cut across all different types of group sequential designs.

Chapter 59 deals with the design and monitoring of trials on an information scale rather than on a sample size scale. By fixing the maximum information but allowing the sample size to float one can ensure that a study will be adequately powered despite poor initial guesses about nuisance parameters like σ^2 .

Chapter 60 describes how one can convert any fixed sample design into a group sequential design. Suppose, for example, that you wish to run a three period cross-over study as a group sequential design with interim looks for early stopping for efficacy and futility. Since East does not at present support this type of design you may first obtain the necessary sample size for a single look design on your own, perhaps with other commercial software. This sample size would be input to East and the single look design would then be converted into a group sequential design with stopping boundaries and a corresponding inflated sample size.

Chapter 61 discusses early stopping for futility.

Chapter 62 describes all the different types of stopping boundary families that are available in East, such as Haybittle-Peto, Wang-Tsiatis, Lan-DeMets etc.

Chapter 63 illustrates through several examples how East may be used to obtain sample sizes that are based on the desired width of a confidence interval for the parameter of interest rather than being based on the desired power of a hypothesis test.

Chapter 64 discusses the various types of simulation tools provided by East.

Chapter 65 explains the concept of predicting the future course of a trial with Predictive Interval Plots.

Chapter 66 discusses the Enrollment/Events Prediction At Design Stage.

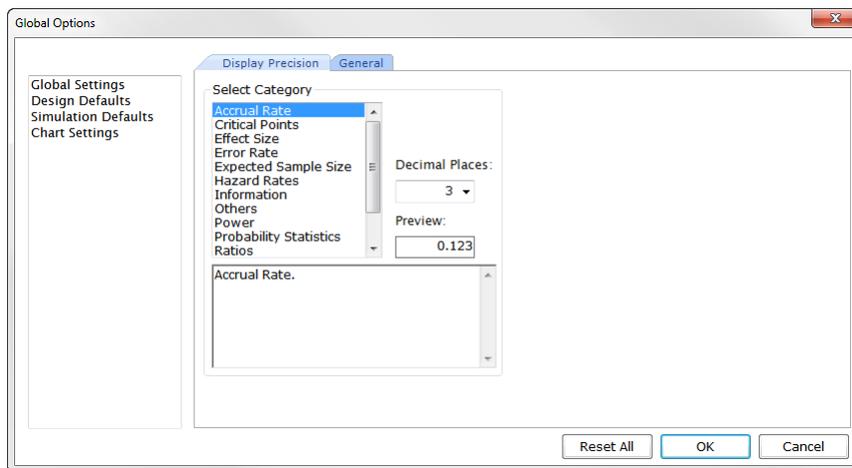
Chapter 67 discusses the Enrollment/Events Prediction At Interim Monitoring Stage using conditional simulations.

Chapter 69 discusses the interaction of East 6 with East PROCs.

58.1 Settings



Click the **Global Options** icon in the **Home** menu to adjust default values in East 6.



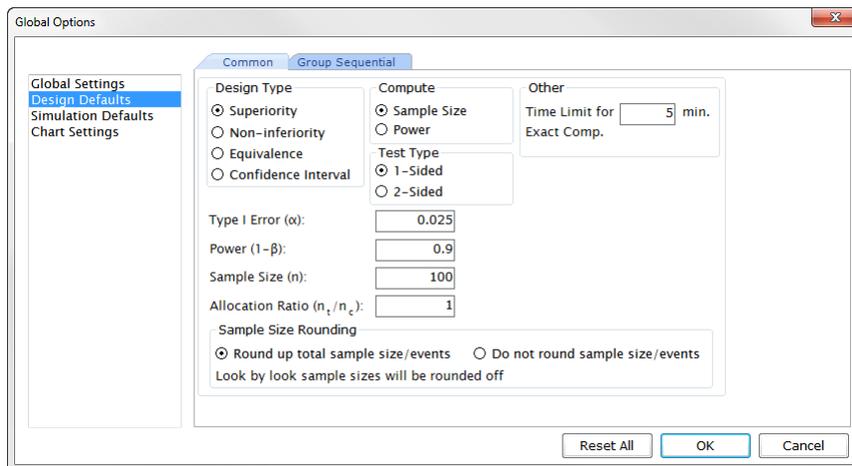
The options provided in the **Display Precision** tab are used to set the decimal places of numerical quantities. The settings indicated here will be applicable to all tests in East 6 under the **Design** and **Analysis** menus.

All these numerical quantities are grouped in different categories depending upon their usage. For example, all the average and expected sample sizes computed at simulation or design stage are grouped together under the category "Expected Sample Size". So to view any of these quantities with greater or lesser precision, select the corresponding category and change the decimal places to any value between 0 to 9.

The **General** tab has the provision of adjusting the paths for storing workbooks, files, and temporary files. These paths will remain throughout the current and future sessions even after East is closed. This is the place where we need to specify the installation directory of the R software in order to use the feature of R Integration in East 6.

58 Introduction to Volume 8

The **Design Defaults** is where the user can change the settings for trial design:

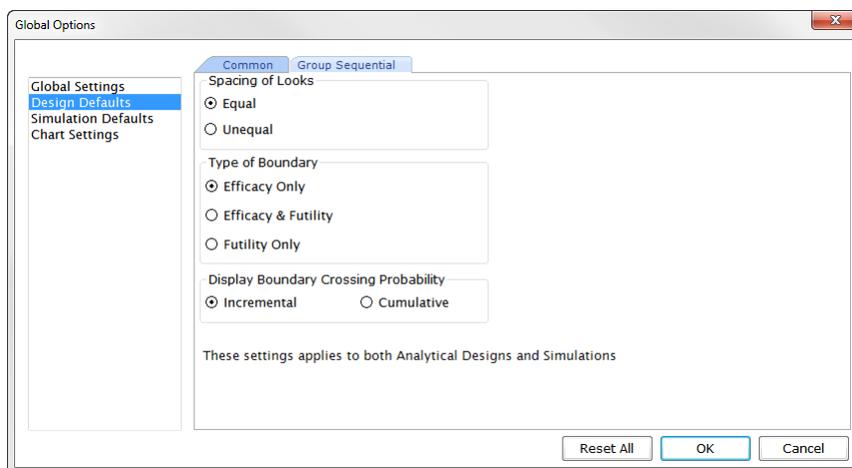


Under the **Common** tab, default values can be set for input design parameters.

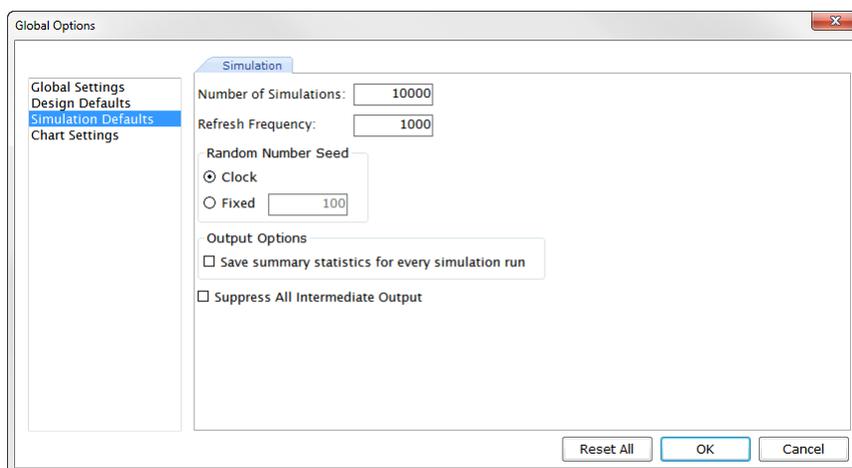
You can set up the default choices for the design type, computation type, test type and the default values for type-I error, power, sample size and allocation ratio. When a new design is invoked, the input window will show these default choices.

- **Time Limit for Exact Computation**
 This time limit is applicable only to exact designs and charts. Exact methods are computationally intensive and can easily consume several hours of computation time if the likely sample sizes are very large. You can set the maximum time available for any exact test in terms of minutes. If the time limit is reached, the test is terminated and no exact results are provided. Minimum and default value is 5 minutes.
- **Type I Error for MCP**
 If user has selected 2-sided test as default in global settings, then any MCP will use half of the alpha from settings as default since MCP is always a 1-sided test.
- **Sample Size Rounding**
 Notice that by default, East displays the integer sample size (events) by rounding up the actual number computed by the East algorithm. In this case, the look-by-look sample size is rounded off to the nearest integer. One can also see the original floating point sample size by selecting the option "Do not round sample size/events".

Under the **Group Sequential** tab, defaults are set for boundary information. When a new design is invoked, input fields will contain these specified defaults. We can also set the option to view the Boundary Crossing Probabilities in the detailed output. It can be either Incremental or Cumulative.



Simulation Defaults is where we can change the settings for simulation:



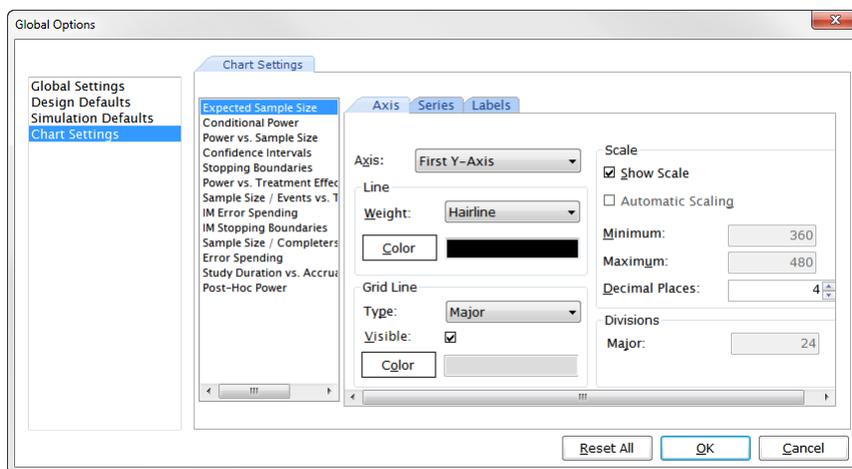
If the checkbox for "Save summary statistics for every simulation" is checked, then East simulations will by default save the per simulation summary data for all the

58 Introduction to Volume 8

simulations in the form of a case data.

If the checkbox for "Suppress All Intermediate Output" is checked, the intermediate simulation output window will be always suppressed and you will be directed to the **Output Preview** area.

The **Chart Settings** allows defaults to be set for the following quantities on East6 charts:



We suggest that you do not alter the defaults until you are quite familiar with the software.

59 *Design and Monitoring of Maximum Information Studies*

This chapter discusses the use of a general tool for designing and monitoring studies on the "information" scale rather than on the "sample size" scale. It is based on the work of Lan and Zucker (1993), Scharfstein, Tsiatis and Robins (1997), Jennison and Turnbull (1997), and Mehta and Tsiatis (2001). It permits a general methodology for group-sequential inference, applicable to any data-generating process with or without covariates. Suppose we wish to detect an effect of magnitude δ with power $1 - \beta$ using a two-sided level- α , K -look group sequential test. The parameter δ may be a binomial probability, a mean from a normal distribution, a difference of two means, a difference of two binomial probabilities, an odds ratio, a hazard ratio, a ratio of Poisson rates, the coefficient of interest in a regression model, or any other univariate "effect size" parameter of interest. The fundamental idea is that no matter what parameter δ we wish to make inferences about, the maximum amount of statistical information, I_{\max} , needed to make the inference is always obtained in the same manner. It is computed by the formula

$$I_{\max} = \left[\frac{z_{\alpha/2} + z_{\beta}}{\delta} \right]^2 \times IF(\alpha, \beta, K, \text{boundary}) \quad (59.1)$$

where $IF(\cdot)$ is an inflation factor that depends on α, β, K and the stopping boundary, but does not depend on δ .

Equation (59.1) tells us, at the design stage, how much information about δ we **need** in order to achieve $1 - \beta$ power. It is applicable in all types of designs, ranging from simple 1-sample normal or binomial designs to more complicated designs based on generalized linear models for discrete categorical or continuous data, parametric survival models, proportional hazard models, mixed effects models, and semi-parametric models for longitudinal data. However, once the trial is underway we need to know how much information about δ has already been accumulated, so as to determine if it is time to terminate the trial. If $\hat{\delta}_j$ is an estimate of δ at the j th interim analysis, the information about δ is estimated by the relationship

$$I_j \approx \text{var}[\hat{\delta}_j]^{-1} \quad (59.2)$$

One could therefore adopt a common design and monitoring strategy for all types of group sequential trials, regardless of the endpoint or the model generating the data.

1. Use equation (59.1) to determine the maximum required information, I_{\max} .
2. At the j -th interim look, estimate I_j , the amount of information currently available about δ , using equation (59.2).
3. If either $I_j \geq I_{\max}$ or the stopping boundary is crossed at information fraction $t_j = I_j/I_{\max}$, terminate the trial. Otherwise continue.

59 Design and Monitoring of Maximum Information Studies

This strategy is appealing both because of its general applicability and because it does not require a priori specification of unknown nuisance parameters. In practice, however, we would be obliged to provide at least an initial estimate of the maximum sample size so that the sponsor of the clinical trial could have some idea of the resources to be committed up-front. For example, suppose that $X_t \sim N(\mu_t, \sigma^2)$, $X_c \sim N(\mu_c, \sigma^2)$ and $\delta = \mu_t - \mu_c$. Then $\hat{\delta}(K) = \bar{X}_t - \bar{X}_c$ and $\text{var}(\hat{\delta}_K) = 4\sigma^2/n_{\max}$, so that finally,

$$n_{\max} = 4\sigma^2 I_{\max} . \tag{59.3}$$

If we were designing the study on the basis of maximum information there would not be any nuisance parameters, whereas if we design the study on the basis of maximum sample size, we would need to know the value of σ^2 . One possibility would be to fix a tentative value for n_{\max} at the design stage, based on our best initial guess at the value of σ^2 . In the previous chapters the group-sequential approach has been utilized exclusively to monitor a study with a view to early stopping. It would seem reasonable, however, to take advantage of the data available at each interim monitoring time-point also to revise our initial estimate of σ^2 and thereby improve the study design adaptively.

Here we will illustrate the procedure with three examples: 1) comparing two binomial distributions where the control response rate is unknown, 2) comparing two normal distributions where the variance is unknown, and 3) comparison of two poisson rates. These examples are intended to demonstrate that the sample size of a study may be revised as data for estimating nuisance parameters become available at the interim monitoring time-points.

59.1 Two Binomials with Unknown Control Response

59.1.1 Information Based Design

59.1.2 Information Based Monitoring

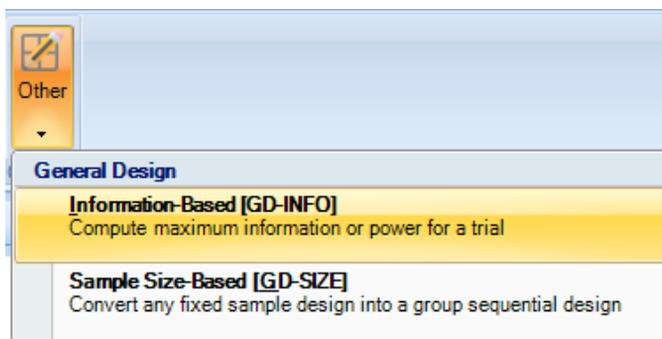
Consider the information based design and monitoring of a randomized clinical trial comparing an experimental therapy with a control therapy based on a dichotomous outcome and equal treatment allocation. Let π_c be the response rate for the control arm, π_t be the response rate for the experimental arm, and $\delta = \pi_t - \pi_c$. We will now design and monitor this study on the information scale.

59.1.1 Information Based Design

Consider a phase III group sequential clinical trial for evaluating the effect of a new drug for prevention of myocardial infarction in patients undergoing coronary artery bypass graft surgery. The study is designed to test the null hypothesis $H_0 : \delta = 0$ against the alternative hypothesis $H_0 : \delta < 0$ using a two sided test at significance level $\alpha = 0.05$. We plan the study to detect a 15% reduction in incidence compared to placebo with 90% power. At the time of designing the study we don't have any reliable estimate of incidence of myocardial infarction in placebo. Therefore, we prefer

information based design that does not rely on the incidence rate of myocardial infarction in placebo.

Single look study Click **Other** on the **Design** tab, and then click **Information Based** as shown below.



A new input window will appear. We will design a study without any interim look. Leave the **Number of Looks** as 1 only. Select **2-Sided** for **Test Type** and enter the values of **Type I Error (α)** and **Power ($1-\beta$)** as 0.05 and 0.9, respectively. Change **Treatment Effect** to -0.15 .

Number of Looks: 1

Test Parameters

Test Type: 2-Sided

Type I Error (α): 0.05 Treatment Effect (δ): -0.15

Power: 0.9

Max. Information: Computed

Click **Compute**. The output is shown as a row in the **Output Preview** located in the

59 Design and Monitoring of Maximum Information Studies

lower pane, with the computed maximum information displayed.

| | ID | No. of Looks | Test Type | Specified α | Power | Max. Information | δ |
|---|------|--------------|-----------|--------------------|-------|------------------|----------|
|  | Des1 | 1 | 2-Sided | 0.05 | 0.9 | 466.996 | -0.15 |

East tells us that the total information required to achieve the operating characteristics of the above study with a fixed sample design is 467 units. This quantity, denoted by I_1 (see Appendix B, Section B.3 for details), was computed by the equation

$$I_1 = \left[\frac{z_{\alpha/2} + z_{\beta}}{\delta_1} \right]^2. \quad (59.4)$$

The subscript ‘1’ indicates that I_1 is the required information for a single look study. Information is approximately equal to the square inverse of the standard error of the estimate of δ . Thus, in a fixed sample trial, the desired power can be achieved if we go on accruing patients until $[\text{se}(\hat{\delta})]^{-2} = 466.996$.

This design has default name Des 1. Save this design in the current workbook by selecting the row corresponding to Des 1 in **Output Preview** and clicking  on the **Output Preview** toolbar.

Multi look study Suppose we actually intend to monitor the study four times. In order to do this, create a new design by selecting Des 1 in the **Library**, and clicking the  icon on the **Library** toolbar. First, change the **Number of Looks** from 1 to 4, to generate a study with three interim looks and a final analysis. Click the **Boundary Info** tab.

Suppose, you have decided to go for a design with 4 interim looks that allows to reject H_0 for efficacy. In order to do this, select **Spending Functions** for **Boundary Family**, **Lan-DeMets** for **Spending Function** and **OF** for **Parameter** in **Efficacy**

box. Select **None** for **Boundary Family** in **Futility** box.

Number of Looks: 4

Test Parameters

Boundary

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.05

Futility

Boundary Family: None

Spacing of Looks: Equal Unequal

Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|--------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.250 | 0.000 | 4.333 | -4.333 |
| 2 | 0.500 | 0.003 | 2.963 | -2.963 |
| 3 | 0.750 | 0.019 | 2.359 | -2.359 |
| 4 | 1.000 | 0.050 | 2.014 | -2.014 |

Click **Compute**. A new row will be added in the **Output Preview**.

| ▲ | ID | No. of Looks | Test Type | Specified α | Power | Max. Information | δ | Spacing of Looks | Efficacy Boundary | Expected Info. (H0) | Expected Info. (H1) |
|---|------|--------------|-----------|-------------|-------|------------------|-------|------------------|-------------------|---------------------|---------------------|
| | Des1 | 1 | 2-Sided | 0.05 | 0.9 | 466.996 | -0.15 | | | | |
| | Des2 | 4 | 2-Sided | 0.05 | 0.9 | 475.533 | -0.15 | Unequal | LD (OF) | 472.874 | 362.996 |

Save this design in the current workbook by selecting the row corresponding to Des 2 in **Output Preview** and clicking on the **Output Preview** toolbar.

East has inflated the maximum information of the single-look study by an appropriate inflation factor to compensate for the power loss of monitoring four times instead of once. The new maximum information, I_K , for a K -look study is shown in Appendix B, Section B.3 to be

$$I_K = I_1 \times \text{IF}(\alpha, \beta, K, \text{boundary})$$

where $\text{IF}(\alpha, \beta, K)$ is an inflation factor that depends on α, β, K and the type of stopping boundary used. The new maximum information is 475.5 units, instead of 467 units. The monitoring strategy for the above sequential trial calls for accruing subjects onto the study until the total information, as measured by $[\text{se}(\hat{\delta})]^{-2}$, equals 475.5 units or until a stopping boundary is crossed, whichever comes first. Now it is difficult to know how long to accrue subjects when the accrual goals are expressed in units of

59 Design and Monitoring of Maximum Information Studies

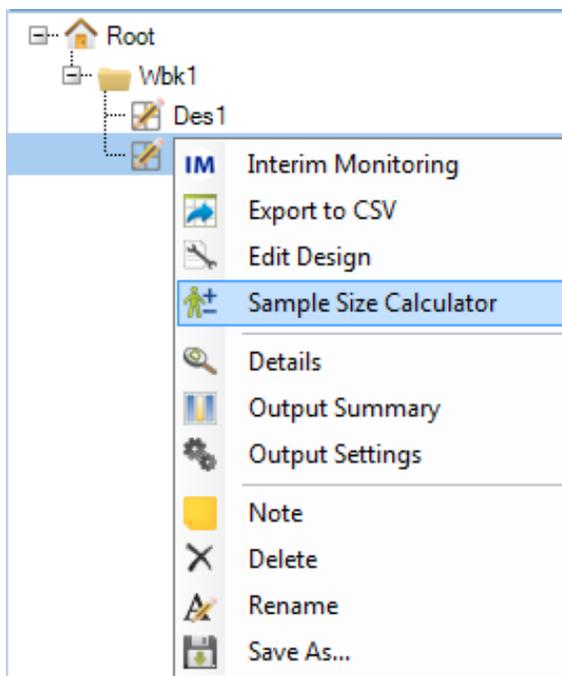
square inverse standard error instead of being expressed in terms of a physical quantity like sample size. We need to translate units of information into sample size units. This is easy to do since the variance of $\hat{\delta}$ is a simple function of the π_c, δ_1 , and the total sample size, n_K . Thus

$$I_K \approx [\text{se}(\hat{\delta})]^{-2} = \left[\frac{(\pi_c)(1 - \pi_c)}{n_K/2} + \frac{(\pi_c + \delta_1)(1 - \pi_c - \delta_1)}{n_K/2} \right]^{-1}.$$

Now since East has already computed $I_K = 475.533$ for $K = 4$, we obtain

$$n_K = 2 \times 475.533 \times [(\pi_c + \delta_1)(1 - (\pi_c + \delta_1)) + (\pi_c)(1 - \pi_c)]. \quad (59.5)$$

East provides you with a convenient sample size calculator for converting the 475.533 units of Fisher information into a sample size, based on equation (38.5). To invoke this calculator, right click on Des2 in the **Library** and select **Sample Size Calculator** from the list.



And select **Difference of Proportions** from the dropdown of *Translate Information From*. The calculator appears as shown below:

The dialog box 'Sample Size Calculator' shows the following settings and results:

- Current Design : Wbk1:Des2
- Translate Information From : Difference of Proportions
- Proportion Control : 0.1

| | Information | Sample Size |
|---------------------|-------------|-------------|
| Maximum : | 475.533 | |
| Expected Under H0 : | 472.874 | |
| Expected Under H1 : | 362.996 | |

Buttons: Recalc, Close

You can alter the control binomial probability in the top cell of the dialog box, and East will compute the corresponding maximum sample size based on maximum Fisher information of 475.533 units. For example, if the baseline response probability is 0.25, the 475.533 units translates into a maximum sample size of 264 subjects (both treatments combined).

The dialog box 'Sample Size Calculator' shows the following settings and results:

- Current Design : Wbk1:Des2
- Translate Information From : Difference of Proportions
- Proportion Control : 0.25

| | Information | Sample Size |
|---------------------|-------------|-------------|
| Maximum : | 475.533 | 264 |
| Expected Under H0 : | 472.874 | 262.524 |
| Expected Under H1 : | 362.996 | 201.505 |

Buttons: Recalc, Close

Based on historical data we assume that the control response rate is 0.3. When you enter 0.3 into the top cell of the dialog box and press the **Recalc** button, East reveals

59 Design and Monitoring of Maximum Information Studies

that the maximum sample size needed for this sequential study is 321.

| | Information | Sample Size |
|---------------------|-------------|-------------|
| Maximum : | 475.533 | 321 |
| Expected Under H0 : | 472.874 | 319.201 |
| Expected Under H1 : | 362.996 | 244.963 |

Thus on the assumption that the control response rate is 0.3, we require an up-front commitment of 321 subjects to meet the operating characteristics of this study. (We can verify this independently by designing a 4-look binomial study using the unpooled

estimate of standard error as shown below. See Section 23.1 for further details.)

| | Des3 |
|--|-------------------|
| Mnemonic | PN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 4 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.15 |
| Proportion under Treatment (π_t) | 0.3 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0.15 |
| Variance | Unpooled Estimate |
| Boundary Parameters | |
| Spacing of Looks | Unequal |
| Efficacy Boundary | LD (OF) |
| Sample Size | |
| Maximum | 321 |
| Expected Under H0 | 319.201 |
| Expected Under H1 | 244.963 |

Of course, if the assumption that the control response rate is 0.30 is incorrect, 321 subjects will not produce the desired operating characteristics. Depending on the actual value of the control response rate, we might have either an under-powered or over-powered study. We shall show in the next section that one of the major advantages of the information based approach is that we can use all the data accrued at any interim monitoring time point to re-estimate the control response rate and, if it differs from what was assumed initially, re calculate the sample size.

59.1.2 Information Based Monitoring

Select Des 2 in the **Library**, and click **IM** from the **Library** toolbar. This will open a interim monitoring dashboard.

If we monitor the data at any chronological time τ , an efficient estimator of δ is $\hat{\delta}(\tau) = \hat{\pi}_t(\tau) - \hat{\pi}_c(\tau)$ and the standard error of this estimator is

$$se(\hat{\delta}(\tau)) = [(\hat{\pi}_t(\tau))(1 - \hat{\pi}_t(\tau))/n_1(\tau) + (\hat{\pi}_c(\tau))(1 - \hat{\pi}_c(\tau))/n_2(\tau)]^{1/2},$$

59 Design and Monitoring of Maximum Information Studies

where $\hat{\pi}_i(\tau)$ is the sample proportion responding to treatment i among the $n_i(\tau)$ individuals assigned to treatment i by time τ , $i = 1, 2$. The information accrued at this time point is

$$I(\tau) = (\text{se}(\hat{\delta}(\tau)))^{-2}$$

and the value of the Wald test statistic is

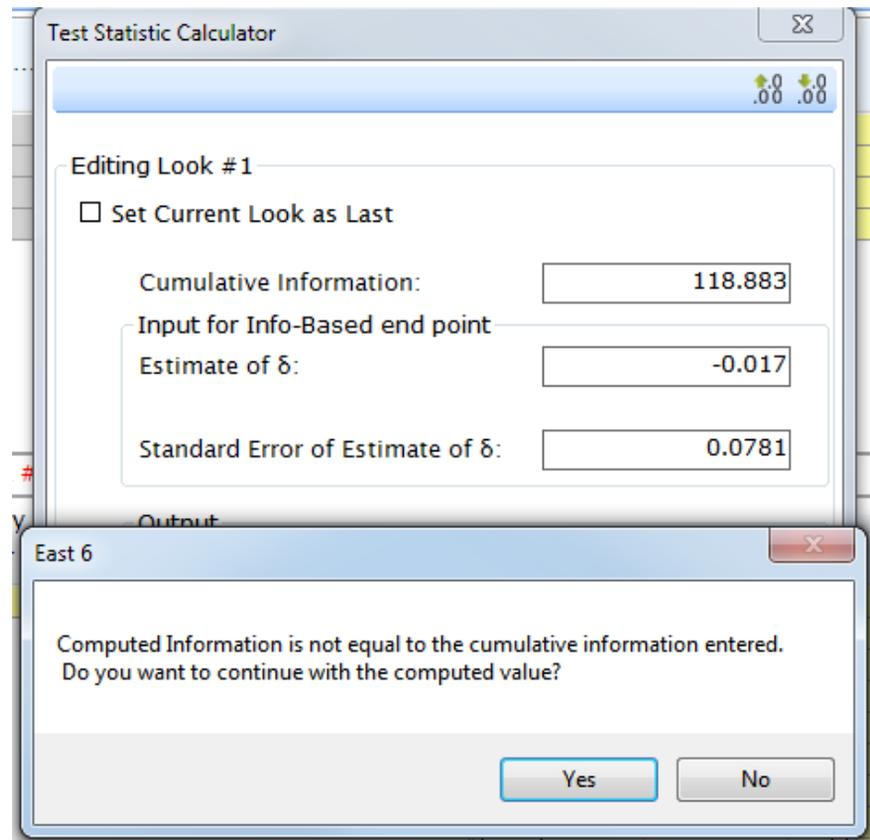
$$T(\tau) = \hat{\delta}(\tau)/\text{se}(\hat{\delta}(\tau)) .$$

The information fraction at chronological time τ is $t(\tau) = I(\tau)/475.5327$. We will stop the study if the test statistic crosses the **LD (OF)** stopping boundary at this information fraction. For future reference, we will also refer to the information fraction as “process time”. In contrast, the time τ will also be referred to as “calendar time”.

Results at the First Interim Monitoring Time Point

Suppose that at the first interim monitoring time point, τ_1 , we observe 15/60 responders on placebo and 14/60 responders on treatment. Then $\hat{\delta}(\tau_1) = -0.017$, $\text{se}(\hat{\delta}(\tau_1)) = 0.0781$. To pass these values to East, click **Enter Interim Data** from the toolbar to invoke the **Test Statistic Calculator**. Enter the information above, and click

Recalc:



Saying ‘Yes’ to this message will update that the current information to 163.945 units and the current value of the test statistic to -0.218. Now click **OK** to continue.

East displays the information fraction, $t(\tau_1) = 163.945/475.533 = 0.345$, and computes the appropriate stopping boundary at that process time. The value of the stopping boundary is ± 3.643 . Since our test statistic did not exceed this boundary, we

59 Design and Monitoring of Maximum Information Studies

continue to the next interim monitoring time point.

| Look # | Information Fraction | Cumulative Info. | Test Statistic | Effect... | Standard Error | Efficacy | | 95% RCI for Effect Si... | | Repeat... p-value | CP | Predicti... Power |
|--------|----------------------|------------------|----------------|-----------|----------------|----------|--------|--------------------------|--------|-------------------|-------|-------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.345 | 163.945 | -0.218 | -0.017 | 0.078 | 3.643 | -3.643 | 0.267 | -0.301 | 1 | 0.024 | 0.16 |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | |

We have accrued 120 subjects out of the 321 required under the design assumption that the nuisance parameter is $\pi_c = 0.30$. The information fraction under this design assumption is thus $120/321 = 0.374$, while the actual information fraction is 0.345. Thus the information appears to be coming in a little slower than anticipated, but this difference does not seem serious enough to alter the sample size requirements of the study.

Results at the Second Interim Monitoring Time Point

Suppose that at the second interim monitoring time point, τ_2 , we observe 29/120 responders on treatment and 41/120 responders on placebo. Therefore, the estimate of $\hat{\delta}$ is -0.1 with standard error as 0.058. Click [Enter Interim Data](#) to bring up **Test Statistic Calculator**. Enter -0.1 for **Estimate of δ** and 0.058 for **Standard Error of**

Estimate of δ .

Click **Recalc**, and then click **Yes**. The information accrued at this time point is 297.265 and the observed value of the test statistic is $T(\tau_2) = -1.724$. Upon pressing the **OK** button, these values are pasted into the interim monitoring dashboard.

The information fraction is 0.625. The required stopping boundary is 2.609. Since the absolute value of test statistic is smaller than 2.609, the stopping boundary is not crossed and, once more, the study continues.

This time the anticipated information fraction under the assumption that $\pi_c = 0.30$ is $240/321 = 0.748$, which is considerably larger than the actual information fraction 0.625. Thus, there is considerable evidence that the information is coming in slower than anticipated. In fact, the data suggest that the value of π_c is close to 0.34, as the

59 Design and Monitoring of Maximum Information Studies

estimate at the first look $14/60 = 0.238$ and the estimate at the second look is $29/120 = 0.242$. It might therefore be prudent to re-estimate the sample size of the study. The new maximum sample size can be obtained by the relationship

$$\frac{n(\tau_2)}{n_{\max}} = \frac{I(\tau_2)}{I_{\max}}.$$

Thus the maximum sample size (rounded up to the nearest integer) is

$$n_{\max} = n(\tau_2) \times \frac{I_{\max}}{I(\tau_2)} = 240 \times \frac{475.533}{297.265} = 389.$$

Therefore we need to commit 389 subjects to the study, not 321 as originally estimated. We see that the original design with 321 subjects would have led to a seriously under-powered study.

Results at the Third Interim Monitoring Time-Point

We continue to accrue subjects beyond the 321 in the original design, and reach the third interim monitoring time point at time τ_3 with 61/180 responders on placebo and 41/180 responders on treatment. Therefore, the estimate of $\hat{\delta}$ is -0.111 with standard error as 0.047. Click on the **Enter Interim Data** icon, and enter -0.111 for

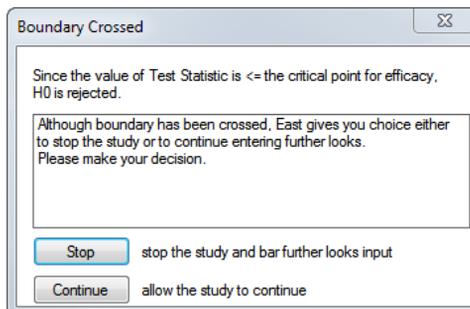
Estimate of δ and 0.047 for Standard Error of Estimate of δ .

The screenshot shows a dialog box titled "Test Statistic Calculator". At the top right, there are two small green arrows and the number ".00". Below the title bar, the text "Editing Look #3" is displayed. A checkbox labeled "Set Current Look as Last" is present and unchecked. The "Input for Info-Based end point" section contains three input fields: "Cumulative Information:" with the value 452.694, "Estimate of δ :" with the value -0.111, and "Standard Error of Estimate of δ :" with the value 0.047. The "Output" section contains one input field: "Test Statistic:" with the value -2.362. At the bottom of the dialog, there are three buttons: "Recalc", "OK", and "Cancel". The "OK" button is highlighted with a dashed border.

Click **Recalc**, and then click **OK**. The information accrued at this time point is 452.69 and the observed value of the test statistic is $T(\tau_2) = -2.362$. Now click **OK** to update the charts and tables in the dashboard. Now the stopping boundary is crossed

59 Design and Monitoring of Maximum Information Studies

and the following dialog box appears.



Click **Stop**. At this look, the total information accrued is 452.694 and the observed value of the test statistic is $T(\tau_3) = -2.362$. Since the absolute value of the test statistic exceeds the corresponding stopping boundary, 2.05, the stopping boundary is crossed and the study terminates with a statistically significant outcome.

You see the IM sheet results as shown below.

| Look # | Information Fraction | Cumulative Info. | Test Statistic | Effect... | Standard Error | Efficacy | | 95% RCI for Effect Si... | | Repeat... p-value | CP | Predicti... Power |
|--------|----------------------|------------------|----------------|-----------|----------------|----------|--------|--------------------------|--------|-------------------|-------|-------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.345 | 163.945 | -0.218 | -0.017 | 0.078 | 3.643 | -3.643 | 0.267 | -0.301 | 1 | 0.024 | 0.16 |
| 2 | 0.625 | 297.265 | -1.724 | -0.1 | 0.058 | 2.609 | -2.609 | 0.051 | -0.251 | 0.224 | 0.613 | 0.59 |
| 3 | 0.952 | 452.694 | -2.362 | -0.111 | 0.047 | 2.05 | -2.05 | -0.015 | -0.207 | 0.023 | NA | NA |

Click the "Edit Interim Data" button to edit the Look # 3 data.

Stopping Boundaries

| Info. | Efficacy Upper | Efficacy Lower |
|---------|----------------|----------------|
| 163.945 | 3.643 | -3.643 |
| 297.265 | 2.609 | -2.609 |
| 452.694 | 2.05 | -2.05 |

Final Inference

| Final Outputs at Look # | | 3 |
|-------------------------------|--|--------|
| Adj. p-value | | 0.023 |
| Adj. Pt. Est. for Effect Size | | -0.109 |
| Adj. 95% CI for Effect Size | | |
| Upper Confidence Bound | | -0.015 |
| Lower Confidence Bound | | -0.202 |
| Post-Hoc Power | | |

Error Spending Function

| Info. Fraction | α |
|----------------|-------|
| 0.345 | 0 |
| 0.625 | 0.009 |
| 0.952 | 0.043 |

Confidence Intervals

| Info. Fraction | RCI Upper | RCI Lower | Naive CI Upper | Naive CI Lower | Adj Up |
|----------------|-----------|-----------|----------------|----------------|--------|
| 0.345 | 0.267 | -0.301 | 0.136 | -0.17 | |
| 0.625 | 0.051 | -0.251 | 0.014 | -0.214 | |
| 0.952 | -0.015 | -0.207 | -0.019 | -0.203 | -0. |

The adjusted p-value is 0.023, with a final adjusted estimate of the difference of

−0.109.

This example highlights the fundamental difference between information based sequential monitoring and conventional sequential monitoring. Had the study been monitored by the conventional method, the maximum sample size would have been fixed from the start at 321 subjects and there would have been no flexibility to change the level of this physical resource over the course of the study. But in an information based approach the maximum information is fixed, not the maximum amount of a physical resource. Thus, the maximum sample size could be altered over the course of the study from 321 subjects to 389 subjects, while the maximum information stayed constant. Without this flexibility, the power of the study would be severely compromised.

59.2 Two Normals with Unknown Variance

59.2.1 Info Based Design

59.2.2 Info Based Monitoring

In this section we will consider the PRIMO study (Pritchett et al., 2011). This was a multinational, multicenter randomized controlled trial to assess the effects of paricalcitol (a selective vitamin D receptor activator) on mild to moderate left ventricular hypertrophy in patients with chronic kidney disease. The primary endpoint was change in left ventricular mass (LVM) index. Let μ_t and μ_c be the change in LVM index in paricalcitol and placebo, respectively. $\delta = \mu_t - \mu_c$ denotes the difference in change in LVM index in paricalcitol compared to placebo. We want to test the hypothesis $H_0 : \delta = 0$ against $H_0 : \delta < 0$. A mean difference of 2.7g/m in LVM index change was considered clinically meaningful. Therefore, we will design a study to detect $\delta_1 = -2.7$ with 90% power.

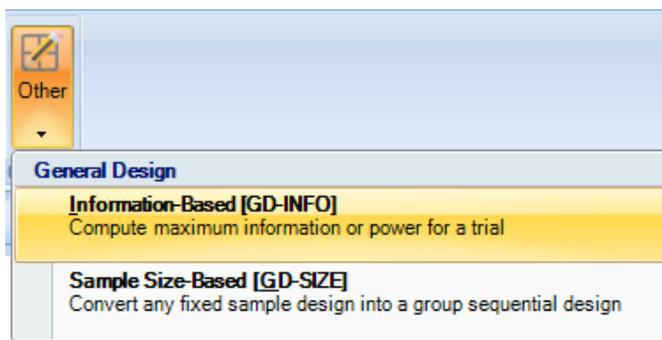
59.2.1 Information Based Design

There is no reliable estimate available for the standard deviation (σ). Therefore, an information based design that does not rely on the standard deviation would be preferable in this case. An unblinded interim analysis was conducted for early termination and to make an informative decision with respect to sample size adjustment. Interim analysis was planned when 90% of subjects are enrolled.

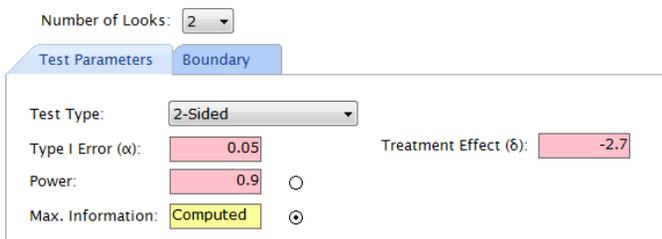
First, click **Other** on the **Design** tab, and then click **Information Based** as shown

59 Design and Monitoring of Maximum Information Studies

below.



Change the **Number of Looks** to 2. This will add a tab with label **Boundary Info**. We will come back to this tab later. In the **Design Parameters** tab, select **Test Type** as **2-Sided** and enter **Type I Error (α)** and **Power ($1-\beta$)** as 0.05 and 0.9, respectively. Change **Effect Size** to -2.7 . The **Design Parameters** tab should appear as below:



Now click **Boundary Info**. Select **Spending Functions** for **Boundary Family**, **Gamma Family** for **Spending Function** and -8 for **Parameter (γ)** in the **Efficacy** box. In the **Futility** box, select **None** for **Boundary Family**. Since we want to have a interim look at 90% of sample size, specify 0.9 for 'Info. Fraction at Interim Look'.

The **Boundary Info** tab should appear as below.

Number of Looks: 2

Test Parameters
Boundary

Efficacy

Boundary Family: Spending Functions

Spending Function: Gamma Family

Parameter (γ): -8

Type I Error (α): 0.05

Futility

Boundary Family: None

Info. Fraction at Interim Look: 0.9 Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|--------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.900 | 0.022 | 2.283 | -2.283 |
| 2 | 1.000 | 0.050 | 1.981 | -1.981 |

Click **Compute**. The output is shown as a row in the **Output Preview** located in the lower pane. The computed maximum information is highlighted in yellow.

| ▲ | ID | No. of Looks | Test Type | Specified α | Power | Info. Fraction at Interim Look | Efficacy Boundary | Max. Information | Expected Info. (H0) | Expected Info. (H1) | δ |
|---|------|--------------|-----------|-------------|-------|--------------------------------|-------------------|------------------|---------------------|---------------------|------|
| | Des1 | 2 | 2-Sided | 0.05 | 0.9 | 0.9 | Gm (-8) | 1.448 | 1.445 | 1.334 | -2.7 |

East tells us that the total information required to achieve the operating characteristics of the above study is 1.448 units. The monitoring strategy for the above 2-look sequential trial calls for accruing subjects onto the study until the total information, as measured by $[se(\hat{\delta})]^{-2}$, equals 1.448 units or until a stopping boundary is crossed, whichever comes first. Now we can translate this information into sample size using the following relationship:

$$n_{\max} = 4\sigma^2 I_{\max} .$$

In the PRIMO study the initial estimate of σ is assumed as 6.39. Save this design in the current workbook by selecting the row corresponding to Des 1 in **Output Preview** and clicking on the **Output Preview** toolbar. Right click on the design node to invoke the **Sample Size Calculator**. Plug in this value in the calculator, the 1.448 units

59 Design and Monitoring of Maximum Information Studies

translates into a maximum sample size of 237 (total sample size).

| | Information | Sample Size |
|---------------------|-------------|-------------|
| Maximum : | 1.448 | 237 |
| Expected Under H0 : | 1.445 | 236.466 |
| Expected Under H1 : | 1.334 | 218.106 |

This calculator should only be used for Difference of Means and Difference of Proportions (Unpooled variance) Superiority designs.

If we design the study for a maximum sample size of 237 patients, we will achieve 90% power so long as our estimate of σ , 6.39, is correct. On the other hand we gain more flexibility by designing the study for maximum information of 1.448 units. This design parameter remains the same whether the standard deviation is 6.39 or something different. As the data accumulate during the interim monitoring phase, we will obtain more accurate estimates of the standard deviation and can revise the sample size on that basis. We shall show in the next section that one of the major advantages of the information based approach is that we can use all the data accrued at any interim monitoring time point to re-estimate the σ and, if it differs from what was assumed initially, re-calculate the sample size.

59.2.2 Information Based Monitoring

We will monitor the study on the information scale. Select Des 1 in the **Library**, and click **IM** from the **Library** toolbar. This will open an interim monitoring dashboard.

Results at the First Interim Monitoring Time-Point

Recall that the study is planned to have an interim look when 90% of sample size are accrued. Therefore a interim look can be planned when 237×0.9 or 214 subjects are evaluated. Suppose that at the first interim monitoring time-point, there were 107

subjects on the placebo arm, and 107 subjects on the treatment arm, $\delta = -2.85$, $s_c = 7.5$ and $s_t = 7.4$. Based on the sample standard deviation, the pooled standard deviation is 7.45 and $[se(\hat{\delta})] = 1.019$.

Click on the **Enter Interim Data** icon to invoke the **Test Statistic Calculator**. Enter -2.85 for **Estimate of δ** and 1.019 for **Standard Error of Estimate of δ** . Click **Recalc**, and then click **Yes**. The information accrued at this time point is 0.963 and the observed value of the test statistic is $T(\tau_1) = -2.797$. Pres OK to update the IM dashboard.

| Look # | Information Fraction | Cumulative Info. | Test Statistic | Effect... | Standard Error | Efficacy | | 95% RCI for Effect Si... | | Repeat... p-value | CP | Predicti... Power |
|--------|----------------------|------------------|----------------|-----------|----------------|----------|--------|--------------------------|--------|-------------------|-------|-------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.665 | 0.963 | -2.797 | -2.85 | 1.019 | 2.927 | -2.927 | 0.133 | -5.833 | 0.076 | 0.995 | 0.983 |
| 2 | | | | | | | | | | | | |

The information fraction is 0.665 . The required stopping boundary is 2.927 . Since the absolute value of test statistic is smaller than 2.927 , the stopping boundary is not crossed and the study continues.

This time the anticipated information fraction under the assumption that $\sigma = 6.4$ is $214/237 = 0.903$, which is considerably larger than the actual information fraction 0.665 . Thus, there is considerable evidence that the information is coming in slower than anticipated. In fact, the data suggest that the value of σ is close to 7.45 . It might therefore be prudent to re-estimate the sample size of the study. The new maximum sample size can be obtained by the relationship

$$\frac{n(\tau_1)}{n_{\max}} = \frac{I(\tau_1)}{I_{\max}}.$$

Thus the maximum sample size (rounded up to the nearest integer) is

$$n_{\max} = n(\tau_1) \times \frac{I_{\max}}{I(\tau_1)} = 214 \times \frac{0.903}{0.665} = 291.$$

Therefore we need to commit 291 subjects to the study, not 237 as originally estimated. Thus it is clear that unless we increase patient accrual from the initial specification of 237, we will have a seriously underpowered study. Let us assume then that the investigators agree at this stage to increase the sample size to 291 patients.

Results at the Final Look

Suppose that at the final look we have accrued 291 patients of which 145 are allocated to placebo and 146 are allocated to new drug paricalcitol. Based on these subjects,

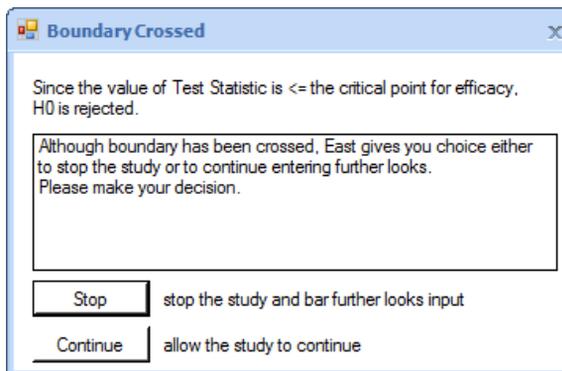
59 Design and Monitoring of Maximum Information Studies

$\delta = -2.93$, $s_c = 7.43$ and $s_t = 7.41$. Thus, the pooled standard deviation is 7.42 and $[se(\hat{\delta})] = 0.870$. Click on the **Enter Interim Data** icon. In the **Test Statistic Calculator**, tick the checkbox of **Set Current Look as Last**. Enter -2.93 for **Estimate of δ** and 0.870 for **Standard Error of Estimate of δ** . Click **Recalc**, and then click **Yes**. The information accrued at this time point is 1.321 and the observed value of the test statistic is $T(\tau_2) = -3.368$. The **Test Statistic Calculator** should look as below.

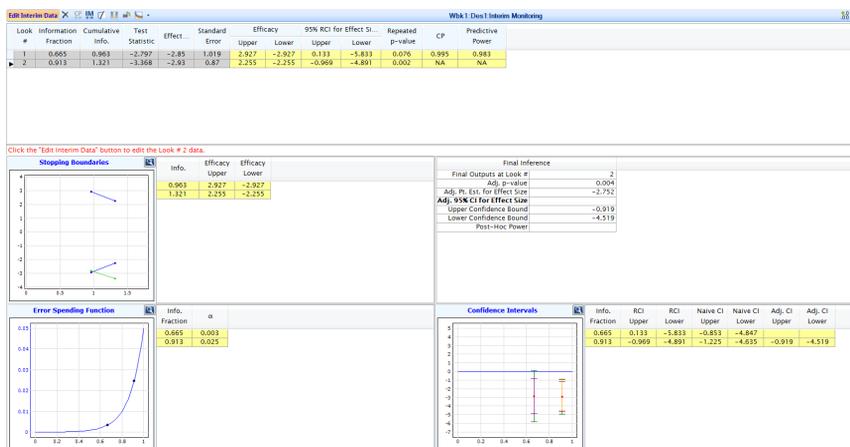
The screenshot shows a dialog box titled "Test Statistic Calculator". At the top right, there are zoom in (+) and zoom out (-) icons. The main content area is titled "Editing Look #2" and contains a checkbox labeled "Set Current Look as Last" which is currently unchecked. Below this are three input fields: "Cumulative Information:" with the value 1.321, "Estimate of δ :" with the value -2.93, and "Standard Error of Estimate of δ :" with the value 0.87. An "Output" section at the bottom shows "Test Statistic:" with the value -3.368 highlighted in yellow. At the very bottom of the dialog are three buttons: "Recalc", "OK", and "Cancel".

Upon pressing the **OK** button a pop-up window will appear notifying you that H_0 is

rejected as the test statistic exceeds the critical boundary.



Click **Stop**. This time East tells us that the stopping boundary of 1.962 has been crossed and the study terminates with the conclusion that the paricalcitol does indeed lower the change in LVM index relative to the placebo.



The adjusted estimate of the difference is -2.752 and the adjusted p-value is 0.004. The 95% adjusted confidence interval for the reduction is [-4.519, -0.919].

59 Design and Monitoring of Maximum Information Studies

59.3 Equality of Two Poisson Rates

59.3.1 Trial Design

59.3.2 Interim Monitoring

We will use an information-based approach to design a stroke prevention study that was previously discussed in detail in Chapter 60, Section 60.1.1. The goal is to design a balanced two arm randomized clinical trial for high risk patients with atrial fibrillation in which the standard treatment (adjusted dose warfarin) has a Poisson event rate of 1.8% per year (i.e., 1.8 ischaemic stroke events per 100 people per year). If the experimental treatment (low-dose warfarin plus aspirin) has a Poisson event rate in excess of 3% per year, we wish to detect this with 90% power using a one sided test conducted at the 5% level of significance. Let λ_c and λ_t denote the Poisson event rates for the control and treatment arms, respectively, and define the risk ratio

$$\gamma = \frac{\lambda_t}{\lambda_c}.$$

59.3.1 Trial Design

We wish to test the null hypothesis that $\gamma = 1$ against the one-sided alternative hypothesis that $\gamma > 1$ using a test at significance at level $\alpha = 0.05$. The test is required to have power $1 - \beta = 0.9$ at the alternative $\gamma = 3/1.8 = 1.667$. In Section 60.1.1, we designed and monitored this study using traditional large-sample methods of unconditional inference. In the present section, we will use an alternative conditional method of inference for comparison purposes. Although there have been no formal studies comparing the conditional and unconditional approaches for Poisson data it is generally believed that the conditional approach has greater accuracy. For example, Breslow and Day (1987) utilize the conditional approach in their monograph on cohort studies.

Suppose that X_c is the number of events observed on the control arm, X_t is the number of events observed on the treatment arm and $N = X_c + X_t$. Then it is well known that the conditional distribution of X_t given N is binomial with parameters (π, N) where

$$\pi = \frac{n_t \gamma}{n_c + n_t \gamma} \tag{59.6}$$

and n_c is the number of person years of follow-up on the control arm and n_t is the number of person years of follow-up on the treatment arm. The present study was designed for equal amounts of follow-up on each arm. Thus, at the design stage we may assume that $n_c = n_t$. The protocol specifies that $\gamma = 1$ under the null hypothesis, and $\gamma = 1.667$ under the alternative hypothesis. Therefore, by equation 59.6, the null and alternative hypotheses may be stated as:

$$H_0: \pi = 0.5 \text{ versus } H_1: \pi = 0.625 .$$

The design has now been formulated in terms of testing the mean of a binomial random variable.

Hence with N , the total number

$$\delta = \pi - 0.5$$

playing the role of effect size. The null and alternative hypotheses can now be specified in terms of δ as

$$H_0: \delta = 0 \text{ versus } H_1: \delta = 0.125 .$$

The maximum value of N for a K look group sequential design is thus

$$N_{\max} = \pi(1 - \pi)I_{\max} \tag{59.7}$$

where I_{\max} is computed by equation (59.1) and can be obtained from East.

Click **Other** on the **Design** tab, and then click **Information Based**. In the ensuing input dialog box, in the **Design Parameters** tab, select **1-Sided** for **Test Type**. Specify **Type I Error (α)** as 0.05, and **Power ($1-\beta$)** as 0.9, respectively. Change **Treatment Effect** to 0.125.

Number of Looks: 1

Test Parameters

Test Type: 1-Sided

Type I Error (α): 0.05 Treatment Effect (δ): 0.125

Power: 0.9

Max. Information: Computed

Click **Compute**. The output is shown as a row in the **Output Preview** located in the lower pane.

| ▲ | ID | No. of Looks | Test Type | Specified α | Power | Max. Information | δ |
|----|------|--------------|-----------|--------------------|-------|------------------|----------|
| 🗑️ | Des1 | 1 | 1-Sided | 0.05 | 0.9 | 548.086 | 0.125 |

59 Design and Monitoring of Maximum Information Studies

This design has default name Des 1. Save this design in the current workbook by selecting the row corresponding to Des 1 in **Output Preview** and clicking  on the **Output Preview** toolbar. For Des 1 $I_{\max} = 548.09$. Equation (59.7) converts I_{\max} to N_{\max} . In applying this equation, we must specify the value of π at which 90% power is desired. With $\pi = 0.625$, we have $N_{\max} = 0.625 \times 0.375 \times 548.09 = 128$ events. This is somewhat lower than the 135 events computed in the general design of Chapter 60, and suggests that the conditional approach used here is more efficient than the unconditional approach.

Suppose we wish to take two interim looks and a final look at the accruing data and utilize the usual Lan and DeMets (1983) α -spending function **LD (OF)**. Now create a new design by right-clicking Des 1 in the **Library**, and edit it by clicking  icon. Change the **Number of Looks** from 1 to 3. In the **Boundary Info** tab, select **Spending Functions** for **Boundary Family**, **Lan-DeMets** for **Spending Function** and **OF** for **Parameter** in the **Efficacy** box. In the **Futility** box, select **None** for **Boundary Family**.

Click **Compute** to generate output for this design. A new row will be added in the **Output Preview**.

| ID | No. of Looks | Test Type | Specified α | Power | Max. Information | δ | Spacing of Looks | Efficacy Boundary | Expected Info. (H0) | Expected Info. (H1) |
|------|--------------|-----------|--------------------|-------|------------------|----------|------------------|-------------------|---------------------|---------------------|
| Des1 | 1 | 1-Sided | 0.05 | 0.9 | 548.086 | 0.125 | | | | |
| Des2 | 3 | 1-Sided | 0.05 | 0.9 | 558.362 | 0.125 | Unequal | LD (OF) | 555.187 | 432.696 |

The maximum information is inflated to $I_{\max} = 558.36$ and the corresponding maximum number of events is inflated to $N_{\max} = 131$. Save this design in the current workbook by selecting the row corresponding to Des 2 in **Output Preview** and clicking  on the **Output Preview** toolbar.

Observe that although the maximum information is slightly inflated, the expected information under H_1 is only 432.696. If H_1 is true then $\pi = 0.625$ so that the corresponding expected number of events is $0.625 \times (.375) \times 432.696 = 101$, a considerable saving over the single look design.

59.3.2 Interim Monitoring

Let us monitor this study using the interim monitoring data published in JAMA (vol 279, No. 16, Table 2). According to this report, the study was monitored after $N = 55$ events were observed. There were $X_c = 11$ events on the control arm over $n_c = 581$

person years of observation. And there were $X_t = 44$ events on the treatment arm over $n_t = 558$ person years of observation. We can estimate γ from the data as

$$\hat{\gamma} = \frac{X_t/n_t}{X_c/n_c} = \frac{44/558}{11/581} = 4.1649$$

whereupon the estimate of π is

$$\hat{\pi} = \frac{558 \times 4.1649}{581 + 558 \times 4.1649} = 0.8,$$

the estimate of effect size is

$$\hat{\delta} = 0.8 - 0.5 = 0.3.$$

and its standard error is

$$\text{se}(\hat{\delta}) = \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{N}} = 0.054.$$

The current information is thus $I = 0.053936^{-2} = 343.75$. We enter this value into the interim monitoring worksheet as described below.

Select Des 2 in the **Library**, and click **IM** from the **Library** toolbar. This will open a interim monitoring dashboard. Click on the **Enter Interim Data** icon to invoke the **Test Statistic Calculator**. Enter 0.3 for **Estimate of δ** and 0.054 for **Standard Error of Estimate of δ** . Click **Recalc**, and then click **Yes**. The information accrued at this

59 Design and Monitoring of Maximum Information Studies

time point is 342.936 and the observed value of the test statistic is $T(\tau_1) = 5.556$.

Finally, click **OK** to paste this information in the monitoring dashboard. Now, the stopping boundary is crossed, and a dialog box appears. Click **Stop**. Since the test statistic, $Z = 0.3/0.054 = 5.556$ exceeds the upper stopping boundary, the trial is terminated. A table for **Final Inference** will appear in the dashboard.

| Final Inference | |
|------------------------------------|-------|
| Final Outputs at Look # | 1 |
| Adj. p-value | 0 |
| Adj. Pt. Est. for Effect Size | 0.3 |
| Adj. 90% CI for Effect Size | |
| Upper Confidence Bound | 0.389 |
| Lower Confidence Bound | 0.211 |
| Post-Hoc Power | |

The lower confidence bound of the adjusted confidence interval for δ is 0.211 implying that π is at least $0.5 + 0.211 = 0.711$ with 95% confidence. Thus the risk ratio γ is

estimated to be at least $\pi/(1 - \pi) = 0.711/0.289 = 2.46$. The risk of stroke is at least 2.46 times greater on the treatment arm than on the control arm. If the event rate on the control arm is 1.8% per year, then the corresponding event rate on the treatment arm is at least 2.46×1.8 , or 4.428% per year.

59.4 Some Non-Statistical Concerns

This chapter demonstrated the monitoring clinical trials which is statistically sound and ensures that the trials will be adequately powered despite inaccurate initial estimates of nuisance parameters that crucially affect the sample size. Provided we are prepared to remain flexible about the final sample size, we can learn as we go, and make appropriate sample size adjustments along the way. The pay-off for adopting this approach is high, both ethically and economically. Many industry trials are over-powered in order to compensate for ignorance about the variability of the data, thereby raising the cost of the trial unnecessarily. Some trials are underpowered because of overly optimistic initial estimates of variability. A promising new therapy might remain undetected despite incurring the high cost running the trial. The information-based approach ensures that we will neither randomize too many subjects nor too few subjects, but just the right number to meet the goals of the trial.

A number of factors, unrelated to the statistical methodology, will determine whether or not this idea is adopted in practice. Here is a list of unresolved issues that must be addressed:

- The time between the intermediate data base lock and the performance of the interim analysis must be shortened, so as to minimize the number of patients being enrolled while the decision to continue or terminate accrual is being made.
- Institutional Review Boards must be educated on the benefits of these trials. They need to understand that an information-based design with a flexible sample size is, in some situations, more ethical than a design that fixes the sample size up-front, despite considerable uncertainty about its adequacy to achieve the desired power.
- When the sample size is a random variable, the sponsor may face logistical challenges related to ensuring that sites have sufficient quantities of the drugs or biologics on hand.
- The sponsor will have to re-think the manner in which the budget is prepared for a trial. Rather than having a fixed budget for each individual trial, it might be necessary to envisage a fixed overall budget for a portfolio of trials which can be allocated to the individual trials in a flexible manner.
- These information-based trials might be subject to additional regulatory scrutiny. The burden will be on the sponsor to demonstrate that the statistical methodology is sound and, by the manner in which the trial was conducted, the

59 *Design and Monitoring of Maximum Information Studies*

interim results were not prematurely unblinded.

60 *Design and Interim Monitoring with General Endpoints*

In the previous chapters, we have shown how to use East to design and monitor group-sequential studies with normal, binomial and survival endpoints. In this chapter, we show how to extend East to design and monitor studies with any general endpoint, including longitudinal studies, equivalence studies, and studies where the endpoint is specified as one of the covariates in a generalized linear regression model. In all these settings, we use East in conjunction with some other design package that is capable of computing the sample-size for the end-point in question when there is no interim monitoring. The fixed sample-size thus obtained is then used as an input to the General Design module provided by East. East inflates this fixed sample-size appropriately based on the planned number of interim analyses, the type of stopping boundary, the desired type-1 error and the desired power. The derivation of the appropriate inflation factor for this purpose is discussed in Appendix B, Section B.3. The resulting group-sequential design may then be monitored flexibly using East's interim monitoring dashboard. We illustrate below with an example involving Poisson data.

60.1 *Poisson Model*

For Stroke Prevention in Atrial Fibrillation, investigators conducted a two arm randomized clinical trial of adjusted-dose warfarin versus low-intensity fixed-dose warfarin plus aspirin for high-risk patients with atrial fibrillation (AF). (See, Lancet, 1996, 348(9028):633-8, for details.) Adjusted-dose warfarin is known to be highly efficacious for prevention of ischaemic stroke in AF patients, with an event rate of only 1.8% per year. This treatment, however, carries a risk of bleeding and requires frequent monitoring. The objective of the study was to determine if low-intensity fixed-dose warfarin plus aspirin, which is safer and easier to administer, might be substituted for adjusted-dose warfarin without resulting in an unacceptably high relative risk of stroke. An event rate in excess of 3% per year with the low-intensity warfarin would be considered unacceptable. We will use East to design and monitor a group-sequential study with two interim looks and one final look.

60.1.1 *Design of Stroke Prevention Study*

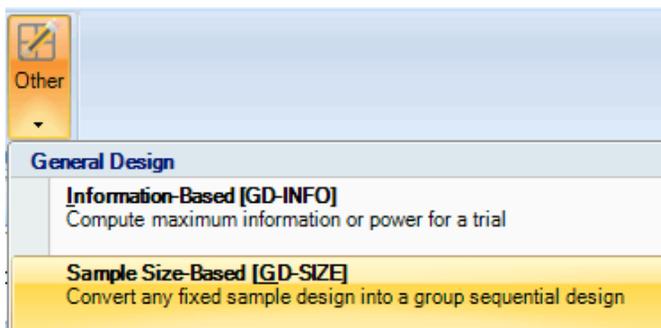
The goal is to design a balanced two-arm randomized clinical trial for high-risk patients with AF in which the standard treatment (adjusted-dose warfarin) has a Poisson event rate of 1.8% per year. If the experimental treatment (low-dose warfarin plus aspirin) has a Poisson event rate in excess of 3% per year, we wish to detect this with 90% power using a one-sided test conducted at the 5% level of significance. One can use a standard sample-size package like **Egret Siz** to determine the total number of events of ischaemic stroke that one must observe in order to detect a difference in Poisson rates of 1.8% per year versus 3% per year (i.e., a risk ratio of $3/1.8 = 1.667$) with 90% power using a one-sided fixed-sample Wald test conducted at the 5% significance level. The desired number works out to be 135 events. More direct

60 Design and Interim Monitoring with General Endpoints

methods of determining the required number of events, rather than relying on output from a statistical software package, are available through the information based approach discussed in Chapter 59, Section 59.3.

The above requirement of 135 events assumed that there would be no interim monitoring for early stopping. This study, however, was intended to be monitored twice during execution, and a third time at the end, each look being taken after equal increments of information. The group-sequential strategies implemented in East are applicable to this problem, and East can determine the amount by which the required number of events for the fixed-sample study should be inflated for the group-sequential design, and then allow to properly monitor the study. The first step is to provide East with the appropriate design parameters.

First, click **Other** on the **Design** tab and then click **General Design: Sample-Size Based** as shown below.



The upper pane displays the several fields with default values. First, change the **Number of Looks** to 3, to generate a study with two interim looks and a final analysis. In the **Design Parameters** tab, select **1-Sided** for **Test Type**. Specify **Type I Error (α)** as 0.05 and **Power ($1-\beta$)** as 0.9, respectively. Enter 135 for **Total SS for Fixed-Sample Study**. The **Design Parameters** tab in the upper pane should appear as

below:

Design: Other: General Design - Sample Size-Based

Number of Looks: 3

Test Parameters | Boundary

Test Type: 1-Sided

Type I Error (α): 0.05

Power: 0.9

Sample Size (n): Computed

Total SS for Fixed-Sample Study: 135

Click **Boundary Info**. In this tab, you will see **Efficacy** and **Futility** boxes. Select **Spending Functions** for **Boundary Family**, **Lan-DeMets** for **Spending Function** and **OF** for **Parameter** in **Efficacy** box. Select **None** for **Boundary Family** in **Futility** boxes. The **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The **Boundary Info** tab should appear as below.

Number of Looks: 3

Test Parameters | Boundary

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.05

Futility

Boundary Family: None

Spacing of Looks: Equal Unequal

Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.333 | 0.001 | 3.200 |
| 2 | 0.667 | 0.016 | 2.141 |
| 3 | 1.000 | 0.050 | 1.695 |

Click **Compute** to generate output for this design. A new row will be added in the **Output Preview** with label Des 1.

| ID | No. of Looks | Test Type | Specified α | Power | Spacing of Looks | Efficacy Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | Fixed SS |
|------|--------------|-----------|--------------------|-------|------------------|-------------------|-------------|------------------|------------------|----------|
| Des1 | 3 | 1-Sided | 0.05 | 0.9 | Equal | LD (OF) | 138 | 137.215 | 106.941 | 135 |

60 Design and Interim Monitoring with General Endpoints

In order to preserve the power of this study at 90% while monitoring the data three times, we must inflate the number of events required for a fixed sample size study from 135 to 138 events. That is, we must commit up front to keeping the study open until 138 ischaemic events are observed. On the other hand, since we will be monitoring the data sequentially, we expect to cross the stopping boundary and stop early after only 107 events, on average, if the alternative hypothesis is true. Thus, the increase in sample size corresponds to a small price to pay in order to benefit from the advantages of potential early stopping.

Save Des 1 in the current workbook by selecting the row corresponding to Des 1 in **Output Preview** and clicking  on the **Output Preview** toolbar. For any chosen design, the study has a certain probability of stopping at any of the looks. In order to see the stopping probabilities select Des 1 in the **Library**, and click .

Design: Other: General Design - Sample Size-Based

| Test Parameters | |
|---------------------------------|----------|
| Design ID | Des1 |
| Number of Looks | 3 |
| Test Type | 1- Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Total SS for Fixed-Sample Study | 135 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |

Sample Size Information

| | Total |
|-------------------------|---------|
| Sample Size (n) | |
| Maximum | 138 |
| Expected H1 | 106.941 |
| Expected H0 | 137.215 |
| Maximum Information (I) | 96.938 |

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries Efficacy Z | Boundary Crossing Probability (Cumulative) | |
|--------|--------------------------|-----------------|---------------------------|-----------------------|--|----------|
| | | | | | Under H0 | Under H1 |
| | | | | | Efficacy | Efficacy |
| 1 | 0.333 | 46 | 6.869E-4 | 3.2 | 6.869E-4 | 0.067 |
| 2 | 0.667 | 92 | 0.016 | 2.141 | 0.016 | 0.608 |
| 3 | 1 | 138 | 0.05 | 1.695 | 0.05 | 0.9 |

The clear advantage of this sequential design resides in the high probability of stopping by the second look, if the alternative is true, with a sample size of 92 patients, which is well below the requirements for a fixed sample study (135 patients). Close the Output window before continuing.

A less conservative approach would be to use stopping boundaries in the spirit of Pocock (1977). To generate stopping boundaries in the spirit of Pocock (1977), create a new design by right-clicking Des 1 in the **Library**, and selecting **Edit Design**. Go to the **Boundary Info** tab. As before, keep **Spending Functions** for **Boundary Family** and **Lan-DeMets** for **Spending Function**. Change the **Parameter** to **PK** in **Efficacy** boxes. Click **Compute**. A new row will be added in the **Output Preview**

with label Des 2.

| ▲ | ID | No. of Looks | Test Type | Specified α | Power | Spacing of Looks | Efficacy Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | Fixed SS |
|---|------|--------------|-----------|--------------------|-------|------------------|-------------------|-------------|------------------|------------------|----------|
|  | Des1 | 3 | 1-Sided | 0.05 | 0.9 | Equal | LD (OF) | 138 | 137.215 | 106.941 | 135 |
|  | Des2 | 3 | 1-Sided | 0.05 | 0.9 | Equal | LD (PK) | 157 | 153.815 | 95.597 | 135 |

Under this sequential scheme, we must commit up front to 157 events, but the expected number of events upon stopping the study is only 96 under the alternative hypothesis. Des 1 requires a smaller up front commitment, but Des 2 will stop with a smaller number of events, on average, if the alternative hypothesis is true. Now select Des 2 in

Output Preview and click  on the **Output Preview** toolbar to save in the **Library**.

The two designs considered can also be compared in terms of the actual stopping probabilities. In order to see the stopping probabilities with the boundaries with the spirit of Pocock, select Des 2 in the **Library**, and click .

Design: Other: General Design - Sample Size-Based

| Test Parameters | |
|---------------------------------|---------|
| Design ID | Des2 |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Total SS for Fixed-Sample Study | 135 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (PK) |

Sample Size Information

| | Total |
|-------------------------|---------|
| Sample Size (n) | |
| Maximum | 157 |
| Expected H1 | 95.597 |
| Expected H0 | 153.815 |
| Maximum Information (I) | 110.327 |

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries Efficacy Z | Boundary Crossing Probability (Cumulative) | |
|--------|--------------------------|-----------------|---------------------------|-----------------------|--|-------------------|
| | | | | | Under H0 Efficacy | Under H1 Efficacy |
| 1 | 0.333 | 52 | 0.023 | 2.002 | 0.023 | 0.428 |
| 2 | 0.667 | 105 | 0.038 | 1.994 | 0.038 | 0.745 |
| 3 | 1 | 157 | 0.05 | 1.98 | 0.05 | 0.9 |

The comparison of stopping probabilities across alternative design options can help in choosing the one with the most desirable properties. In particular, designs that require a larger maximum sample size are usually those that have rather high stopping probabilities at early analyses. Indeed, although Des 2 may require as many as 157 events if the alternative hypothesis is indeed true, there is a higher chance of stopping at the first analysis with this design (stopping probability = 0.428 with 52 events) than with Des 1 (stopping probability = 0.067 with 46 events).

Although the trial report did not mention which monitoring strategy, we will assume

60 Design and Interim Monitoring with General Endpoints

that the decision was made to use Des 1, with stopping boundaries in the spirit of O'Brien and Fleming, and we shall now proceed with the interim monitoring of the study. The inflation factor, $IF(\alpha, \beta, K, \text{boundaries})$, for Des 1 is

$$\frac{N_{max}}{N_1} = \frac{138}{135} = 1.022$$

The $IF = IF(\alpha, \beta, K, \text{boundaries})$ and $\eta = \eta(\alpha, \beta, K, \text{boundaries})$ are related as

$$IF = \left(\frac{\eta}{z_\alpha + z_\beta}\right)^2$$

With $IF = 1.022$, $\alpha = 0.05$ and $\beta = 0.1$, $\eta = 2.958$. We have obtained η through back calculation. In fact, East calculates IF and N_{max} from η . Although this parameter was not specified at the design stage, it is implied by the choice of power, type 1 error, number and spacing of looks and spending function. Specifically, a process of independent increments of the form $W(t) \sim N(\eta t, t)$ (as defined by equations (B.8), (B.9), and (B.10) in Section B.1 of Appendix B) in which $\eta = 2.958$, will cross the stopping boundary of the above study design at one of the three equally spaced monitoring times ($t_1 = 1/3, t_2 = 2/3$, or $t_3 = 1$) with probability $1 - \beta = 0.9$. The parameter η generated at the design stage is an abstract quantity of no inherent interest to the end user. However, as we shall see in the next two sections, point and interval estimates of η obtained from the data at the interim monitoring stage can be of great interest to the end user, for they can be transformed into corresponding estimates of the relevant treatment difference δ .

60.1.2 Interim Monitoring of Stroke Prevention Study

Select Des 1 in the **Library**, and click **IM** from the **Library** toolbar. Alternatively, right-click on Des 1 and select **Interim Monitoring**.

60.1.3 First Interim Analysis

The report does not mention how many events were observed at the first interim analysis and what the value of the test statistic was at that time. We shall suppose that the study was first monitored after 25 events. Suppose in addition that the treatment group was followed for 210 person years producing 20 events, and the control group was followed for 218 person years producing 5 events. With these data, we can test the null hypothesis that the event rate for ischaemic stroke is the same in the treatment and control groups. Before proceeding with this test, however, it is useful to review some basic theory about the Poisson distribution. Let (λ_c, λ_t) be the Poisson event rates for the treatment and control groups, respectively. It is convenient to characterize the treatment difference in terms of the logarithm of the risk ratio

$$\delta = \ln\left(\frac{\lambda_t}{\lambda_c}\right).$$

Then the test statistic of interest for testing $H_0: \delta = 0$ is the Wald statistic

$$Z = \frac{\hat{\delta}}{\text{se}(\hat{\delta})}. \tag{60.1}$$

This statistic is $N(0, 1)$ under the null hypothesis and has the appropriate covariance structure for group sequential inference provided $\hat{\delta}$ is an efficient estimate of δ . At the time of the interim analysis, let n_c denote the number of person years of follow-up in the control group, and let x_c be the corresponding number of events that are observed in the control group. Similarly, let n_t denote the number of person years of follow-up in the treatment group and let x_t be the corresponding number of events that are observed in the treatment group. An efficient estimator for δ is now given by

$$\hat{\delta} = \ln(x_t/n_t) - \ln(x_c/n_c). \tag{60.2}$$

In order to compute the standard error, $\text{se}(\hat{\delta})$, we need to derive the variance of the random variable

$$\ln(T_n) = \ln\left(\frac{X}{n}\right)$$

where X is a Poisson random variable with density

$$f(x) = \frac{(\lambda n)^x e^{-\lambda n}}{x!}.$$

By Poisson theory,

$$E(T_n) = \lambda$$

and

$$\text{var}(T_n) = \frac{\lambda}{n}.$$

Thus, under the null hypothesis $H_0: \delta = 0$,

$$\sqrt{n}(T_n - \lambda) \xrightarrow{d} N(0, \lambda).$$

Therefore, by the delta method (see for example, Agresti, 1990, page 420)

$$\sqrt{n}[g(T_n) - g(\lambda)] \xrightarrow{d} N(0, \lambda[g'(\lambda)]^2).$$

Here $g(\lambda) = \ln(\lambda)$. Therefore

$$\lambda[g'(\lambda)]^2 = \frac{1}{\lambda}$$

and hence

$$\sqrt{n}[\ln(T_n) - \ln(\lambda)] \xrightarrow{d} N\left(0, \frac{1}{\lambda}\right).$$

60 Design and Interim Monitoring with General Endpoints

It follows that

$$\text{var}[\ln(T_n)] = \frac{1}{n\lambda}. \quad (60.3)$$

Substituting this result into equation (60.2) we have

$$\text{var}(\hat{\delta}) = \frac{1}{n_c \lambda_c} + \frac{1}{n_t \lambda_t}.$$

Replacing the Poisson event rates λ_c and λ_t by their corresponding maximum likelihood estimates x_c/n_c and x_t/n_t we finally obtain

$$\text{se}(\hat{\delta}) = \sqrt{\frac{1}{x_c} + \frac{1}{x_t}} \quad (60.4)$$

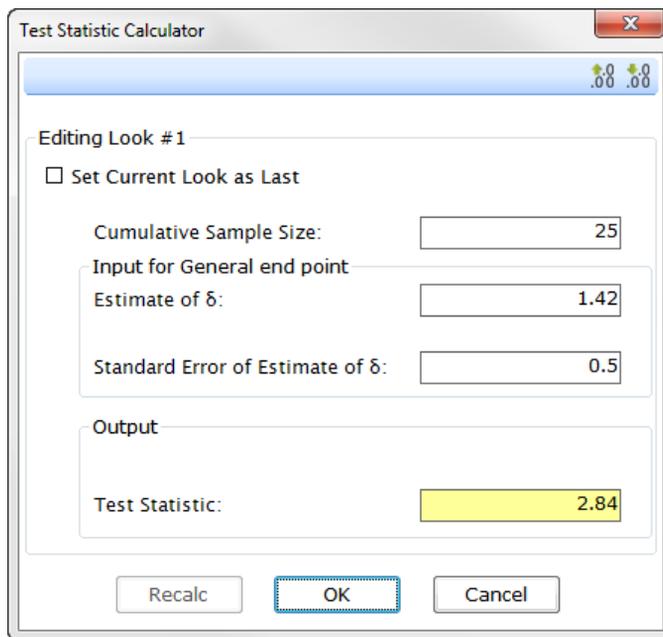
so that the test statistic (60.1) becomes

$$Z = \frac{\ln\left(\frac{x_t}{n_t}\right) - \ln\left(\frac{x_c}{n_c}\right)}{\sqrt{\frac{1}{x_c} + \frac{1}{x_t}}}. \quad (60.5)$$

Substituting the observed values of x_c, x_t, n_c, n_t into equation (60.2) and (60.4), we obtain $\hat{\delta} = 1.423682$ and $\text{se}(\hat{\delta}) = 0.5$. Thus the first interim analysis is performed after observing 55 events with the value of the test statistic being $1.42/0.5 = 2.84$.

At the top of the Interim Monitoring sheet, click [Enter Interim Data](#) from the toolbar to invoke the **Test Statistic Calculator**. In this dialog box, enter 25 in **Cumulative Sample Size**, 1.42 for **Estimate of δ** and 0.5 for **Standard Error of Estimate of δ** .

Then, click **Recalc**.



Click **OK**. East displays the information fraction, $t(\tau_1) = 25/135 = 0.181$, test statistic, $T(\tau_1) = 1.42/0.5 = 2.84$ and efficacy boundary as 4.458. Thus, we can stop the study if the value of test statistic exceeds 4.458. Since this is not the case, we continue to the next interim monitoring time point.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 95% RCI for δ | | Repeat ... p-value | CP | Predicti... Power |
|--------|----------------------|------------------------|----------------|------|----------------|----------|---------------|--------|--------------------|----|-------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | 0.181 | 25 | 2.84 | 1.42 | 0.5 | 4.458 | Infinity | -0.809 | 0.194 | 1 | 0.992 |
| 2 | | | | | | | | | | | |
| 3 | | | | | | | | | | | |

The lower 95% confidence bound on η is -3.803. We can convert this estimate into a lower 95% confidence bound for δ by using the relationship

$$\eta = \delta \sqrt{I_{\max}}$$

derived in Section B.1 of Appendix B. Now observe that

$$I_{\max} = \frac{I_{\max}}{I_1} I_1 = t_1^{-1} [\text{se}(\hat{\delta}_1)]^{-2} .$$

60 Design and Interim Monitoring with General Endpoints

Therefore

$$\delta = \eta \sqrt{t_1} [\text{se}(\hat{\delta}_1)] \quad (60.6)$$

Thus, the lower confidence bound for δ is $-3.803 \times \sqrt{0.181} \times 0.5 = -0.809$. We can conclude that based on the current data, the ratio of treatment event rate to control event rate is at least $\exp(-0.809) = 0.445$. There is not yet sufficient evidence to exclude a ratio of 1.0.

60.1.4 Second Interim Analysis

A published report (JAMA, vol 279, No 16, Table 2) shows that this study was indeed monitored after 55 events were observed. There were only 11 events on the adjusted dose arm (control) with 581 patient years of observation. On the other hand there were 44 events on the fixed dose plus aspirin arm (treatment) with 558 patient years of observation. Entering these data into equations (60.2), (60.4) and (60.5) we obtain $\hat{\delta} = 1.427$, $\text{se}(\hat{\delta}) = 0.337$ and $Z = 4.234$.

In the top part of the IM dashboard, enter 55 for **Cumulative Sample Size**, 1.427 for **Estimate of δ** , and 0.337 for **Standard Error of Estimate of δ** . Click **OK**.

Test Statistic Calculator

Editing Look #2

Set Current Look as Last

Cumulative Sample Size: 55

Input for General end point

Estimate of δ : 1.427

Standard Error of Estimate of δ : 0.337

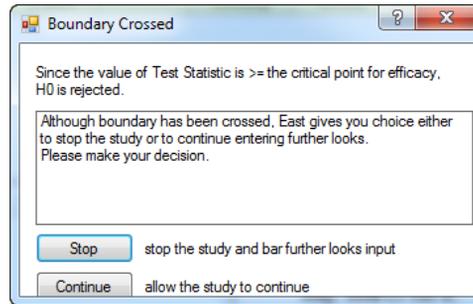
Output

Test Statistic: 4.234

Recalc OK Cancel

Click **OK** to update the charts and tables in the dashboard. Now, the stopping

boundary is crossed, and the following window appears.



Click **Stop**. The left side of the dashboard will show the stopping boundaries and the error spending function.

The right side of the dashboard will show a table for final inference, and the confidence intervals:

With 39.9% of the information, East was able to reach an early decision in favor of the alternative hypothesis, that fixed-dose warfarin plus aspirin is insufficient for stroke prevention.

61 *Early Stopping for Futility*

Group sequential methods were developed originally for early stopping if the experimental treatment showed a statistically significant therapeutic advantage at an interim look. In many clinical trials, however, there is limited interest in stopping early for a positive efficacy outcome. This is usually because the investigators wish to continue the trial all the way to the end and gather additional safety data for the experimental arm. Nevertheless, there is a great deal of interest in stopping early for futility if the interim analysis reveals that, with high probability, the trial will end up negative. In that case, the investigators might wish to cut their losses and possibly divert their resources to a more promising study.

East provides two ways to stop early for futility: (a) informal – based on conditional power and (b) formal – based on futility stopping boundaries. Industry trials have typically adopted the informal approach, stopping early if the conditional power at an interim analysis is extremely low. We consider this approach to be informal because it is not necessary to specify ahead of time how low the conditional power should be in order to declare futility and terminate the study. The futility threshold can be determined at the time of the interim analysis itself, possibly using both internal data from the trial and external information about other similar trials. It is easy to see that the informal approach will not inflate the type-1 error, provided the only decisions possible at each interim monitoring time point are to either continue the study or stop and declare futility. On the other hand, the informal approach may not preserve the type-2 error (and thus, the study may lose power) as the decision to stop for futility is based on an ad hoc determination that the conditional power is too low. In contrast, the use of a futility boundary guarantees the preservation of power. This is because the boundary is constructed by using the spending function methodology of Lan and DeMets (1983). However, in this case one spends β , the type-2 error, rather than spending α , the type-1 error. The technical details are available in Appendix B.

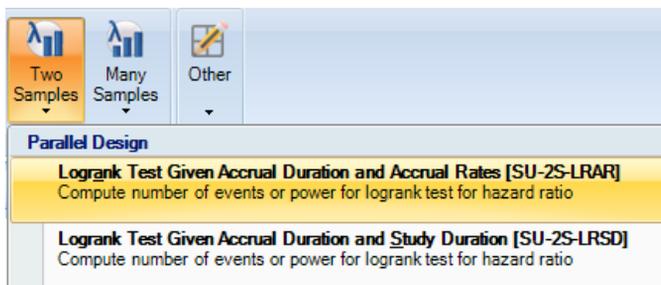
61.1 *Example: Survival in patients with advanced melanoma*

A phase III trial was conducted to compare overall survival (OS) in Tremelimumab, a fully human anti-cytotoxic T lymphocyte-associated antigen 4 (CTLA4) monoclonal antibody with standard, single-agent chemotherapy (Ribas et. al., 2008). Primary endpoint was OS. Let λ_t and λ_c be the overall survival rate in Tremelimumab and standard chemotherapy, respectively. Here, the treatment effect δ is represented in terms of $\ln(\lambda_t/\lambda_c)$ or the log hazard ratio. Therefore, $\delta < 0$ indicates the beneficial effect of new treatment, Tremelimumab. The study was designed to provide 90% power to detect a 33% improvement in true median OS with an unstratified log-rank test at overall 2-sided significance level of 0.05. Two equally spaced interim analyses were planned based on the group sequential design using the Lan-DeMets alpha and beta spending approach to an O'Brien-Fleming boundary. Improvement of 33% in true

median OS can be translated to ratio of medians as 1.33. In other words, we are considering a hazard ratio of $\ln \frac{1}{1+0.33}$ or 0.752. In the study, a median survival time of 10.7 months was observed in the standard chemotherapy group.

61.2 Single-Look Design with No Early Stopping

Suppose initially that no interim monitoring is contemplated. First, click **Survival: Two Samples** on the **Design** tab, and then click **Parallel Design: Logrank Test Given Accrual Duration and Accrual Rates**.



In the input window, leave the **Number of Looks** as 1. In the **Design Parameters** tab, select **Design Type** as **Superiority**, **Test Type** as **2-Sided**, and the values for **Type I Error (α)** and **Power ($1-\beta$)** as 0.05 and 0.9, respectively. Select **# of Hazard Pieces** as 1 which implies that hazard rates remain constant over time in both Tremelumab and standard chemotherapy. Select the **Input Method** as **Median Survival Times**. Tick the check box for **Hazard Ratio (Optional)** and select the radio-button **Ratio of Medians (m_t/m_c)**. Enter 1.33 for **Ratio of Medians (m_t/m_c)**. In the table below, enter 10.7 for **Med. Surv. Time (Control)**. The **Design**

61 Early Stopping for Futility

Parameters tab should now appear as below:

Design Type: Number of Looks:

Test Parameters **Accrual / Dropouts**

Test Type: # of Hazard Pieces: Input Method:

Type I Error (α): Hazard Ratio (Optional) Alternative

Power: Hazard Ratio (λ_t / λ_c)

No. of Events: Ratio of Medians (m_t / m_c)

| Med.Surv.Time | |
|-----------------------------------|-------------------------------------|
| Control | Treatment: Alt. |
| <input type="text" value="10.7"/> | <input type="text" value="14.231"/> |

Variance of Log Hazard Ratio

Null Alternative

Move to the **Accrual / Dropout Info** tab. The original study does not report about accrual information. However, we will assume that the patients arrive in the study at the rate of 48 per month. For this example, select **1** for # of **Accrual Periods** and enter

48 in the **Accrual Rate** column of the ensuing table.

Test Parameters
Accrual / Dropouts

Subjects are followed: Until End of Study

Accrual Info

of Accrual Periods: 1

Accrual Rate: 48

Accrual

| | Min. | Comtd. | Sugg. Max. |
|--|--------|--------|------------|
| <input type="radio"/> Duration: | 10.771 | 17.333 | 23.875 |
| <input checked="" type="radio"/> Subjects: | 517 | 832 | 1146 |

Click **Compute** to obtain the number of events required to have the desired operating characteristics. This will add a row in the **Output Preview**. The computed maximum number of events (517) is highlighted in yellow.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Accrual Rate | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events | Exp. Events (H0) | Exp. Events (H1) |
|------|-------------|--------------|-----------|--------------------|-------|-------|--------------|-------------|------------------|------------------|----------------|------------------|------------------|
| Des1 | Superiority | 1 | 2-Sided | 0.05 | 0.9 | 1 | 48 | 832 | 832 | 832 | 517 | 517 | 517 |

Select Des 1 in **Output Preview** and click . This will display the design details

61 Early Stopping for Futility

in the **Output Summary**.

| | Des 1 |
|---|--------------------|
| Mnemonic | SU-2S-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.752 |
| Var (Log HR) | Null |
| Accrual & Dropout Parameters | |
| Accrual Rate | 48 |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 832 |
| Expected Under H0 | 832 |
| Expected Under H1 | 832 |
| Events | |
| Maximum | 517 |
| Expected Under H0 | 517 |
| Expected Under H1 | 517 |
| Study Duration | |
| Maximum | 26.658 |
| Expected Under H0 | 24.462 |
| Expected Under H1 | 26.658 |
| Accrual Duration | |
| Maximum | 17.333 |
| Expected Under H0 | 17.333 |
| Expected Under H1 | 17.333 |

Click on the icon  to go back to the **Output Preview** window. Select Des 1 by clicking anywhere along the row in the **Output Preview** and click  to save this design in the **Library**. Des 1 shows that, in order to achieve the desired 90% power, we must keep the study open until 517 events are observed. Half of these events need to be observed in Tremeliumab arm, and another half in the standard chemotherapy arm. You can see the exact number of events required in each arm by double-clicking on Des 1 in the **Library**. In this design, there is no provision for interim monitoring to stop the trial early.

61.3 Group Sequential Design with Early Stopping for Efficacy

Recall from section 61.1 that the study was originally planned with two interim looks with the Lan-DeMets spending approach to an O'Brien-Fleming boundary. In this section, we will consider early stopping boundaries for efficacy only. Create a new design by selecting Des 1 in the **Library**, and clicking the  icon on the **Library** toolbar. First, change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. A new tab with label **Boundary Info** will appear. The left side contains details for the **Efficacy** boundary, and the right side contains details for the **Futility** boundary. Select **Spending Functions** for **Boundary Family**, **Lan-DeMets** for **Spending Function** and **OF** for **Parameter** in **Efficacy** box. Select **None** for **Boundary Family** in **Futility** box. Select **None** for **Boundary Family** in **Futility** box.

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary Accrual / Dropouts

Efficacy
 Boundary Family: Spending Functions
 Spending Function: Lan-DeMets
 Parameter: OF
 Type I Error (α): 0.05

Futility
 Boundary Family: None

Spacing of Looks Equal Unequal Efficacy Boundary: Z Scale  

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|--------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.333 | 0.000 | 3.710 | -3.710 |
| 2 | 0.667 | 0.012 | 2.511 | -2.511 |
| 3 | 1.000 | 0.050 | 1.993 | -1.993 |

Click **Compute** to generate output for this design. A new row will be added in the **Output Preview**. Save this design in the current workbook by selecting the row corresponding to Des 2 in **Output Preview** and clicking  on the **Output Preview** toolbar. Des 2 requires a larger up-front commitment than Des 1. To compare Des 1 and Des 2, select both rows in **Output Preview** using the Ctrl key and click 

61 Early Stopping for Futility

icon. Both designs will be displayed in the **Output Summary**.

| | Des 1 | Des 2 |
|---|--------------------|--------------------|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 3 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Hazard Ratio (Alt.) | 0.752 | 0.752 |
| Var (Log HR) | Null | Null |
| Boundary Parameters | | |
| Spacing of Looks | | Unequal |
| Efficacy Boundary | | LD (OF) |
| Accrual & Dropout Parameters | | |
| Accrual Rate | 48 | 48 |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 0 | 0 |
| Sample Size | | |
| Maximum | 832 | 832 |
| Expected Under H0 | 832 | 831.947 |
| Expected Under H1 | 832 | 824.54 |
| Events | | |
| Maximum | 517 | 523 |
| Expected Under H0 | 517 | 520.851 |
| Expected Under H1 | 517 | 419.466 |
| Study Duration | | |
| Maximum | 26.658 | 27.003 |
| Expected Under H0 | 24.462 | 24.674 |
| Expected Under H1 | 26.658 | 22.306 |
| Accrual Duration | | |
| Maximum | 17.333 | 17.333 |
| Expected Under H0 | 17.333 | 17.332 |
| Expected Under H1 | 17.333 | 17.178 |

In order to achieve the desired 90% power, the study in Des 2 should be kept open until 523 events are obtained. However, under H_1 , the required number of events is 420 with expected study duration of 22 months only, compared to 517 events and 26.6 months for Des 1. To see the probability of crossing the stopping boundaries at one of the interim looks, and thus terminating the study earlier, double-click on Des 2 in the

Library.

⊖ Events, Sample Size, Pipeline and Analysis Times: Look by Look (Under H0)

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (s) | Pipeline (n-s) | Analysis Time | Boundary Crossing Probability (Cumulative) | |
|--------|--------------------------|-----------------|------------|----------------|---------------|--|----------|
| | | | | | | Efficacy | |
| | | | | | | Upper | Lower |
| 1 | 0.333 | 574 | 174 | 400 | 11.938 | 1.019E-4 | 1.019E-4 |
| 2 | 0.667 | 832 | 349 | 483 | 17.864 | 0.006 | 0.006 |
| 3 | 1 | 832 | 523 | 309 | 24.759 | 0.025 | 0.025 |

⊖ Events, Sample Size, Pipeline and Analysis Times: Look by Look (Under H1)

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (s) | Pipeline (n-s) | Analysis Time | Boundary Crossing Probability (Cumulative) | |
|--------|--------------------------|-----------------|------------|----------------|---------------|--|-------|
| | | | | | | Efficacy | |
| | | | | | | Upper | Lower |
| 1 | 0.333 | 609 | 174 | 435 | 12.678 | 1.103E-8 | 0.033 |
| 2 | 0.667 | 832 | 349 | 483 | 19.014 | 1.251E-7 | 0.561 |
| 3 | 1 | 832 | 523 | 309 | 27.003 | 1.914E-7 | 0.9 |

You can increase the decimal precision by clicking on the  icon and displaying Probability Statistics up to four decimal places. Under H_1 there is a 3.34% chance of crossing a boundary at the first look, and 56% chance of crossing at the second look (this column is cumulative). This is why the expected study duration is about 4.5 months less than the study duration with Des 1. However, Des 2 has no formal mechanism for stopping the trial early if the two treatments are similar. Under the null hypothesis, the expected study duration under H_0 is nearly the same as for a single look design.

61.4 Informal Use of Conditional Power for Futility Stopping

One can use conditional power as an informal guide for terminating a study at an interim monitoring time point. To see how this works, recall that the study has been designed for two interim looks: first, when one-third of deaths are observed and second, when two-thirds of deaths are observed.

Right-click Des 2 in the **Library**, and select **Interim Monitoring**.

First interim monitoring

Click **Enter Interim Data** from the toolbar to invoke the **Test Statistic Calculator**. In this dialog box, enter 175 for **Cumulative Events**, 0.143 as **Estimate of δ** and 0.477 as **Standard Error of Estimate of δ** . Click **Recalc**. The test statistic value is computed and is displayed as 0.3. This appears to be a rather disappointing value for the test statistic half-way through the study, and suggests that the study might not end

61 Early Stopping for Futility

up positive after all.

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Read from Analysis Node

Select Workbook: [Dropdown]

Select Analysis Node: [Dropdown]

Cumulative Events: [Input: 175]

Input for Survival end point

Estimate of δ : [Input: 0.143]

$\delta = \ln(\lambda_t / \lambda_c)$

Standard Error of Estimate of δ : [Input: 0.477]

Output

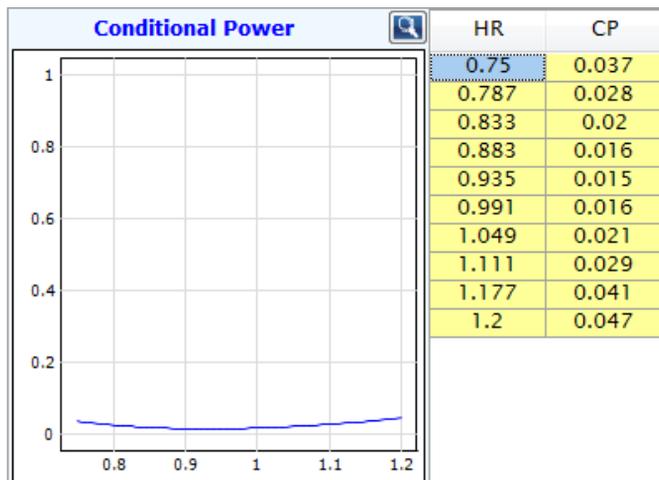
Test Statistic: [Input: 0.3]

Recalc OK Cancel

Click **OK** to continue. This will paste the information in the monitoring dashboard.

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of ... | Est. of δ | Std. Error of Est. of δ | Efficacy | | 95% RCI for HR | | Repeat ... p-value | CP | Predicti... Power |
|--------|----------------------|-------------------|----------------|-------------|------------------|--------------------------------|----------|--------|----------------|-------|--------------------|-------|-------------------|
| | | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.335 | 175 | 0.3 | 1.154 | 0.143 | 0.477 | 3.703 | -3.703 | 6.747 | 0.197 | 1 | 0.037 | 0.186 |
| 2 | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | |

Examine the **Conditional Power** section of the monitoring sheet.



Conditional powers are calculated at different effect sizes. The conditional power corresponding to HR of 0.1.2 (which is very close to observed HR of 1.1545) is only 0.047. This means that if we were to perform an analysis of the data at 523 events, there is only a 4.7% chance of crossing the upper stopping boundary and declaring statistical significance. Is this chance sufficiently small to warrant terminating the study? There are no objective criteria for making this determination. Recall that the conditional power approach to stopping early for futility is informal. Thus, the low conditional power would have to be considered by the DMC, along with other factors such as toxicity, rate of accrual and parallel developments in other trials.

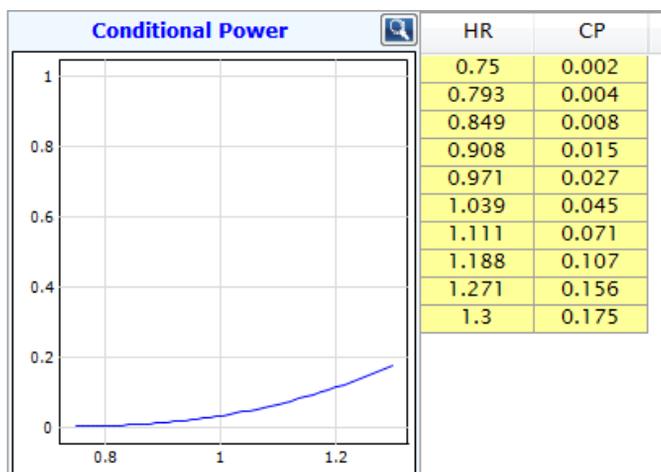
Second interim monitoring Suppose the trial continues and a second interim analysis is performed when almost two-thirds of the events are observed. Assume that the total number of events is 350, and the estimates of $\hat{\delta}=0.237$ and $SE(\hat{\delta})=0.206$. Enter these values in the test statistic calculator to post the results into the interim monitoring dashboard.

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of HR | Est. of δ | Std. Error of Est. of δ | Efficacy | | 95% RCI for HR | | Repeat... p-value | CP | Predict... Power |
|--------|----------------------|-------------------|----------------|------------|------------------|--------------------------------|----------|--------|----------------|-------|-------------------|-------|------------------|
| | | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.335 | 175 | 0.3 | 1.154 | 0.143 | 0.477 | 3.703 | -3.703 | 6.747 | 0.197 | 1 | 0.037 | 0.186 |
| 2 | 0.669 | 350 | 1.15 | 1.267 | 0.237 | 0.206 | 2.506 | -2.506 | 2.124 | 0.756 | 0.443 | 0.154 | 0.202 |
| 3 | | | | | | | | | | | | | |

Although the value of the test statistic has increased considerably from the value at the previous look, the conditional power has only marginally increased, from 0.047 to

61 Early Stopping for Futility

0.175. Because we are very close to the end of the study, there is only a 17.5% chance of crossing the upper stopping boundary at the final look. Should the study continue or be terminated? Again, the decision is a subjective one.



61.5 Combined Efficacy and Futility Stopping Boundaries

61.5.1 Two-Sided Tests

61.5.2 One-Sided Test

61.5.3 Conservative Futility Boundaries

One way to remove the subjectivity from the decision to stop early based on low conditional power is to use formal futility stopping boundaries. East has the provision to simultaneously create efficacy boundaries for rejecting H_0 and futility boundaries for rejecting H_1 . The efficacy boundaries are generated by an α -spending function that spends the type-1 error. The futility boundaries are generated by a β -spending function that spends the type-2 error. Moreover the two sets of boundaries are forced to meet at the last look so as to ensure that either H_0 or H_1 is rejected.

61.5.1 Two-Sided Tests

Recall that the advanced melanoma study we are considering in this section was implemented using the Lan-DeMets alpha and beta spending approach to an O'Brien-Fleming boundary. We will first consider a two-sided design with both efficacy and futility boundaries. In order to do this, create a new design by selecting Des 2 in the **Library**, and clicking the  icon on the **Library** toolbar. Click the **Boundary Info** tab. Select **Spending Functions** for **Boundary Family**, **Lan-DeMets** for **Spending Function** and **OF** for **Parameter** in both **Efficacy** and **Futility** boxes. In the right of the **Futility** box there is a field where you have to choose either **Non-Binding** or **Binding**. Binding futility boundary refers to a situation where the trial must be terminated once the test statistic falls within the futility

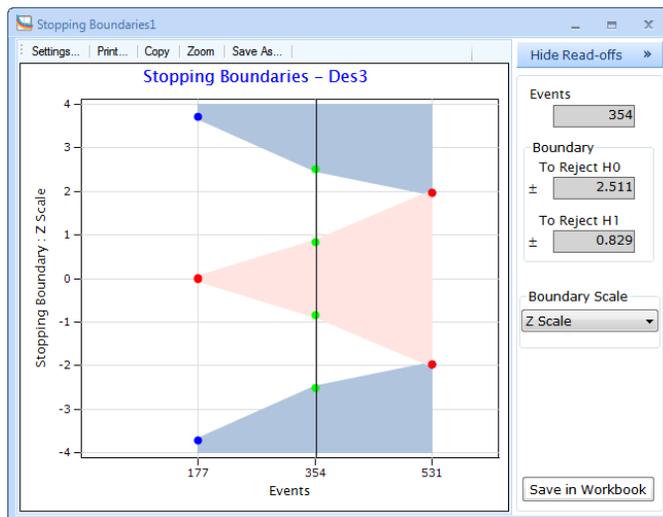
boundaries; otherwise overall type I error might be inflated. Non-Binding futility boundaries do not have this constraint. For now, select the radio-button corresponding to **Binding**. The cumulative α and β spent along with the boundary values are shown in the table in the **Boundary Info** tab. The columns **Stop for Efficacy** and **Stop for Futility** in the table provide the flexibility of excluding either efficacy of futility boundaries in certain interim looks, by unchecking the corresponding cells. For this example, leave all the boxes in columns **Stop for Efficacy** and **Stop for Futility** checked. Click **Compute**.

A new row will be added in the **Output Preview** labeled as Des 3. Save this design in the current workbook by selecting the row corresponding to Des 3 in **Output Preview** and clicking  on the **Output Preview** toolbar. To compare Des 1, Des 2, and Des 3, select all three rows in **Output Preview** using the Ctrl key and click  icon. All three designs will be displayed in the **Output Summary**.

| | Wbk1:Des1 | Wbk1:Des2 | Wbk1:Des3 |
|---|--------------------|--------------------|--------------------|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR | SU-2S-LRAR |
| Test Parameters | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 1 | 3 | 3 |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.9 | 0.9 | 0.9 |
| Model Parameters | | | |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| Hazard Ratio (Alt.) | 0.752 | 0.752 | 0.752 |
| Var (Log HR) | Null | Null | Null |
| Boundary Parameters | | | |
| Spacing of Looks | | Unequal | Unequal |
| Efficacy Boundary | | LD (OF) | LD (OF) |
| Futility Boundary | | | LD (OF) (B) |
| Accrual & Dropout Parameters | | | |
| Accrual Rate | 48 | 48 | 48 |
| Subjects are Followed | Until End of Study | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 | 1 |
| No. of Dropout Pieces | 0 | 0 | 0 |
| Sample Size | | | |
| Maximum | 832 | 832 | 848 |
| Expected Under H0 | 832 | 831.947 | 845.71 |
| Expected Under H1 | 832 | 824.54 | 839.527 |
| Events | | | |
| Maximum | 517 | 523 | 531 |
| Expected Under H0 | 517 | 520.851 | 422.044 |
| Expected Under H1 | 517 | 419.466 | 418.193 |
| Study Duration | | | |
| Maximum | 26.658 | 27.003 | 27.08 |
| Expected Under H0 | 24.462 | 24.674 | 20.648 |
| Expected Under H1 | 26.658 | 22.306 | 22.081 |
| Accrual Duration | | | |
| Maximum | 17.333 | 17.333 | 17.667 |
| Expected Under H0 | 17.333 | 17.332 | 17.619 |
| Expected Under H1 | 17.333 | 17.178 | 17.49 |

61 Early Stopping for Futility

Select Des 3 in the **Library**, and click  , then select **Stopping Boundaries**.



Des 3 requires a commitment to keep the study open until either 531 events are observed or a boundary is crossed. However, by providing upper and lower stopping boundaries and an inner wedge, Des 3 has lower expected study durations under both the null and alternative hypotheses. If the test statistic enters:

- the pink zone (the inner wedge), the trial stops, the alternative hypothesis is rejected, and futility is declared.
- the lower blue zone, the trial stops, the null hypothesis is rejected, and the new treatment Tremelimumb is declared to be beneficial relative to the standard chemotherapy.
- the upper blue zone, the trial stops, the null hypothesis is rejected, and the Tremelimumb is declared to be harmful relative to the standard chemotherapy.

These boundaries are constructed in such a way that:

- if the null hypothesis is true (i.e., $\delta = \ln \lambda_t / \lambda_c = 0$), the test statistic will enter the pink inner wedge region with probability $1 - \alpha = 0.95$, the upper blue zone with probability 0.025 and the lower blue zone with probability 0.025.
- if the alternative hypothesis is true with $\delta = \ln \lambda_t / \lambda_c \leq \ln 0.752 = -0.285$, the test statistic will enter the pink zone with probability $\beta = 0.1$ and the lower blue zone with probability almost equal to 0.9.
- if the alternative hypothesis is true with $\delta \geq 0.285$ the test statistic will enter the

pink zone with probability $\beta = 0.1$ and the upper blue zone with probability almost equal to 0.9.

The inner wedge boundaries give us the chance to stop early if H_0 is true. Notice that with Des 3, the expected study duration under H_0 is only 20.639 months, as compared to 24.668 months with Des 2. Close this chart before continuing.

61.5.2 One-Sided Test

In Des 3 we utilized a total of four boundaries – two-sided upper and lower boundaries for rejecting H_0 , and two-sided upper and lower boundaries for rejecting H_1 . Such boundaries are only necessary if we wish to actually continue the trial until we have demonstrated that the new treatment is significantly worse than the standard treatment; i.e., until the test statistic enters the lower blue zone and rejects H_0 in favor of $H_1 : \delta \leq 0$. If, however, we are willing to stop the study early if equivalence rather than actual harm is demonstrated, a more efficient design consisting of only two boundaries can be devised. Create a new design by selecting Des 3 in the **Library**, and clicking the  icon on the **Library** toolbar. Click the **Design Parameters** tab. Replace **2-Sided** by **1-Sided**, and replace the significance level, $\alpha = 0.05$ by $\alpha = 0.025$.

| Test Parameters | Boundary | Accrual / |
|----------------------------|---------------|-----------|
| Test Type: | 1-Sided | |
| Type I Error (α): | 0.025 | |
| Power: | 0.9 | |
| No. of Events: | Computed | |
| Allocation Ratio: | 1 | |
| | (n_t / n_c) | |

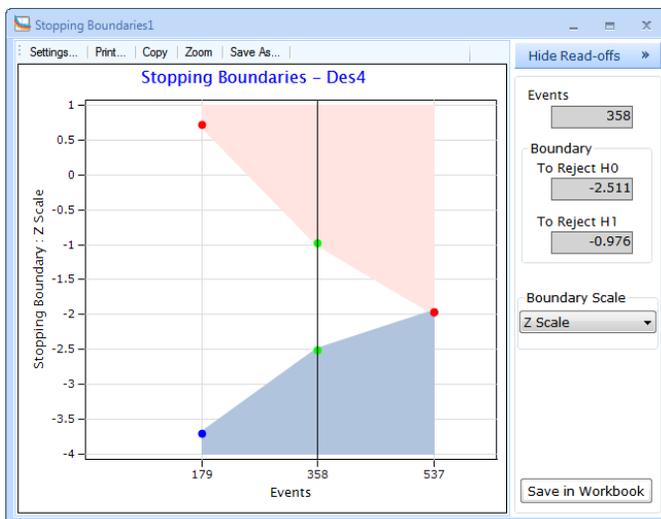
Go to the **Boundary Info** tab. Select **Spending Functions** for **Boundary Family**, **Lan-DeMets** for **Spending Function** and **OF** for **Parameter** in both

61 Early Stopping for Futility

Efficacy and **Futility** boxes. Select the radio-button corresponding to the **Binding**.

| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Cum. β Spent | Futility Boundary |
|--------|----------------|-------------------------------------|-------------------------------------|---------------------|-------------------|--------------------|-------------------|
| 1 | 0.333 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.000 | -3.710 | 0.004 | 0.713 |
| 2 | 0.667 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.006 | -2.511 | 0.044 | -0.976 |
| 3 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.025 | -1.959 | 0.100 | -1.959 |

Click **Compute**. This will add a new row to the **Output Preview**. Save this design in the current workbook by selecting the row corresponding to Des 4 in **Output Preview** and clicking  on the **Output Preview** toolbar. Select Des 4 in the **Library**, and then click , and select **Stopping Boundaries**.



Des 4 requires a commitment to keep the study open until either 537 events are observed or one of the two boundaries is crossed. If the test statistic crosses:

- the upper boundary and enters the pink zone the trial stops, the alternative

- hypothesis is rejected, and futility is declared.
- the lower boundary and enters the blue zone the trial stops, the null hypothesis is rejected, and the new treatment is declared to be beneficial over the standard chemotherapy.

These boundaries are forced to meet at the end of 537 events, thus ensuring that either H_0 or H_1 will be rejected. They are constructed so that:

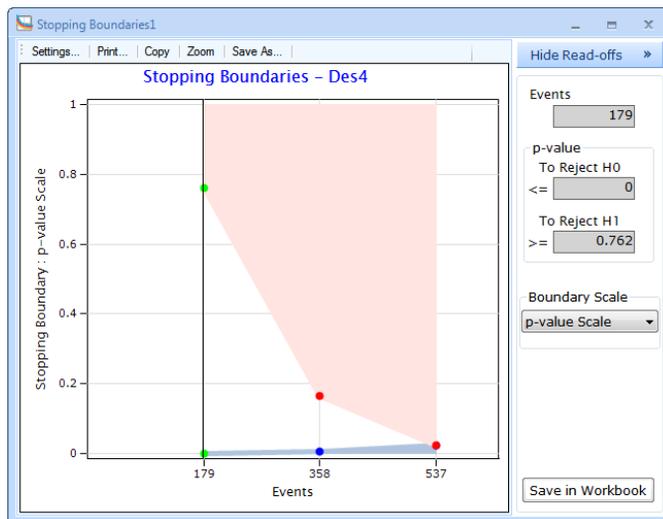
- if the null hypothesis is true (i.e., $\delta = \ln \lambda_t / \lambda_c = 0$), the test statistic will enter the pink zone with probability $1 - \alpha = 0.975$ and the blue zone with probability 0.025
- if the alternative hypothesis is true (i.e., $\delta = \ln \lambda_t / \lambda_c = -0.285$), the test statistic will enter the pink zone with probability $\beta = 0.1$ and the blue zone with probability 0.9.

Des 4 therefore meets the regulatory requirement that the false positive rate for a one sided test should not exceed 0.025. It also meets the sponsor's requirement that the study be designed for 90% power. In terms of shortening the expected study duration, however, Des 4 completely dominates the other three designs. Under H_0 the expected study duration is less than 18 months, a saving of over 6.5 months compared to Des 1. There is also over 4.5 months of expected saving relative to Des 1 if H_1 is true.

Unlike the informal approach, based on conditional power, Des 4 utilizes a formal futility boundary. Since the futility boundary is derived from a β -spending function, the type-2 error (and hence the power of the study) is fully controlled. A drawback of this approach is the loss of flexibility to keep the study open if the futility boundary is crossed. In this case, we must terminate the study. If we keep on accruing patients even after crossing a futility boundary, we are no longer assured of preserving the type-1 error. For this reason, it is important to examine the futility boundary from every angle before making the commitment. Accordingly, let us examine the stopping boundaries again, this time on the p-value scale. To display the boundaries on the p-value scale

61 Early Stopping for Futility

you must select this scale from the drop-down list in the Stopping Boundaries chart.

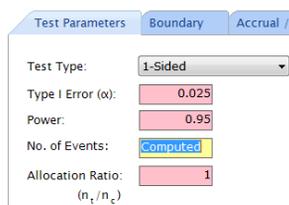


If the p-value (one-sided) at the first look exceeds 0.7622 the study should be terminated for futility. At the second look the futility criterion is $p = 0.1646$ and at the final look it is $p = 0.0251$. These values reveal several psychological drawbacks of the selected futility boundary. For instance, even though the overall power of the study is preserved, most investigators would be unwilling to terminate a study and declare futility at an interim analysis where the p-value was 0.1646; they would prefer to complete the study in hopes of a further decline in the p-value. Also, since the boundaries meet at the final look, one could technically reject the null hypothesis and claim that the trial is a success if the final p-value is less than 0.0251. This could appear counter-intuitive because one expects to pay a penalty for having taken multiple looks at the data. Usually the penalty amounts to requiring the cut-off for the final p-value to be less than $\alpha = 0.025$ in order to declare significance and reject H_0 . Here, however, the cut-off for the final p-value exceeds 0.025. It appears that we have been rewarded rather than penalized for having designed a multiple-look study.

The reason is that the presence of a futility boundary reduces the risk of crossing the efficacy stopping boundary. If the study were designed with an efficacy boundary only, it would be at risk of crossing the efficacy boundary at each interim look. This would elevate the overall type-1 error unless we imposed a suitable penalty on the final p-value to compensate. On the other hand if the study were designed with a futility boundary only, it would be at risk of crossing the futility boundary at each interim

look. This would reduce the overall type-1 error unless we rewarded the final look p-value by a suitable amount to compensate. When both efficacy and futility boundaries are present, the efficacy boundaries tend to lower the cut-off for the final p-value to below α whereas the futility boundaries tend to increase the cut-off for the final p-value to above α . Depending on the choice of stopping boundaries the number and timing of the looks and the values of α and β , one or other of these opposing forces dominates, resulting in a cut-off for the final p-value that is sometimes greater than α and sometimes less.

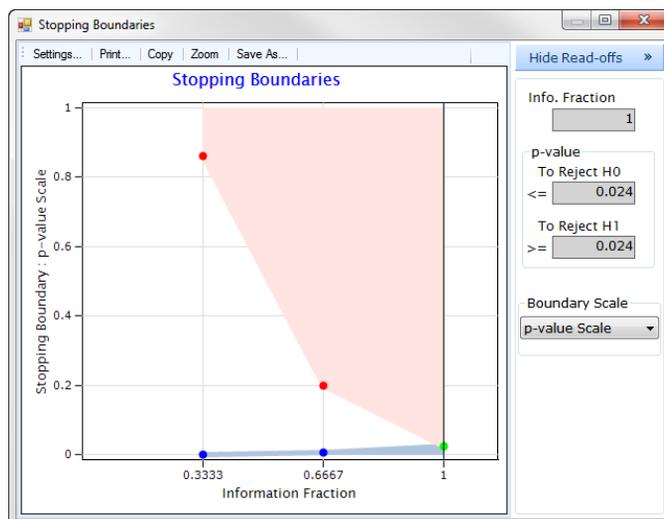
Select Des 4 in the **Library**, and click the  icon on the **Library** toolbar. Change the power from 90% to 95% in the **Design Parameters** tab.



Now go to the **Boundary Info** tab, and click  . Change the **Boundary Scale** to **p-value**. Look at the display of the stopping boundary in p-value scale. In this case the penalty imposed by the efficacy boundary has overcome the reward imposed by the futility boundary and the cut-off for the final p-value required to reject H_0 and declare

61 Early Stopping for Futility

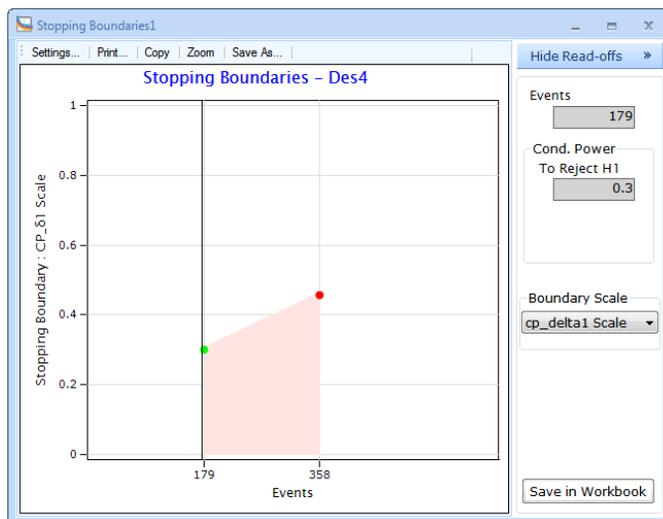
statistical significance is less than $\alpha = 0.025$.



61.5.3 More Conservative Futility Boundaries

It is useful to view futility stopping boundaries on a conditional power scale since that permits us to directly compare a formal futility boundary with an alternative informal early stopping criterion where both criteria are based on low conditional power. Select Des 4 in the **Library**, and then select **Stopping Boundaries** after clicking the 

icon. Select **cp_delta1 Scale**.



We are required to terminate the study at the first interim look if the conditional power is less than 0.2999, and at the second interim look if the conditional power is less than 0.4581. These are fairly large conditional power values. The trial investigators might not be willing to commit in advance to stop the study and declare futility if the conditional power is as high as 45%. Consequently, they might prefer to adopt an informal approach to early stopping for futility. However, as we have already discussed, the informal approach cannot ensure that the type-2 error will be preserved and the study might lose power.

The availability of a rich family of flexible spending functions in East enables us to pick formal futility boundaries with substantially lower conditional power for futility stopping, within the range of conditional power values that we might use with the informal approach. For example, suppose that the trial investigators do not wish to terminate this trial for futility unless the conditional power is less than 20% at the first interim look, and less than 10% at the second interim look. These rather conservative criteria for early stopping are more realistic than the 30% and 45% conditional power criteria implied by the stopping boundaries of Des 4. Close this chart before continuing.

Gamma family β spending function Create a new design by selecting Des 4 in the

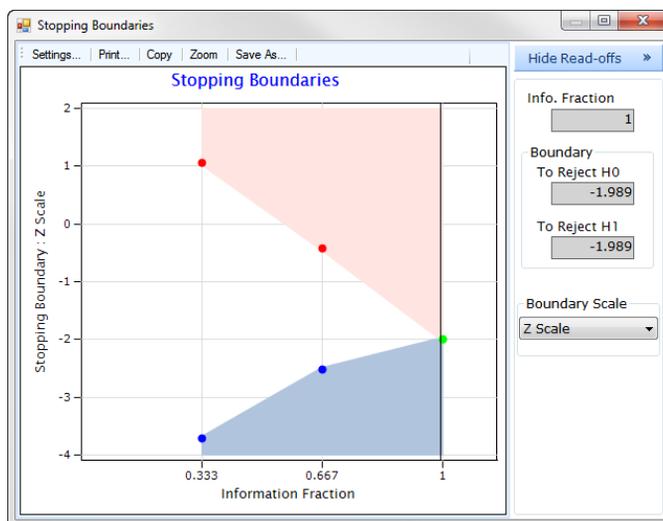
61 Early Stopping for Futility

Library, and then by clicking the  icon on the **Library** toolbar. Click the **Boundary Info** tab. In the **Futility** box, change the **Spending Function** to **Gamma Family** from the drop-down list.

We must choose a parameter value, γ , to identify a specific member of this family. The value $\gamma = -4$ will yield a spending function roughly similar to the **LD (OF)** spending function. Smaller values of γ will yield more conservative spending functions. Since the **LD (OF)** function (which was used to spend type-2 error in Des 4) yielded unsatisfactory futility boundaries on the conditional power scale, let us be more conservative. Type in -6 as the value of **Parameter** γ . Select the radio-button next for **Binding**.

| Test Parameters | Boundary | Accrual / Dropouts |
|----------------------------|---|--------------------|
| Efficacy | | |
| Boundary Family: | Spending Functions | |
| Spending Function: | Lan-DeMets | |
| Parameter: | OF | |
| Type I Error (α): | 0.025 | |
| Futility | | |
| Boundary Family: | Spending Functions | |
| Spending Function: | Gamma Family | |
| Parameter (γ): | -6 | |
| Type II Error (β): | 0.1 | |
| | <input type="radio"/> Non-Binding <input checked="" type="radio"/> Binding | |

Click  to show the boundary chart.

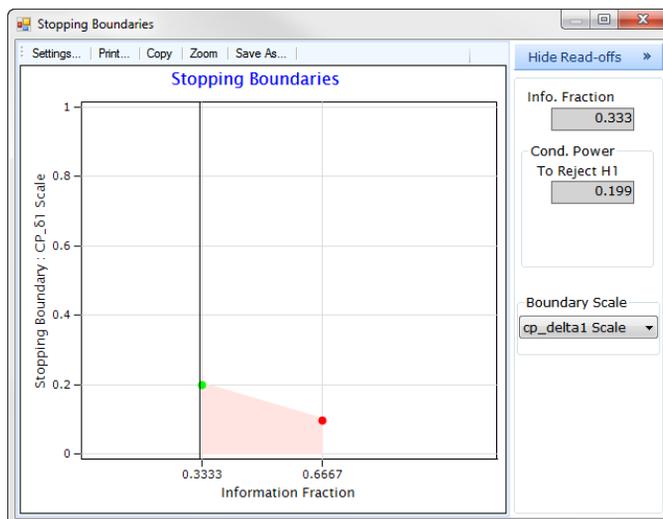


The stopping boundary for rejecting H_0 at the final look is now -1.9889 . As this value

is smaller than -1.96 there is indeed a penalty being paid for the multiple looks. Thus, the psychological difficulty encountered in Des 4, where the final stopping boundary for rejecting H_0 was less than -1.96, has been resolved. Click **Compute** to generate Des 5.

We can try to be more conservative in terms of β spending function. Change the parameter for the Gamma spending function from $\gamma = -6$ to $\gamma = -8$

Click  and change the **Boundary Scale** to the **cp_delta1 Scale**.

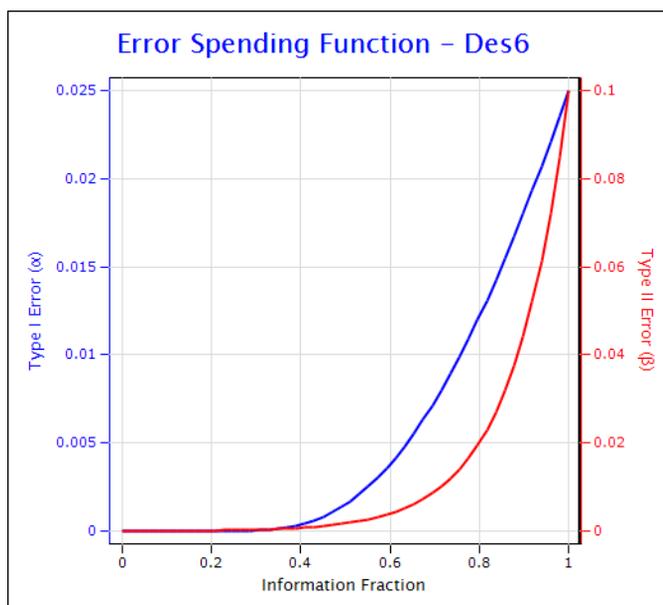


By viewing the futility boundary on the cp delta scale, the first and second-look values of conditional power required to stop early are, respectively, 0.1991 and 0.0973. These values are within a range where the trial investigators would be willing to guarantee in advance that they would stop the trial and declare futility. The advantage of using the formal futility boundary is, of course, that the type-2 error (and hence the power) is guaranteed to be preserved. Click **Compute**. This will add a new row to the **Output Preview** labeled as Des 6. Save this design in the current workbook by selecting the row corresponding to Des 6 in **Output Preview** and clicking  on the **Output Preview** toolbar.

61 Early Stopping for Futility

61.6 Early Stopping for Futility Only

Under Des 6 there is the possibility of rejecting H_0 and stopping early for efficacy if the upper stopping boundary is crossed. The α -spending function used to generate the upper efficacy stopping boundary is the **LD (OF)** spending function proposed by Lan and DeMets (1983). This function is popular because it spends the type-1 error conservatively in the beginning, but still provides a reasonable opportunity for premature termination once the trial gets underway. In contrast, the **Gm (-8)** β -spending function used by Des 6 to generate the futility boundary is much more conservative and provides considerably less opportunity for premature termination until the study close to completion. To examine these two spending functions together, first select Des 6 in **Library**. Click  in the **Library** toolbar and then select **Error Spending**.



For the first 40% of the trial, both spending functions are extremely conservative, spending a negligible amount of error. Thereafter, however, the α -spending function starts to spend the type-1 error at a much faster rate making it easier to stop early for efficacy. Let us examine the stopping probabilities for Des 6 under H_0 and H_1 . Select

Des 6 in the **Library** and double-click on it.

⊖ Events, Sample Size, Pipeline and Analysis Times: Look by Look (Under H0)

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (s) | Pipeline (n-s) | Analysis Time | Boundary Crossing Probability (Cumulative) | |
|--------|--------------------------|-----------------|------------|----------------|---------------|--|----------|
| | | | | | | Efficacy | Futility |
| 1 | 0.334 | 575 | 175 | 400 | 11.977 | 1.051E-4 | 0.076 |
| 2 | 0.666 | 840 | 349 | 491 | 17.857 | 0.006 | 0.576 |
| 3 | 1 | 840 | 524 | 316 | 24.66 | 0.025 | 0.975 |

⊖ Events, Sample Size, Pipeline and Analysis Times: Look by Look (Under H1)

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (s) | Pipeline (n-s) | Analysis Time | Boundary Crossing Probability (Cumulative) | |
|--------|--------------------------|-----------------|------------|----------------|---------------|--|----------|
| | | | | | | Efficacy | Futility |
| 1 | 0.334 | 611 | 175 | 436 | 12.718 | 0.034 | 4.501E-4 |
| 2 | 0.666 | 840 | 349 | 491 | 18.989 | 0.56 | 0.007 |
| 3 | 1 | 840 | 524 | 316 | 26.87 | 0.9 | 0.1 |

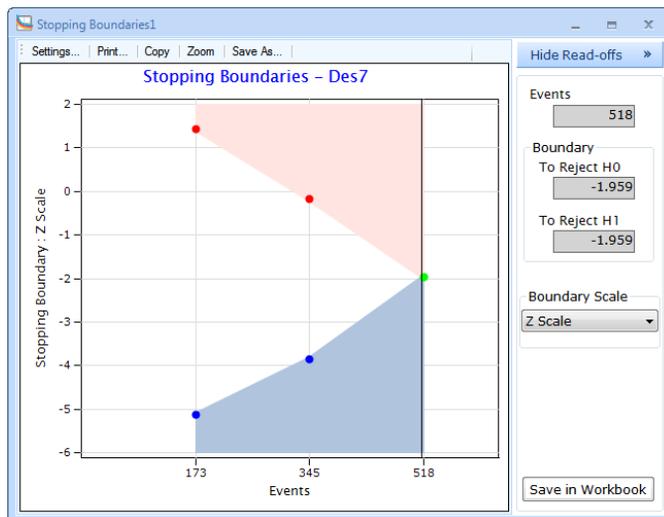
Under H_1 , the efficacy boundary would be crossed with probability 0.034 at the first interim analysis, one-third-way through the trial. By the time two-thirds of the trial has been completed, the probability of early stopping for efficacy under H_1 at the second look is 0.56 (cumulative). In some studies, however, the investigators have no desire to stop early for efficacy, but only wish to stop early for futility. Early efficacy stopping for a promising new therapy might not be desirable, for instance, if the investigators wish to continue the trial and monitor safety. Early futility stopping under H_0 , on the other hand, is desirable since it is better to kill a study that is going nowhere and spend the resources elsewhere. We can discourage early efficacy stopping by using stopping boundaries that are considerably more conservative than the **LD (OF)** boundary used in Des 6. Let us consider using the Gamma spending function with parameter $\gamma = -18$.

Create a new design by selecting Des 6 in the **Library**, and clicking the  icon on the **Library** toolbar. Click the **Boundary Info** tab. In the **Efficacy** box, change the **Spending Function** to **Gamma Family** from the drop-down list. Type in -18 as the value of **Parameter** (γ), and click **Compute**

This will add a row in the **Output Preview** with label Des 7. Select Des 7 by clicking anywhere along the row in the **Output Preview** and click  to save this design in the **Library**. Select Des 7 in the **Library**, and click , then select **Stopping**

61 Early Stopping for Futility

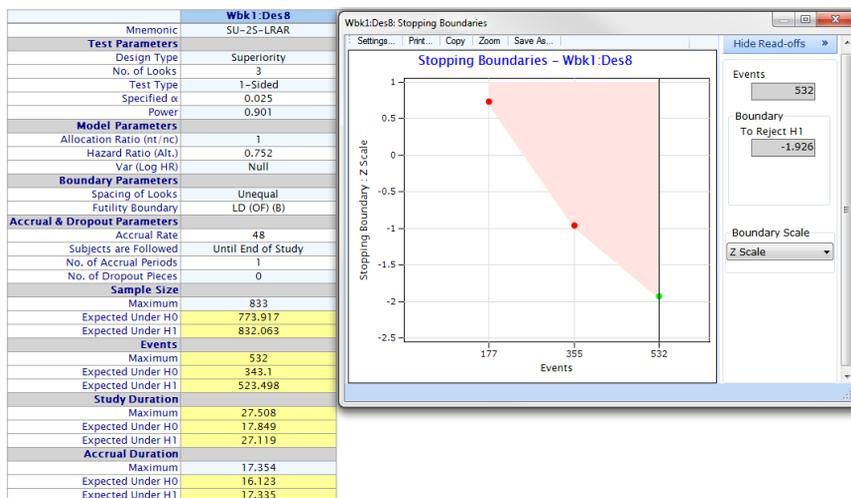
Boundaries.



Notice how hard it is to stop early for efficacy. Even as late as the second interim look the efficacy boundary value is -3.841 on the standardized difference scale. We would need to see a one-sided p-value smaller than 0.0001 in order to stop early for efficacy. Thus, except in very extreme situations, Des 7 will not permit early stopping for efficacy. An interesting feature of Des 7 is that the p-value required at the final look in order to reject the null hypothesis and declare statistical significance is 0.0251 . Although we have designed the study for multiple looks at the data, the cut-off p-value for rejecting H_0 at the final look is greater than $\alpha = 0.025$; i.e., we have been rewarded rather than penalized for the multiple looks. We explained the reason for this seeming anomaly in Section 61.5.2. The final cut-off p-value required to preserve the type-1 error is determined by balancing the penalty due to the presence of an efficacy boundary against the reward due to the presence of a futility boundary. Because of the specific choice of γ parameters, this balance ended up favoring a tiny reward. It might however be important, in an industry trial, to obtain the approval of the regulatory reviewers for using 0.0252 as the final cut-off for rejecting H_0 . The simulation tools of East may be used to demonstrate that this cut-off does indeed preserve the type-1 error. Close the chart before continuing.

It would be interesting to compare Des 7 with a design that has the same futility boundary but no efficacy boundary whatsoever. To achieve this aim create a new design by selecting Des 7 in the **Library**, and clicking the  on the **Library**

toolbar. Click the **Boundary Info** tab. In the **Efficacy** box, change the **Boundary Family** to **None** from the drop-down list, and click **Compute**. This will add a row in the **Output Preview** with label Des 8. The design summary and stopping boundaries (futility only) of this design are displayed below.



In this design, the trial stops for futility if the value of the test statistic is less than the corresponding boundary value. The value of the boundary at the final look is -1.9264. Therefore if the value of the test statistic is less than -1.9264 at the final look one could technically reject H_0 and claim efficacy. The type-1 error of this procedure is 0.025 even though this final boundary value is greater than -1.96. This follows from the same reasoning as we provided in Section 61.5.2. The type-1 error is decreased because there is a chance of being absorbed into the futility boundary at an earlier look. To compensate, the critical value of the test statistic for rejecting H_0 at the final look is determined to be -1.9264 rather than -1.96.

62 *Flexible Stopping Boundaries in East*

East provides considerable flexibility for generating stopping boundaries with different shapes and varying levels of conservatism for early stopping for efficacy, safety or futility. Suppose, for instance that a trial will be monitored at regular intervals for safety. For ethical reasons, one might wish to choose safety stopping boundaries that possess a very low threshold for early stopping. On the other hand, there might be some reluctance to stopping a trial early for efficacy. If the new treatment looks promising there is often a desire to go to completion and thereby gather overwhelmingly strong evidence of treatment benefit rather than stopping prematurely. In that case, one might wish to choose extremely conservative stopping boundaries with a high threshold for early stopping at the early interim looks. The boundaries that are available in East run the gamut between extreme conservatism and extreme liberality for early stopping. They fall into three main categories: p-value boundaries, power boundaries and spending function boundaries. Furthermore, a boundary may serve either to stop a trial and reject the null hypothesis or to stop a trial and reject the alternative hypothesis. Boundaries that facilitate early stopping to reject the null hypothesis are by far the more common of the two types. They are further classified into efficacy boundaries and safety boundaries. Boundaries that facilitate early stopping to reject the alternative hypothesis are known as futility boundaries. They play a role in early termination of trials in which the treatment effect is too small to confer a therapeutic advantage to the experimental arm. They may be used either in conjunction with, or as an alternative to, conditional power for futility stopping.

P-value boundaries are discussed in Section [62.1](#). Power boundaries are discussed in Section [62.2](#). As originally conceived of, p-value boundaries offer less flexibility than power boundaries in terms of boundary shape. However, as described in Section [62.1](#), p-value boundaries have been generalized in this version of East to accommodate many more situations. Still, spending function boundaries offer the most flexibility for trial design. They are discussed in Section [62.3](#). Our recommendation is to use the spending function boundaries whenever possible.

The theory underlying the actual construction of stopping boundaries is developed in Appendix [B](#). The purpose of the present chapter is to document how the various boundaries can be invoked in East and to demonstrate, through examples, the flexibility they confer for trial design.

62.1 P-Value (or Haybittle-Peto) Boundaries

62.1.1 Use of Haybittle-Peto boundaries

62.1.2 SPARCL trial

P-value boundaries, also known as Haybittle-Peto boundaries, have a very simple structure. One specifies a fairly small p-value, say 0.0001, for early stopping at the first $K - 1$ looks. East then uses recursive integration to compute the last-look p-value needed to achieve an overall type-1 error of α . Historically these boundaries were conceived by Haybittle (1971) as a fairly straightforward way of being permitted to take interim looks without having any substantial impact on the final p-value one would need in order to attain a statistically significant outcome.

In East, we have generalized the original Haybittle-Peto boundaries so that the p-values specified at the first $K - 1$ looks need not be equal. We call such boundaries Generalized Haybittle-Peto boundaries. The following two examples illustrate how to use the original and the generalized Haybittle-Peto boundaries in East. In addition to designing a trial with these types of boundaries, the second example shows how such a trial can be simulated and monitored using East.

62.1.1 Use of Haybittle-Peto boundaries in a hypertension trial

A randomized, placebo-controlled trial were conducted to evaluate the efficacy of arthroscopy for osteoarthritis of the knee (Moseley et al., 2002). Primary endpoint was patient-reported pain in the study knee 24 months after intervention on a scale range from 0 to 100, with higher score indicating the more sever pain. Let $X_{ic} \sim N(\mu_c, \sigma^2)$ be the pain score for the i th subject in the placebo group, $X_{it} \sim N(\mu_t, \sigma^2)$ be the pain score for the i th subject in the treatment group, and $\delta = \mu_t - \mu_c$. Null hypothesis was that the patients in the two groups report the same amount of knee pain after two years. That is, $H_0 : \mu_t = \mu_c$. The trial was designed to detect a moderate effect size $\delta_1 = 0.55$ with 90% power with a two-sided level-0.04 test. This was the group-sequential design with Haybittle-Peto stopping boundaries of $p=0.001$ for the interim analyses. For this study, the standard deviation for placebo arm was reported as 18.5 and we will use this as common standard deviation for both the group. We will illustrate designing of this study considering maximum of $K=4$ equally spaced looks.

First, click **Continuous: Two Samples** on the **Design** tab, and then click **Parallel Design: Difference of Means**. The upper pane of this window displays several fields with default values. First, change the **Number of Looks** to 4. This will add a tab with label **Boundary Info**. We will come back to this tab later. In the **Design Parameters** tab, select **Superiority** for **Design Type** and **2-Sided** for **Test Type**. Since the study was planned to detect a moderate effect size of 0.55, select **Standardized Diff. of Means** for **Input Method** and specify **Standardized Diff.** $((\mu_t - \mu_c)/\sigma)$ as 0.55. Enter 0.04 for **Type I Error (α)**, and 0.9 for **Power ($1-\beta$)**. The

62 Flexible Stopping Boundaries in East

Design Parameters tab should appear as below:

Design Type: Superiority Number of Looks: 4

Test Parameters Boundary

Test Type: 2-Sided Input Method: Standardized Diff. of Means Test Statistic: Z

Type I Error (α): 0.04 Specify Alternative Hypothesis

Power: 0.9 Standardized Diff. $((\mu_t - \mu_c)/\sigma)$: 0.55

Sample Size (n): Computed

Allocation Ratio: 1

(n_t/n_c)

Click the **Boundary Info** tab. In this tab, you will see **Efficacy** and **Futility** boxes, where you can select efficacy and futility boundary families. Select **Haybittle Peto (p-value)** for **Boundary Family** in the **Efficacy** box and select **None** for **Boundary Family** in the **Futility** box.

For the Haybittle Peto boundary family, East allows you to fix either overall type I error or the p-value at the final look. In both the cases, p-values for the interim looks need to be specified. To use the original Haybittle-Peto boundaries, all the interim looks should have equal p-value.

Fixed p-value at final look First we will illustrate how to fix the p-value at the final look instead of overall type I error (α). This is the case when one would like to specify a constant p-value boundary at the first 3 looks as well as any desired final p-value boundary for the 4th look. Suppose, for example, that we specify 0.001 at each of the first 3 looks and 0.04 at the 4th look. Select the radio-buttons corresponding to the **Last Look p-value**, and **Unequal p-values at looks**. The **Boundary Info** tab should

appear as below:

Test Parameters | **Boundary**

Efficacy

Boundary Family: Haybittle Peto (p-value)

Fix

Total Type I Error (α) p-values at looks Equal

Last Look p-value Unequal

Type I Error (α): 0.041

Spacing of Looks

Equal Unequal

| Look # | Info. Fraction | Efficacy p-value |
|--------|----------------|------------------|
| 1 | 0.250 | 0.001 |
| 2 | 0.500 | 0.001 |
| 3 | 0.750 | 0.001 |
| 4 | 1.000 | 0.04 |

Click **Compute**. This will add a row to the **Output Preview** with label Des 1.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | Input Method | Standardized Diff. | Test Statistic |
|------|-------------|--------------|-----------|--------------------|-------|-------|------------------|-------------------|-------------|------------------|------------------|-----------------------------|--------------------|----------------|
| Des1 | Superiority | 4 | 2-Sided | 0.041 | 0.9 | 1 | Unequal | HP | 147 | 146.801 | 124.507 | Standardized Diff. of Means | 0.55 | Z |

The overall type I error is now 0.041, which is slightly higher than the desired type I error of 0.04. The increase in overall power is due to the 3 interim looks. Maximum sample size required for this design is 147.

Fixed overall type I error Recall that the study we are considering in this section was designed to maintain an overall type I error of 0.04 with constant Haybittle-Peto boundaries of $p=0.001$ for the interim analyses. In **Boundary Info** tab, select the radio-buttons corresponding to the **Total Type I Error (α)**, and **Unequal p-values at looks**. Then go to the **Design Parameters** tab and set the **Type I error (α)** at 0.04.

62 Flexible Stopping Boundaries in East

The **Boundary Info** tab should appear as below:

Boundary

Efficacy

Boundary Family: Haybittle Peto (p-value) ▾

Fix

Total Type I Error (α) p-values at looks

Last Look p-value Unequal

Type I Error (α): 0.04

Spacing of Looks

Equal Unequal

| Look # | Info. Fraction | Efficacy p-value |
|--------|----------------|------------------|
| 1 | 0.250 | 0.001 |
| 2 | 0.500 | 0.001 |
| 3 | 0.750 | 0.001 |
| 4 | 1.000 | 0.039 |

The p-value corresponding to final look has been updated to 0.0391. Upon clicking the **Compute** button, we will see that a maximum sample size of 148 would be needed.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | Input Method | Standardized Diff. | Test Statistic |
|------|-------------|--------------|-----------|--------------------|-------|-------|------------------|-------------------|-------------|------------------|------------------|-----------------------------|--------------------|----------------|
| Des1 | Superiority | 4 | 2-Sided | 0.041 | 0.9 | 1 | Unequal | HP | 147 | 146.801 | 124.507 | Standardized Diff. of Means | 0.55 | Z |
| Des2 | Superiority | 4 | 2-Sided | 0.04 | 0.901 | 1 | Unequal | HP | 148 | 147.799 | 125.148 | Standardized Diff. of Means | 0.55 | Z |

It is more common to use Haybittle-Peto boundaries as shown in Des 2 – to specify a common p value for the first $K - 1$ looks, and adjust the final p value to satisfy an overall α . To see the sample size required for a single look design, change the **Number of Looks** to 1. Click **Compute** to obtain the fixed sample size.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | Input Method | Standardized Diff. | Test Statistic |
|------|-------------|--------------|-----------|--------------------|-------|-------|------------------|-------------------|-------------|------------------|------------------|-----------------------------|--------------------|----------------|
| Des1 | Superiority | 4 | 2-Sided | 0.041 | 0.9 | 1 | Unequal | HP | 147 | 146.801 | 124.507 | Standardized Diff. of Means | 0.55 | Z |
| Des2 | Superiority | 4 | 2-Sided | 0.04 | 0.901 | 1 | Unequal | HP | 148 | 147.799 | 125.148 | Standardized Diff. of Means | 0.55 | Z |
| Des3 | Superiority | 1 | 2-Sided | 0.04 | 0.902 | 1 | | | 148 | | | Standardized Diff. of Means | 0.55 | Z |

The sample size for Des 2 and that of Des 3, the fixed sample plan, are nearly the same as shown above. This was the original motivation for Haybittle-Peto boundaries. They are easy to specify, permit interim looks with very little chance of stopping the trial, and resemble the fixed sample trial at the final look.

62.1.2 Use of the Generalized Haybittle-Peto boundaries in the SPARCL

trial

The Stroke Prevention by Aggressive Reduction in Cholesterol Levels (SPARCL) group of investigators conducted a large multi-center placebo-controlled trial to evaluate the safety and efficacy of High-Dose Avorstatin after Stroke or Transient Ischemic Attack (TIA) (SPARCL, 2006). The primary hypothesis of the study was that treatment with 80 mg of Avorstatin per day would reduce the risk of fatal or non-fatal stroke among patients with a history of stroke or TIA. The study was designed to have a statistical power of 90% to detect an absolute one third increase in the primary endpoint (time to first fatal or non-fatal stroke) in the Avorstatin group as compared with the placebo group during a median follow-up of five years with a two-sided significance level of 5%. The assumed annual rate in the placebo group was 3.5% or a cumulative survival rate of 96.5%. Seven interim analyses of efficacy were planned with a stopping boundary corresponding to a two-sided significance level of $p_1 = 0.0001$ for the first analysis and $p_j = 0.001, j = 2, \dots, 7$ thereafter. Patients were enrolled between September 1998 and March 2001 for a total of 4200 (implying an accrual rate of 140 patients per month).

Trial Design

Using the generalized Haybittle-Peto boundaries available in East, we will now design this trial.

Start East afresh. Click **Survival: Two Samples** on the **Design** tab and then click **Parallel Design: Logrank Test Given Accrual Duration and Accrual Rates**. Set the **Number of Looks** to 8, to generate a study with seven interim looks and a final analysis.

In the **Design Parameters** tab, select **Design Type** as **Superiority**, **Test Type** as **2-Sided**, and enter **Type I Error (α)** and **Power ($1-\beta$)** as 0.05 and 0.9, respectively. Leave the **# of Hazard Pieces** as 1, which implies that hazard rates remain constant overtime in both Avorstatin and placebo groups. Change the **Input Method** to **Cum. % Survival**. Tick the check box for **Hazard Ratio (Optional)**, select the radio-button for **Hazard Ratio (λ_t/λ_c)** and enter 0.75. Finally, the **Cum. % Survival (Control)**

62 Flexible Stopping Boundaries in East

should be 96.5 at 12 months. The **Design Parameters** tab should appear as below:

Design Type: Superiority Number of Looks: 8

Test Parameters Boundary Accrual / Dropouts

Test Type: 2-Sided # of Hazard Pieces: 1 Input Method: Cum. % Survival

Type I Error (α): 0.05 Hazard Ratio (Optional) Alternative

Power: 0.9 Hazard Ratio (λ_1/λ_2) 0.75

No. of Events: Computed Ratio of % Survivals at Period # 1 (S_1/S_2) 1.009

Allocation Ratio: 1 (n_1/n_2)

| | Cum. % Survival | |
|----|-----------------|-----------------|
| | Control | Treatment: Alt. |
| 12 | 96.5 | 97.363 |

Note: Period 1 hazard rates apply after time 12.

Variance of Log Hazard Ratio

Null Alternative

Move to the **Boundary Info** tab. Select **Haybittle Peto (p-value)** for **Boundary Family in Efficacy**, and the radio-buttons corresponding to the **Total Type I Error (α)**, and **Unequal p-values at looks**. Enter the p-value as 0.0001 for the first look, and 0.001 for the next six looks, and click **Recalc**. The **Boundary Info** tab should appear as below:

Test Parameters Boundary Accrual / Dropouts

Efficacy Futility

Boundary Family: Haybittle Peto (p-value)

Fix p-values at looks

Total Type I Error (α) Equal

Last Look p-value Unequal

Type I Error (α): 0.05

Spacing of Looks Equal Unequal

| Look # | Info. Fraction | Efficacy p-value |
|--------|----------------|------------------|
| 1 | 0.125 | 0.0001 |
| 2 | 0.250 | 0.001 |
| 3 | 0.375 | 0.001 |
| 4 | 0.500 | 0.001 |
| 5 | 0.625 | 0.001 |
| 6 | 0.750 | 0.001 |
| 7 | 0.875 | 0.001 |
| 8 | 1.000 | 0.0488 |

The p-value corresponding to the final look has been updated to 0.0488.

Finally, move to the **Accrual / Dropout Info** tab. Select 1 for **# of Accrual Periods**, and enter 140 in the **Accrual Rate** column, and change the **Comtd.** number of subjects to 4200.

| Accrual | | | |
|--|------|--------|------------|
| | Min. | Comtd. | Sugg. Max. |
| <input type="radio"/> Duration: | 3.65 | 30 | 54.221 |
| <input checked="" type="radio"/> Subjects: | 511 | 4200 | 7591 |

Click **Compute**. Select Des 1 in **Output Preview** and click the  icon. This will

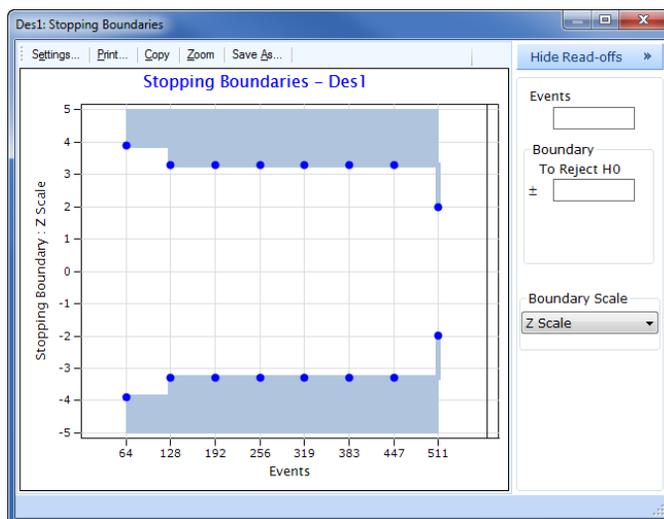
62 Flexible Stopping Boundaries in East

display the design details in the **Output Summary**.

| | Des 1 |
|---|--------------------|
| Mnemonic | SU-25-LRAR |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 8 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.901 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Hazard Ratio (Alt.) | 0.75 |
| Var (Log HR) | Null |
| Boundary Parameters | |
| Spacing of Looks | Unequal |
| Efficacy Boundary | HP |
| Accrual & Dropout Parameters | |
| Accrual Rate | 140 |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| Sample Size | |
| Maximum | 4200 |
| Expected Under H0 | 4199.166 |
| Expected Under H1 | 4174.983 |
| Events | |
| Maximum | 511 |
| Expected Under H0 | 509.999 |
| Expected Under H1 | 418.798 |
| Study Duration | |
| Maximum | 65.103 |
| Expected Under H0 | 58.718 |
| Expected Under H1 | 55.768 |
| Accrual Duration | |
| Maximum | 30 |
| Expected Under H0 | 29.994 |
| Expected Under H1 | 29.821 |

According to this design, 511 events are needed to appropriately power the study. Select Des 1 in **Output Summary**, click  , and select **Stopping Boundaries**.

The boundaries on the Z-scale are shown below:



62.2 Power Boundaries

62.2.1 Wang-Tsiatis Boundaries

62.2.2 Pampallona-Tsiatis Boundaries

East provides two types of power boundaries – Wang-Tsiatis boundaries (Wang and Tsiatis, 1987) for early rejection of H_0 , and Pampallona-Tsiatis boundaries (Pampallona and Tsiatis, 1994) for early rejection of H_0 or H_1 .

62.2.1 Wang-Tsiatis Boundaries

The Wang-Tsiatis boundaries permit early stopping to reject H_0 . They are used to stop a trial early for efficacy only (1-sided boundaries), safety only (1-sided boundaries) or to stop early either for efficacy or safety (two-sided case).

Group sequential boundaries of this type were first proposed by Pocock (1977) and O’Brien and Fleming (1979). Subsequently Wang and Tsiatis (1987) incorporated both the Pocock and O’Brien-Fleming boundaries into a family of “power boundaries” characterized by a shape parameter Δ . For a K -look group sequential trial the power boundary for the standardized test statistic Z_j at look j is of the form

$$c_j = C(\Delta, \alpha, K)t_j^{\Delta-0.5}, \quad j = 1, 2, \dots, K,$$

where $t_j = n_j/n_{\max}$, n_j is the sample size at look j , and n_{\max} is the maximum sample size we must commit up-front to this study in order to achieve the desired power. For technical details on the computation of $C(\Delta, \alpha, K)$ refer to Appendix B. The study is terminated, and the null hypothesis rejected, the first time that $Z_j > c_j$ for

62 Flexible Stopping Boundaries in East

one-sided tests and $|Z_j| > |c_j|$ for two-sided tests. The constant $C(\Delta, \alpha, K)$ is computed by recursive integration as described in Appendix F. When $\Delta = 0$, the stopping boundaries decrease in proportion to the square root of the current information fraction, these are the O'Brien-Fleming boundaries. When $\Delta = 0.5$, the stopping boundaries are constant at each look, these are the Pocock boundaries.

East permits shape parameters in the range $-0.5 < \Delta < 0.5$. The smaller the value of Δ , the more difficult it is to stop the trial at an interim look. The maximum sample size requirements increase progressively with increasing values of the shape parameter Δ . On the other hand, the expected sample sizes under the alternative hypothesis decrease with increasing values of Δ . Depending on availability of patients and the importance to the trial sponsor of trading off a larger maximum sample size commitment in exchange for a smaller expected sample size, one can select an appropriate value of Δ .

62.2.2 Pampallona-Tsiatis Boundaries

The Wang-Tsiatis power boundaries were developed for early stopping to reject H_0 . Subsequently Pampallona and Tsiatis (1994) extended these power boundaries to cover the case of early stopping to reject either H_0 or H_1 . The Pampallona-Tsiatis boundaries are characterized by two shape parameters, Δ_1 for the boundaries that facilitate early rejection of H_0 and Δ_2 for the boundaries that facilitate early rejection of H_1 . At the j th look the boundaries for early stopping to reject of H_0 are of the form

$$c_j = C_1(\Delta_1, \alpha, \beta, K),$$

and the boundaries for early stopping to reject H_1 are of the form

$$c_j = C_2(\Delta_2, \alpha, \beta, K) - \delta_1 \sqrt{n_j}$$

where $1 - \beta$ is the power and δ_1 is the treatment effect under H_1 . For technical details on the computation of $C_1(\cdot)$ and $C_2(\cdot)$ refer to Appendix B.

The one-sided version consists of a pair of boundaries that meet at the last look. In their most common application, one member of the pair facilitates stopping early for efficacy by rejecting H_0 and the other member facilitates stopping early for futility by rejecting H_1 . The two-sided version consists of a pair of outer boundaries and an inner wedge. Usually one outer boundary is for early stopping to reject H_0 in favor of efficacy and the other outer boundary is used for early stopping to reject H_0 and conclude that the new treatment is worse than the standard, hence that it is unsafe. If the test statistic enters the inner wedge, the alternative hypothesis H_1 is rejected and the trial stops for futility. We shall discuss efficacy, safety and futility stopping boundaries in greater detail in the next section where we introduce the spending function boundaries.

62.3 Spending Function Boundaries

The most general way to generate stopping boundaries is through α - and β -spending functions. The idea of using an α -spending function to derive stopping boundaries for early rejection of H_0 was first introduced in a landmark paper by Lan and DeMets (1983). Subsequently, Pampallona, Tsiatis and Kim (1995), (2001) developed the notion of a β -spending function to derive stopping boundaries for early rejection of H_1 . In East, one may use an α -spending function to generate efficacy, safety or non-inferiority boundaries and a β -spending function to generate futility boundaries. Also one may combine both α - and β -spending in a single trial, with one-sided or two-sided boundaries. All these options are discussed in the sections below. The theory underlying these spending functions is given in Appendix C.

62.3.1 The Alpha Spending Function

Suppose the type-1 error of a trial is fixed at α . An α -spending function is any monotone function of the information fraction $t \in [0, 1]$, with $\alpha(t) = 0$ and $\alpha(1) = \alpha$. The value $\alpha(t)$ may be interpreted as the probability, under H_0 , of crossing a stopping boundary by time t ; i.e., of committing a type-1 error by time t . Thus one can think of the α -spending function as a way of budgeting how the overall type-1 error is to be spent over the course of the trial.

Lan-DeMets Spending Functions

A conservative spending function will spend the type-1 error very sparingly in the beginning but will rapidly increase the pace of spending as the trial nears completion. An example of such a spending function, proposed by Lan and DeMets (1983) for two-sided tests, has the functional form

$$\alpha(t) = 4 - 4\Phi\left(\frac{z_{\alpha/4}}{\sqrt{t}}\right). \quad (62.1)$$

We shall see that this spending function generates stopping boundaries that are very similar to the O'Brien-Fleming boundaries. The function is displayed below. Notice how slowly the α is spent in the early phase of the trial. In East we use the mnemonic **LD (OF)** to denote this spending function where **LD** stands for Lan-DeMets and **OF** stands for O'Brien-Fleming.

Lan and DeMets (1983) proposed the following function for spending the type-1 error more aggressively.

$$\alpha(t) = \alpha \ln\{1 + (e - 1)t\} \quad (62.2)$$

This function is displayed below. Notice that it is a concave function. We shall see that this function generates stopping boundaries that closely resemble the Pocock boundaries. In East we use the mnemonic **LD (PK)** to denote this spending function where **LD** stands for Lan-DeMets and **PK** stands for Pocock.

62 Flexible Stopping Boundaries in East

At any time t that an interim look is taken, it is possible to invert the corresponding value of the $\alpha(t)$ and thereby generate the stopping boundary. Suppose, for instance, that a study is designed for two interim looks and one final look, at information fractions t_1, t_2 and $t_3 = 1$, not necessarily equally spaced. The two-sided symmetric boundary $\pm c_1$ at look-1 is obtained as the solution to

$$P_0(|Z(t_1)| \geq |c_1|) = \alpha(t_1)$$

Having already utilized $\alpha(t_1)$ of the total available error to compute c_1 , one can generate c_2 recursively as the solution to

$$\alpha(t_1) + P_0(|Z(t_1)| < |c_1|, |Z(t_2)| \geq |c_2|) = \alpha(t_2)$$

At the time of the last look, we will have utilized $\alpha(t_2)$ of the total available error and will know the values of the first two stopping boundaries, c_1 and c_2 . Thus, the final stopping boundary, c_3 , is obtained recursively as the solution to

$$\alpha(t_2) + P_0(|Z(t_1)| < |c_1|, (|Z(t_2)| < |c_2|, |Z(t_3)| \geq |c_3|) = \alpha.$$

Notice from the above that the probability of crossing a boundary for the first time at either the first, second or third looks is

$$\alpha(t_1) + [\alpha(t_2) - \alpha(t_1)] + [\alpha - \alpha(t_2)] = \alpha \quad (62.3)$$

In other words, this strategy for generating the stopping boundaries is guaranteed to preserve the type-1 error.

We will now see how to obtain stopping boundaries in East based on α spending. Suppose we want to generate two-sided stopping boundaries based on three equally spaced looks, derived from the **LD (OF)** spending function specified by equation (62.1).

Start East afresh. Click **Continuous: Two Samples** on the **Design** tab, and then click **Parallel Design: Difference of Means**. Change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. Accept the default values in **Design Parameters** tab. Move to the **Boundary Info** tab. Select **Spending Functions** for **Boundary Family**, **Lan-DeMets** for **Spending Function** and **OF** for **Parameter** in **Efficacy** box. In the **Futility** box, select **None** for **Boundary**

Family. Stopping boundaries will be displayed in the table below in this tab.

Design Parameters
Boundary Info

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.025

Futility

Boundary Family: None

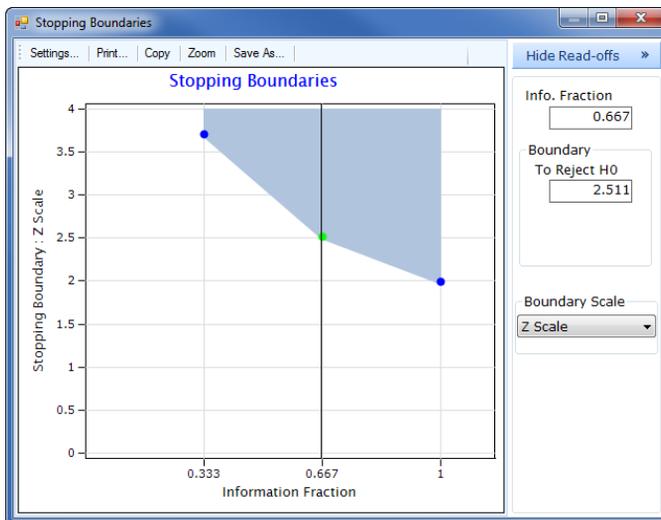
Spacing of Looks

Equal Unequal

Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.333 | 0.000 | 3.710 |
| 2 | 0.667 | 0.006 | 2.511 |
| 3 | 1.000 | 0.025 | 1.993 |

Click the  icon. This will show the boundary chart on the Z scale.



The stopping boundaries closely resemble the O’Brien-Fleming boundaries discussed in Section 62.2.1. East allows us to see stopping boundaries on different scales: Select from different options in the drop-down list under **Boundary Scale**.

Now compare these charts with those from Pocock-like boundaries. Change the

62 Flexible Stopping Boundaries in East

Parameter to **PK** in **Boundary Info** tab, and click  . The stopping boundaries derived from the **LD (PK)** spending function specified by equation (62.2) closely resemble the Pocock stopping boundaries.

Although one usually specifies the number and timing of the interim looks at the design stage, it might not be administratively convenient to adhere to these two design parameters at the interim monitoring stage. The great appeal of the spending function approach for regulatory purposes is that it gives us the freedom to alter both the number and timing of the interim looks while still preserving the overall type-1 error, α . Suppose, for instance that we were to introduce an unplanned interim analysis in between the second and third looks. Thus, suppose that a total of four looks were taken, even though the study was designed for only three looks. Let these looks be taken at times t'_1, t'_2, t'_3 , and t'_4 , where these times need not be the same as any of the three time points t_1, t_2, t_3 specified at the design stage. If we use the above recursive method to compute the stopping boundaries c'_1, c'_2, c'_3 , and c'_4 at the four looks, the probability of crossing a stopping boundary must be

$$\alpha(t'_1) + [\alpha(t'_2) - \alpha(t'_1)] + [\alpha(t'_3) - \alpha(t'_2)] + [\alpha(t'_4) - \alpha(t'_3)] = \alpha(t'_4) \leq \alpha$$

For further details and for a discussion of how to compute sample size for a given power using spending function boundaries, refer to Appendix B.

Published Spending Function Families

Two single-parameter spending function families are available in East. One such family is the ρ -family (Kim and DeMets, 1987; Jennison and Turnbull, 2000) whose spending functions are given by

$$\alpha(t) = \alpha t^\rho, \quad \rho > 0. \tag{62.4}$$

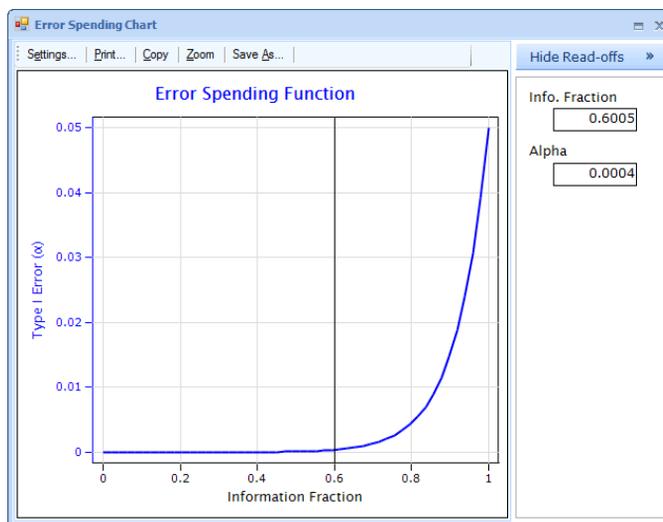
When $\rho = 1$, the corresponding stopping boundaries resemble the Pocock stopping boundaries. When $\rho = 3$, the boundaries resemble the O'Brien-Fleming boundaries. Larger values of ρ yield increasingly conservative boundaries.

Even greater flexibility is available through γ -family of spending functions (Hwang, Shih and DeCani, 1990) whose spending functions are given by

$$\alpha(t) = \begin{cases} \alpha \frac{(1-e^{-\gamma t})}{(1-e^{-\gamma})}, & \text{if } \gamma \neq 0 \\ \alpha t & \text{if } \gamma = 0. \end{cases} \tag{62.5}$$

Here negative values of γ yield convex spending functions that increase in conservatism as γ decreases, while positive values of γ yield concave spending

functions that increase in aggressiveness as γ increases. The choice $\gamma = 0$ spends the type-1 error linearly. The choice $\gamma = -4$ produces stopping boundaries that resemble the O'Brien-Fleming boundaries. The choice $\gamma = 1$ produces stopping boundaries that resemble the Pocock boundaries. The spending function below was produced with $\gamma = -12$.

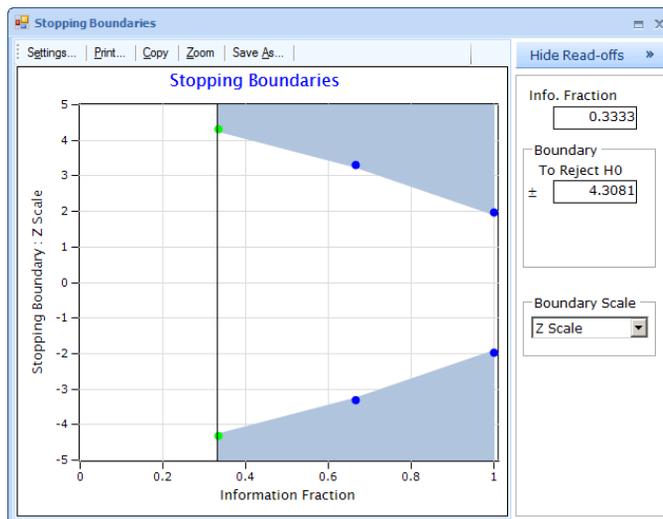


Notice that hardly any error is spent until the study has progressed 80% of the way through.

Below we display the 3-look stopping boundary on the standardized Z -statistic scale for a 2-sided design. Go to **Test Parameters** tab and change the **Test Type** to s-Sided, **Alpha** to 0.05. Also change the Spending Function to Gamma (-12) on **Boundary** tab. Notice that the test statistic must equal ± 4.3 standard deviations to stop at the first look and ± 3.32 standard deviations at the second look. This might be an appropriate stopping boundary for situations in which it is desirable to take interim looks primarily

62 Flexible Stopping Boundaries in East

for safety, but it is not desirable to stop the trial early for efficacy.



We stated that large values of γ result in spending functions that spend the error very aggressively. For example if we were to select $\gamma = 4$, we would obtain a spending function that is even more aggressive at the first look than the **LD (PK)** function

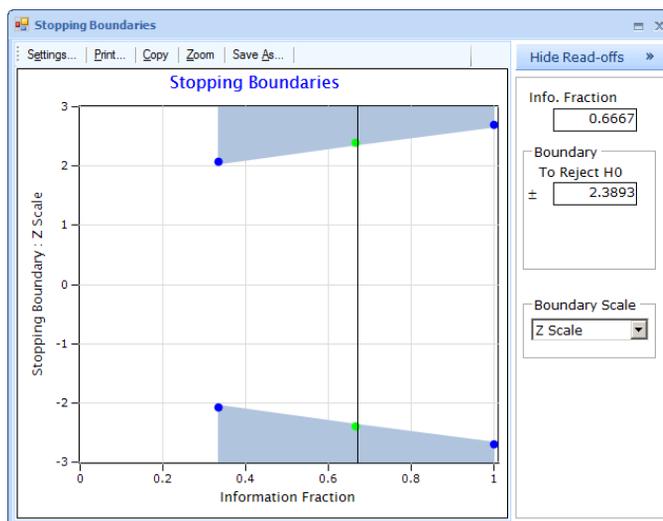
proposed by Lan and DeMets (1983).



The stopping boundaries generated by this spending function are displayed below. These boundaries actually widen over succeeding looks, unlike the Pocock boundaries that stay constant, or the O'Brien-Fleming boundaries that decrease. These might be

62 Flexible Stopping Boundaries in East

appropriate boundaries for stopping early for serious adverse events.



Interpolated Spending Functions

East permits users to specify arbitrary spending functions of their own choosing by defining the amount of α to be spent at various time points and interpolating linearly in between the time points. Interpolated spending functions can be used when it is of interest to use a published spending function and modify it. For instance, some trials use a truncated Lan and DeMets O'Brien-Fleming alpha spending function where the early boundary values are more aggressive than that generated by a regular Lan and DeMets (O'Brien-Fleming) alpha spending function. Suppose we want to take 4 equally spaced looks at the data and use a truncated Lan and DeMets O'Brien-Fleming boundary, which sets the first 2 boundary points close to each other.

Go back to **Test Parameters** tab. Change the **Number of Looks** to 4. In the **Boundary** tab, select **Spending Functions for Boundary Family**, **Lan-DeMets** for **Spending Function** and **OF** for **Parameter** in **Efficacy** box.

Choose **Spacing of Looks** as **Equal**.

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|---------|
| | | | Upper | Lower |
| 1 | 0.25 | 0 | 4.3326 | -4.3326 |
| 2 | 0.5 | 0.0031 | 2.9631 | -2.9631 |
| 3 | 0.75 | 0.0193 | 2.3590 | -2.3590 |
| 4 | 1 | 0.05 | 2.0141 | -2.0141 |

The cumulative α spent in the second look is 0.0031. As we want to spend equal amount of α in the first two looks, the α to be spent in the first look is $0.0031/2 = 0.00165$. That is, we are looking for a interpolated spending function with 4 equally spaced looks like below:

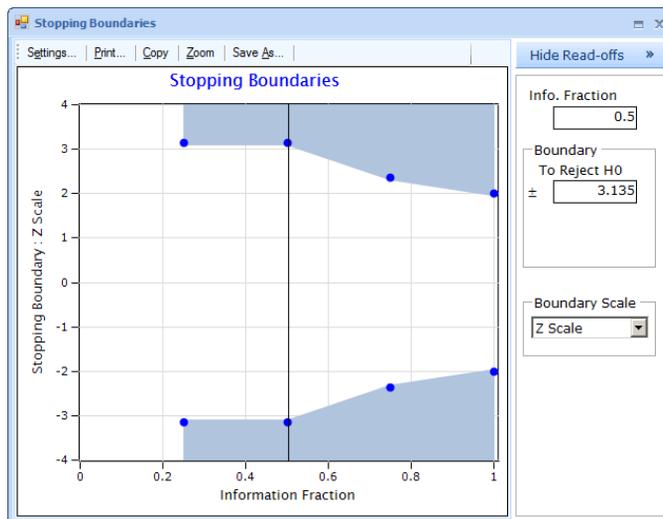
| t | $\alpha(t)$ |
|------|-------------|
| 0.25 | 0.00165 |
| 0.50 | 0.0031 |
| 0.75 | 0.0193 |
| 1.0 | 0.05 |

Change the **Spending Functions** to **Interpolated** and enter the values 0.00165, 0.0031 and 0.0193 in the first 3 cells of **Cum. α Spent**. Click **Recalc**.

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|---------|
| | | | Upper | Lower |
| 1 | 0.25 | 0.00165 | 3.1469 | -3.1469 |
| 2 | 0.5 | 0.0031 | 3.1350 | -3.1350 |
| 3 | 0.75 | 0.0193 | 2.3698 | -2.3698 |
| 4 | 1 | 0.05 | 2.0169 | -2.0169 |

62 Flexible Stopping Boundaries in East

To see the stopping boundaries for this modified α -spending function, click .



These boundaries as you can observe are more aggressive at the first look than a regular Lan and DeMets O'Brien-Fleming boundary.

Spending the α Error Asymmetrically

It is sometimes desirable to spend the total type-1 error asymmetrically. Thus, suppose that we wish to split the total type-1 error, α , of a two-sided test into two components α_l and α_u , with $\alpha_l + \alpha_u = \alpha$ in such a way that the probability, under H_0 , of crossing the upper boundary is α_u and the probability, under H_0 , of crossing the lower boundary is α_l . The algorithm for constructing these asymmetric boundaries is given in Section B.2.4 of Appendix B.

We will now illustrate the use of these asymmetric two-sided α -spending function boundaries through an example. The CRASH trial (Lancet, 2004) was a very large multicenter clinical trial to determine the efficacy and safety of administering intravenous corticosteroids to subjects with significant head injury. Subjects with a Glasgow Coma Score of 14 or less were randomized to placebo or corticosteroids. The primary endpoint was death within 14 days. The public health implications of the conclusions from this study were expected to be significant. On the one hand, there was evidence from previous randomized studies that the use of corticosteroids is beneficial. On the other hand, evidence from meta-analysis suggested the possibility of

harm. The CRASH trial was intended to settle this issue. A large sample size was needed because any benefit was likely to be small. The risk of death in patients allocated to placebo was expected to be around 15%. Because even a 2% survival difference would be clinically important, the trial had to be large enough to detect a difference of this size. Accordingly, the trial planned to enroll a maximum of 20,000 patients. A sample size this large would be able to detect a 2% benefit with over 90% power while limiting the (two-sided) type-1 error to 0.01. A five-look group sequential design with a Lan-DeMets (O'Brien-Fleming) spending function was adopted since it would be desirable to terminate the trial early if a statistically significant result emerged.

First, click **Discrete: Two Samples** on the **Design** tab, and then click **Parallel Design: Difference of Proportions**. Change the **Number of Looks** to 5. In the **Design Parameters** tab, select **Superiority** as **Design Type** and **2-Sided (Asymmetric)** for **Test Type**. East will ask you to specify the upper and lower α . This is where we can specify that we wish to spend the total type-1 error asymmetrically. Suppose that we split the 0.01 type-1 error into two components each equal to 0.005. This implies that we are equally interested in detecting harm or detecting benefit. Therefore, enter 0.005 for both upper and lower α . Select the radio-button corresponding to **Power (1- β)** and enter 20000 for **Sample Size (n)**. Specify **Prop. under Control (π_c)** as 0.15 and **Prop. under Treatment (π_t)** as 0.13. The **Design Parameters** tab should appear as below:

Design Type: Superiority Number of Looks: 5

Test Parameters Boundary

Test Type: 2-Sided (Asymmetric)

Type I Error (α): Upper 0.005 Lower 0.005

Power: 0.9

Sample Size (n): Computed

Allocation Ratio: 1 (n_1/n_2)

Specify Proportion Response

Prop. under Control (π_c): 0.15

Specify Alternative Hypothesis

Prop. under Treatment (π_t): 0.13

Diff. in Prop. ($\delta_1 = \pi_t - \pi_c$): -0.02

Specify Variance

Pooled Estimate

Unpooled Estimate

Use Casagrande-Pike-Smith Correction (Ignored if alloc. ratio is not 1)

Click the **Boundary Info** tab. It is reasonable to suppose that if the corticosteroids are harmful, one would wish to detect this fact early in the trial, and terminate it before half of the 20,000 subjects are randomized to a harmful product. Therefore, one might prefer to spend the available type-1 error aggressively, using, say a Pocock type spending function, for the upper stopping boundary. On the other hand, if the corticosteroids are beneficial, it might be desirable to apply the more conservative O'Brien-Fleming type spending function for the lower stopping boundary so that

62 Flexible Stopping Boundaries in East

stronger evidence of benefit is obtained before the trial is terminated. Select **Spending Functions** for **Efficacy Boundary Family**. Choose **Lan-DeMets** as **Spending Function** in both **Upper Efficacy Boundary** and **Lower Efficacy Boundary** boxes. For **Parameter**, select **PK** and **OF** in **Upper Efficacy Boundary** and **Lower Efficacy Boundary** boxes, respectively.

Design Parameters
Boundary Info

Efficacy Boundary Family: Spending Functions
Total Type I Error (α): 0.010

Upper Efficacy Boundary

Spending Function: Lan-DeMets

Parameter: PK

Upper α : 0.005

Lower Efficacy Boundary

Spending Function: Lan-DeMets

Parameter: OF

Lower α : 0.005

Spacing of Looks: Equal Unequal
Boundary Scale: Z Scale

| Look # | Info. Fraction | Upper α | Lower α | Efficacy Boundary | |
|--------|----------------|----------------|----------------|-------------------|---------|
| | | | | Upper | Lower |
| 1 | 0.200 | 0.0015 | 0.0000 | 2.9725 | -6.1680 |
| 2 | 0.400 | 0.0026 | 0.0000 | 2.9899 | -4.2867 |
| 3 | 0.600 | 0.0035 | 0.0003 | 2.9919 | -3.4436 |
| 4 | 0.800 | 0.0043 | 0.0017 | 2.9916 | -2.9470 |
| 5 | 1.000 | 0.0050 | 0.0050 | 2.9911 | -2.6153 |

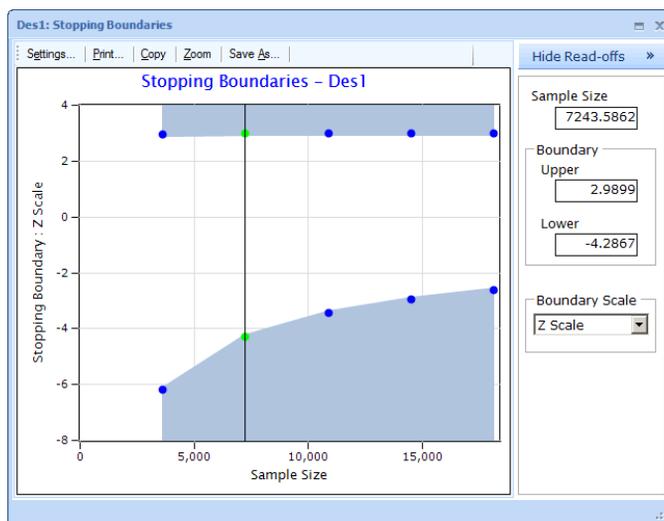
Click **Compute**. Select Des1 in the **Output Preview** and click  .

| | Des 1 |
|--|----------------------|
| Mnemonic | PN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 5 |
| Test Type | 2-Sided (Asymmetric) |
| Specified α | 0.005,0.005 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_C) | 0.15 |
| Proportion under Treatment (π_T) | 0.13 |
| Diff. in Prop. ($\pi_T - \pi_C$) | -0.02 |
| Variance | Unpooled Estimate |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (PK), LD (OF) |
| Sample Size | |
| Maximum | 18109 |
| Expected Under H0 | 18058.448 |
| Expected Under H1 | 14245.509 |

Although the design requires an up-front commitment of 18,109 patients, if in fact the corticosteroids do reduce the mortality rate by 2%, then the trial is likely to terminate early with an expected sample size of 14246. To see the stopping boundaries, select this Design, click  in the **Output Summary** toolbar and then select **Stopping**

62 Flexible Stopping Boundaries in East

Boundaries.



The asymmetry in the lower and upper stopping boundaries ensures that if corticosteroids are harmful, this fact will be detected more quickly than would be the case with symmetric two sided boundaries.

62.3.2 The Beta Spending Function

Suppose we wish to design a group sequential trial with α as the type-1 error and β as the type-2 error; i.e., with $1 - \beta$ as the power. Just as we can use an α -spending function to generate efficacy boundaries, we can use a β -spending function to generate futility boundaries, or boundaries for early stopping in favor of the null hypothesis. The idea of designing trials with futility boundaries was developed by Pampallona and Tsiatis (1994). The further idea of using β -spending functions to create such boundaries both at the design and interim monitoring stages was developed by Pampallona, Tsiatis and Kim (2001). These boundaries are crossed with probability β under the alternative hypothesis. Moreover, the probability of crossing these boundaries increases as the treatment effect decreases towards the null hypothesis until, at the null hypothesis itself, the probability of crossing is $1 - \alpha$. Futility boundaries may be used either by themselves or in conjunction with efficacy boundaries. When an efficacy boundary and a futility boundary are both present in the same study, they are forced to meet at the last look, so that either H_0 is rejected or H_1 is rejected by the end of the study. Refer to Appendix B, Section B.2.4 for the technical details concerning the use of β -spending functions and the construction of futility boundaries.

Trials with Early Stopping for Efficacy or Futility

Consider a hypothetical two-arm hypertension clinical trial in which $X_{ic} \sim N(\mu_c, 1)$ is the blood pressure reduction for the i th subject in the control group, $X_{it} \sim N(\mu_t, 1)$ is the blood pressure reduction for the i th subject in the treatment group, and $\delta = \mu_t - \mu_c$. The trial should have 90% power to detect $\delta_1 = 0.3$ using a maximum of $K=5$ equally spaced looks, and we will assume that all measurements are made on a standardized scale so that $\sigma^2 = 1$. We wish to construct a one-sided level-0.025 test with both an efficacy and a futility boundary. These boundaries should be such that if H_1 is true ($\delta = \delta_1 = 0.3$) the upper efficacy boundary will be crossed with probability 0.9, whereas if H_0 is true ($\delta = 0$), the lower futility boundary will be crossed with probability $1 - 0.025 = 0.975$. The efficacy boundary is generated by specifying an α -spending function. The futility boundary is generated by specifying a β -spending function.

Start East afresh. First, click **Continuous: Two Samples** on the **Design** tab, and then click **Parallel Design: Difference of Means**. Change the **Number of Looks** to 5. In the **Design Parameters** tab, select **Superiority** as **Design Type** and **1-Sided** as **Test Type**. Select **Difference of Means** for **Input Method** and specify **Difference in Means ($\mu_t - \mu_c$)** as 0.3. Enter 1 for **Std. Deviation (σ)**. Enter values for **Type I Error (α)** and **Power (1- β)** as 0.025 and 0.9, respectively. The **Design Parameters** tab should appear as below:

Design Type: Superiority Number of Looks: 5

Test Parameters Boundary

Test Type: 1-Sided Input Method: Difference of Means Test Statistic: Z

Type I Error (α): 0.025

Power: 0.9

Sample Size (n): Computed

Allocation Ratio: 1 (n_t/n_c)

Diff. in Means ($\delta = \mu_t - \mu_c$): 0.3

Std. Deviation (σ): 1

Click the **Boundary Info** tab. In this tab, we must specify both the α - and β -spending functions. Select **Spending Functions** for **Boundary Family** in both **Efficacy** and **Futility** boxes. The next field asks you choose the type of spending function. There is complete flexibility to select any member of any of the four available spending function families (**Rho Family**, **Gamma Family**, **Lan-DeMets Family**, **Power Family**) for spending α and independently for spending β . Suppose we decide that we will use the **Gm (-8)** spending function for spending α and the **Gm (-4)** spending function for spending the β . This might be a good choice, for instance, if the sponsor wants to set a very high hurdle for early stopping for efficacy,

62 Flexible Stopping Boundaries in East

but wants to have a reasonable chance of pulling out early if the trial is going nowhere. Select **Gamma Family** as a type of **Spending Function** in both **Efficacy** and **Futility** boxes. Specify **Parameter (γ)** as -8 and -4 for efficacy and futility, respectively. Notice that in the **Futility** box you are given a further choice between **Binding** and **Non Binding** radio-buttons. The default selection is **Non Binding** and implies that the futility boundary will be constructed in such a way that it can be overruled if desired without inflating the type-1 error. This flexibility is important, since the sponsor or the data monitoring committee might well prefer to keep the trial going to gather additional information, despite crossing the futility boundary. A **Binding** futility boundary is generally not recommended. It interacts with the corresponding efficacy boundary in such a way that unless it is strictly enforced (i.e., unless the trial is terminated if the futility boundary is crossed) the type-1 error might be inflated. Thus, for the present, select the default **Non Binding** radio button. We will compare the operating characteristics of binding and non binding futility boundaries at the end of the present section. A more detailed technical discussion is available in Appendix B, Section B.2.4. The **Boundary Info** tab will look as shown below:

Test Parameters
Boundary

Efficacy

Boundary Family: Spending Functions

Spending Function: Gamma Family

Parameter (γ): -8

Type I Error (α): 0.025

Futility

Boundary Family: Spending Functions

Spending Function: Gamma Family

Parameter (γ): -4

Type II Error (β): 0.1

Non-Binding
 Binding

Spacing of Looks Equal Unequal

Boundary Scale: Z Scale

| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Cum. β Spent | Futility Boundary |
|--------|----------------|-------------------------------------|-------------------------------------|---------------------|-------------------|--------------------|-------------------|
| 1 | 0.200 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.000 | 3.989 | 0.002 | -1.362 |
| 2 | 0.400 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.000 | 3.581 | 0.007 | -0.436 |
| 3 | 0.600 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.001 | 3.121 | 0.019 | 0.384 |
| 4 | 0.800 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.005 | 2.596 | 0.044 | 1.169 |
| 5 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.025 | 1.971 | 0.100 | 1.971 |

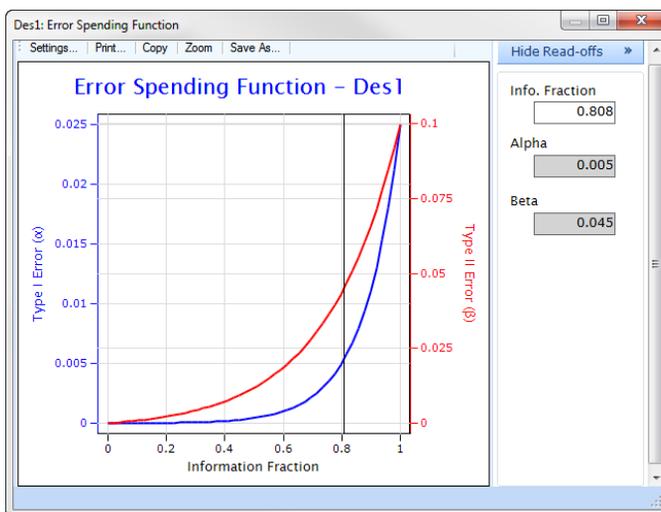
Note - In the **Spacing of Looks** table of the **Boundary Info** tab, notice that there are ticked checkboxes under the columns **Stop for Efficacy** and **Stop for Futility**. East gives you the flexibility to remove one of the stopping boundaries at certain looks, subject to the following constraints: (1) both boundaries must be included at the final two looks, (2) at least one boundary, either efficacy or futility, must be present at each look, (3) once a boundary has been selected all subsequent looks must include this boundary as well and (4) efficacy boundary for the penultimate look cannot be absent.

Click **Compute**. Select Des 1 by clicking anywhere along the row in the **Output Preview** and click the icon to save this design in the **Library**. Select Des 1 in

Output Preview or in **Library** and click the  icon. This will display the design details in the **Output Summary**.

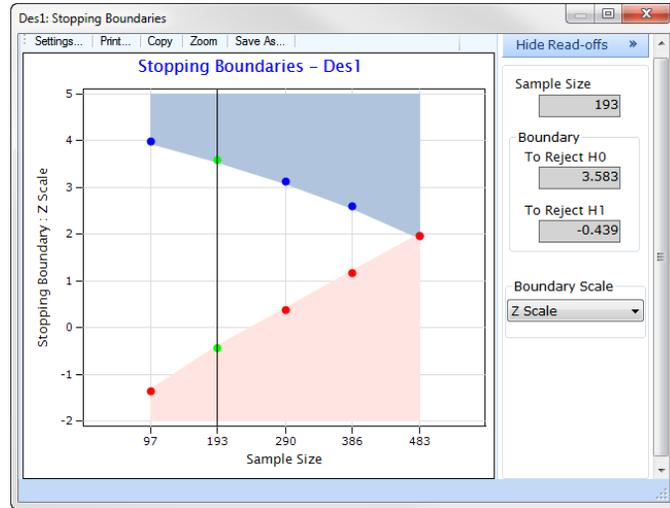
| Des 1 | |
|---|---------------------|
| Mnemonic | MN-25-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 5 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Attained α | 0.024 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Input Method | Difference of Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 0.3 |
| Std. Deviation (σ) | 1 |
| Test Statistic | Z |
| Boundary Parameters | |
| Spacing of Looks | Unequal |
| Efficacy Boundary | Gm (-8) |
| Futility Boundary | Gm (-4) (NB) |
| Sample Size | |
| Maximum | 483 |
| Expected Under H0 | 291.451 |
| Expected Under H1 | 378.875 |

To see the spending functions, click on the  icon from the **Output Summary** toolbar and then select **Error Spending**.



62 Flexible Stopping Boundaries in East

Notice how much slower the **Gm (-8)** function spends the α error than the **Gm (-4)** function spends the β error. Close the spending function chart and select **Stopping Boundaries** after clicking on the  icon.



An important feature of the stopping boundaries in Des 1 is that they meet at the final look. East forces this property on all **H0 or H1** boundaries. The computational details are given in Appendix B. By forcing the boundaries to meet, one is guaranteed to decide to either reject H_0 or reject H_1 . There is no area of indecision. This leads to a slight increase in the maximum sample size relative to a boundary corresponding to H_0 rejection. For comparison purposes, create a new design by right-clicking Des 1 in the **Library**, and clicking  icon. Go to the **Boundary Info** tab and change the **Boundary Family** to **None** in the **Futility** box. Click **Compute**. Select both Des 1

and Des 2 in the **Output Preview** and click .

| | Des 1 | Des 2 |
|---|---------------------|---------------------|
| Mnemonic | MN-2S-DI | MN-2S-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 5 | 5 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Attained α | 0.024 | |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Input Method | Difference of Means | Difference of Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 0.3 | 0.3 |
| Std. Deviation (σ) | 1 | 1 |
| Test Statistic | Z | Z |
| Boundary Parameters | | |
| Spacing of Looks | Unequal | Unequal |
| Efficacy Boundary | Gm (-8) | Gm (-8) |
| Futility Boundary | Gm (-4) (NB) | |
| Sample Size | | |
| Maximum | 483 | 469 |
| Expected Under H0 | 291.451 | 468.412 |
| Expected Under H1 | 378.875 | 377.752 |

For a very small increase in the up-front sample size commitment, Des 1 produces about the same saving in expected sample size as Des 2 if $\delta = 0.3$ and a considerably larger saving if $\delta = 0$. Moreover, as stated earlier, the futility boundary of Des 1 is non-binding; it can be overruled whenever desired without causing the type-1 error to exceed α , and without decreasing the power. Thus, all in all, Des 1 would appear to be superior to Des 2.

Futility boundaries derived from β -spending functions were introduced initially by Pampallona, Tsiatis and Kim (1995), (2001). The boundaries proposed in those papers had the serious drawback of being mandatory or binding. They interacted with the corresponding efficacy boundaries in such a way that one could not overrule them without the risk of inflating the type-1 error. For this reason, they were not very practical. Data monitoring committees (DMCs) prefer to use group sequential boundaries as guidance rather than as mandatory stopping rules. Efficacy boundaries pose no difficulty in this regard. If an efficacy boundary is crossed but the DMC votes nevertheless to keep the trial going to gain some additional information (on a secondary endpoint, say), there might be some loss of power, but there is no risk of inflating the type-1 error. Futility boundaries, as derived by Pampallona, Tsiatis and Kim (2001) are a different matter. They cannot be overruled without the risk of inflating the type-1 error. The modification to these boundaries that we have proposed

62 Flexible Stopping Boundaries in East

in Appendix B, Section B.2.4 overcomes this difficulty.

To compare non-binding and binding futility boundaries, create a new design by right-clicking Des 1 in the **Output Preview**, and clicking  icon. Go to the **Boundary Info** tab, and select the radio button corresponding to **Binding** in the **Futility** box, and click **Compute**. Select both Des 1 and Des 3 and click .

| | Des 1 | Des 3 |
|---|---------------------|---------------------|
| Mnemonic | MN-25-DI | MN-25-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 5 | 5 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Attained α | 0.024 | |
| Power | 0.9 | 0.901 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Input Method | Difference of Means | Difference of Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 0.3 | 0.3 |
| Std. Deviation (σ) | 1 | 1 |
| Test Statistic | Z | Z |
| Boundary Parameters | | |
| Spacing of Looks | Unequal | Unequal |
| Efficacy Boundary | Gm (-8) | Gm (-8) |
| Futility Boundary | Gm (-4) (NB) | Gm (-4) (B) |
| Sample Size | | |
| Maximum | 483 | 477 |
| Expected Under H0 | 291.451 | 289.528 |
| Expected Under H1 | 378.875 | 375.537 |

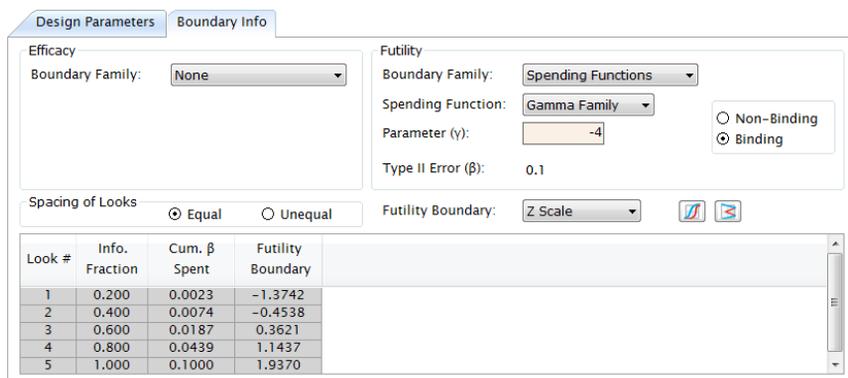
Des 3 is very similar to Des 1 in terms of maximum and expected sample sizes. The two designs differ in one important respect, however. The upper efficacy boundary of Des 3 is different from the upper efficacy boundary of Des 1, whereas the upper efficacy boundary of Des 1 is identical to the upper efficacy boundary of Des 2. Thus, the attained α for Des 1 is slightly lower than the specified α : the futility boundary will capture a small proportion of trials that would otherwise have crossed the efficacy boundary as type-1 errors.

Trials with Early Stopping for Futility Only

Let us consider, once again, the hypertension clinical trial introduced at the beginning of the ongoing subsection. Suppose the trial is designed for a test of $H_0 : \mu_t - \mu_c = 0$ at one-sided significance level $\alpha = 0.025$ and 90% power at the alternative hypothesis $H_1 : \mu_t - \mu_c = 0.3$ with an assumed variance $\sigma^2 = 1$. There will be five equally spaced looks at the data with a futility boundary for terminating the trial early with the declaration that H_0 cannot be rejected. The futility boundary is required to have the

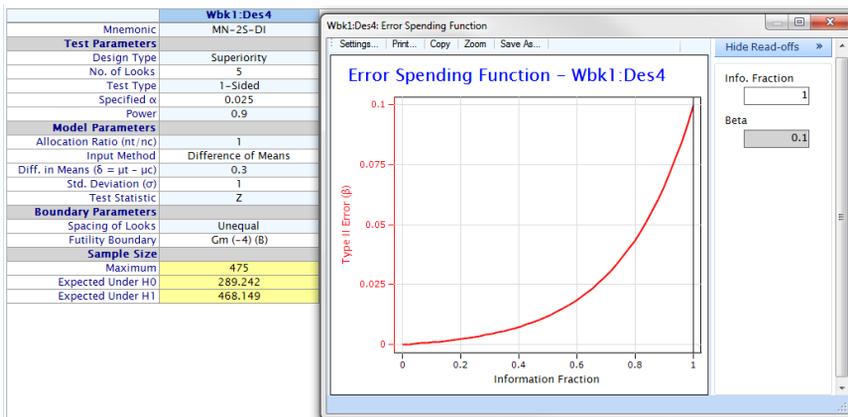
property that the overall boundary crossing probability under H_1 is 0.1. There is no intention to stop the trial early for efficacy.

Create a new design by right-clicking Des 1 in the **Library**, and clicking  icon. Go to the **Boundary Info** tab. In the **Efficacy** box, change the **Boundary Family** to **None** from the drop-down list. In the **Futility** box, set the **Boundary Family** to **Spending Function** and select **Gamma Family** in the ensuing field. Type in -4 as the value of **Parameter (γ)**. Select the radio-button corresponding to **Binding**.



| Look # | Info. Fraction | Cum. β Spent | Futility Boundary |
|--------|----------------|--------------------|-------------------|
| 1 | 0.200 | 0.0023 | -1.3742 |
| 2 | 0.400 | 0.0074 | -0.4538 |
| 3 | 0.600 | 0.0187 | 0.3621 |
| 4 | 0.800 | 0.0439 | 1.1437 |
| 5 | 1.000 | 0.1000 | 1.9370 |

Click **Compute** to obtain size for this 'Futility only' design. East will create this design with label Des 4. A summary of Des 4 and the associated β spending function are displayed below.



| Wbk1:Des4 | |
|---|---------------------|
| Mnemonic | MN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 5 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Input Method | Difference of Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 0.3 |
| Std. Deviation (σ) | 1 |
| Test Statistic | Z |
| Boundary Parameters | |
| Spacing of Looks | Unequal |
| Futility Boundary | Gm (-4) (B) |
| Sample Size | |
| Maximum | 475 |
| Expected Under H0 | 289.242 |
| Expected Under H1 | 468.149 |

62 Flexible Stopping Boundaries in East

Edit Des 4 to create a corresponding a **Single-look** study (Des 5), designed for the same effect size, type-1 error and power. In Des 5, we are forced to continue until the maximum sample size is reached, unless it is terminated due to low conditional power. We have pointed out in Chapter 61 that use of low conditional power to terminate a trial early is rather ad hoc, and gives us no assurance that the overall unconditional power of the study will be preserved.

| | Des4 | Des5 |
|---|------------------|------------------|
| Mnemonic | MN-25-DI | MN-25-DI |
| Test Parameters | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 5 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.9 | 0.9 |
| Model Parameters | | |
| Allocation Ratio (nt/nc) | 1 | 1 |
| Input Method | Individual Means | Individual Means |
| Diff. in Means ($\delta = \mu_t - \mu_c$) | 0.3 | 0.3 |
| Mean Control (μ_c) | 0 | 0 |
| Mean Treatment (μ_t) | 0.3 | 0.3 |
| Std. Deviation (σ) | 1 | 1 |
| Test Statistic | Z | Z |
| Boundary Parameters | | |
| Spacing of Looks | Equal | |
| Futility Boundary | Gm (-4) (B) | |
| Sample Size | | |
| Maximum | 475 | 467 |
| Expected Under H0 | 289.242 | |
| Expected Under H1 | 468.149 | |

Des 5 requires a commitment of 467 patients. However, there is no option under Des 5 to stop the trial early if the effect size is smaller than was anticipated at the design stage. In contrast Des 4 requires an up-front commitment of 475 patients, five more than Des 5. But this is a small price to pay for the flexibility to take interim looks and stop early if the futility boundary is crossed. The expected sample size of Des 1 is 289 patients if H_0 is true.

63 Confidence Interval Based Design

During the design of an experiment such as a clinical trial, when researchers consider a hypothesis test for a parameter of interest, say δ , either the unknown sample size for the desired power or the unknown power for a fixed sample size must be determined. A confidence interval based design calculates the sample size based on the desired width of a confidence interval for the parameter of interest rather than the power of the hypothesis test. In previous versions of East, a user could employ a confidence interval based approach only via a labor intensive process of trial and error by generating repeated confidence interval charts. East now allows the computation of such a sample size for many single look designs based on analytical methods without the need to use such charts. The result is a quick and efficient way to compute the sample size required to achieve a desired width for a confidence interval for δ , given the confidence level $1 - \alpha$.

Definitions

- $1 - \alpha$ denotes the confidence level
- ω is the measure of precision for δ (width of confidence interval)
- $\hat{\delta}$ is the empirical estimate of δ

The estimated sample size n must satisfy the following:

For a two-sided confidence interval

$$P(\hat{\delta} - \omega \leq \delta \leq \hat{\delta} + \omega) = 1 - \alpha$$

For a one-sided confidence interval

$$P(\delta \geq \hat{\delta} - \omega) = 1 - \alpha$$

or

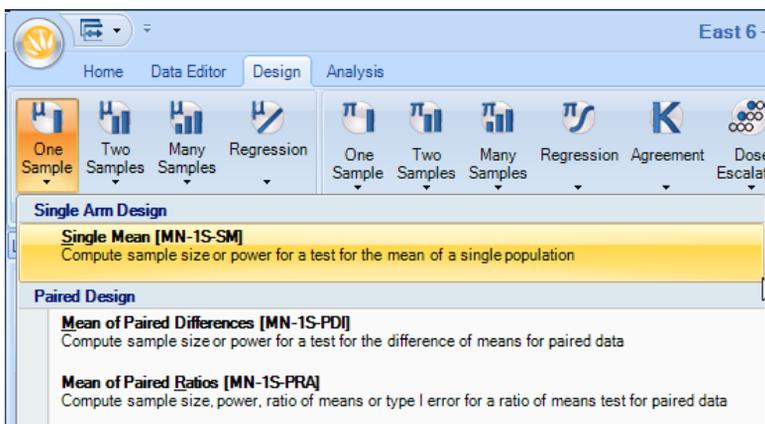
$$P(\delta \leq \hat{\delta} + \omega) = 1 - \alpha$$

63.1 One Sample Test for a Single Mean for Continuous Data

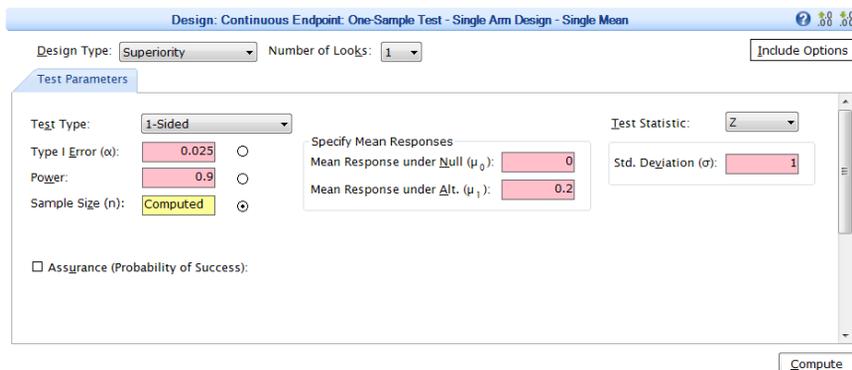
Consider the problem of comparing the mean of the distribution of observations from a single random sample of continuous data to a specified constant. Suppose it is required to estimate the sample size for obtaining a 95% two-sided confidence interval for the population mean with a precision of 5 units, when the population standard deviation is known to be 20 units.

63 Confidence Interval Based Design

To illustrate this example, in East under the **Design** ribbon for **Continuous** data, click **One Sample** and then click **Single Arm Design: Single Mean** as shown:

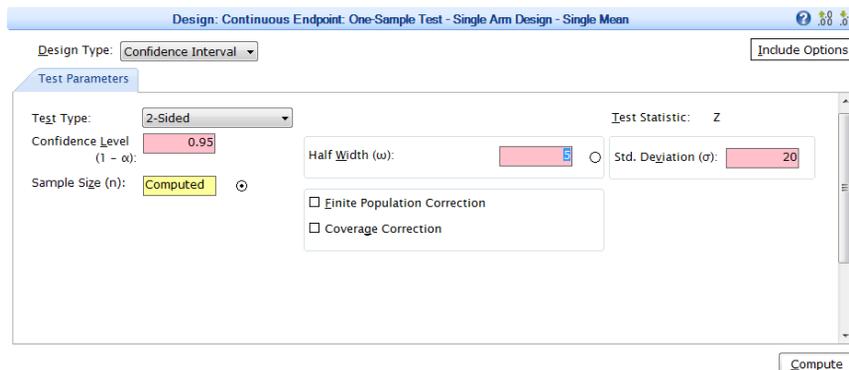


This will launch the following input window:



Choose **Confidence Interval** in the **Design Type** dropdown box and enter the following design parameters:

- Test Type: 2 sided
- Confidence Level ($1 - \alpha$): 0.95
- Sample Size (n): Computed (select radio button)
- Half Width (ω): 5.0
- Standard Deviation (σ): 20



The Confidence Interval based design for this particular test also allows the user to specify whether or not a **Finite Population Correction** for a fixed **Population Size** is used. In addition, the user can also determine if a **Coverage Correction** is to be used for a given **Coverage Probability**. This coverage correction may become necessary when the population standard deviation is unknown and is to be estimated from the sample. For now leave these boxes unchecked and click **Compute**. The sample size for this design is calculated and the output is shown as a row in the **Output Preview** window:

| ID | Design Type | Test Type | Sample Size | ω | $(1 - \alpha)$ | σ | Test Statistic |
|------|---------------------|-----------|-------------|----------|----------------|----------|----------------|
| Des1 | Confidence Interval | 2-Sided | 62 | 5 | 0.95 | 20 | Z |

As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview toolbar. The design summary will be displayed labeled **Output**

63 Confidence Interval Based Design

Summary.

| Des 1 | |
|-----------------------------|---------------------|
| Mnemonic | MN-1S-SM |
| Test Parameters | |
| Design Type | Confidence Interval |
| Test Type | 2-Sided |
| Confidence Level | 0.95 |
| Model Parameters | |
| Width (ω) | 5 |
| Std. Deviation (σ) | 20 |
| Test Statistic | Z |
| Sample Size | |
| Maximum | 62 |

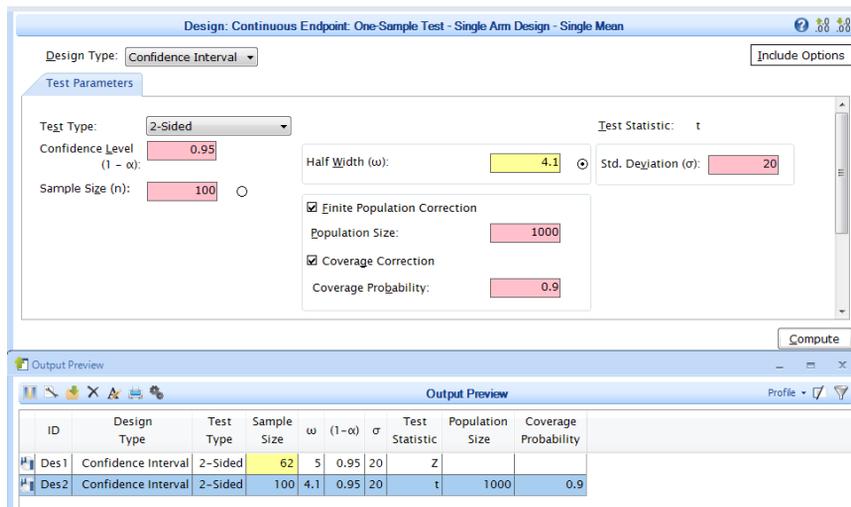
This design can be saved to the **Library** by selecting the **Des 1** in the **Output Preview** window and clicking the  icon. This test can easily be repeated for a **One-sided** confidence interval and with various values for ω and σ , as well as any desired differences in **Population Size** and **Coverage Probability**.

| | |
|--|-----------------------------------|
| <input checked="" type="checkbox"/> Finite Population Correction | |
| Population Size: | <input type="text" value="1000"/> |
| <input checked="" type="checkbox"/> Coverage Correction | |
| Coverage Probability: | <input type="text" value="0.9"/> |

Alternatively East can compute the precision level ω given a fixed sample size using a confidence interval based design. Following the example above, to determine the precision of the estimate of population parameter where the sample size is fixed, the user has to only enter the desired value, for example $n = 100$.

Enter the following in the **Design Input** screen and click **Compute**:

Test Type: 2 sided
 Confidence Level ($1 - \alpha$): 0.95
 Sample Size (n): 100
 Half Width (ω): Computed (select radio button)
 Standard Deviation (σ): 20



The precision parameter ω is calculated to be 4.1. As the sample size is increased the resulting estimate of precision increases, which is to say the precision limit decreases, providing a tighter confidence interval for the parameter of interest. The output for all parameters is again shown as a row in the **Output Preview** window and the design can be saved to the **Library** using the standard method as with all tests in East.

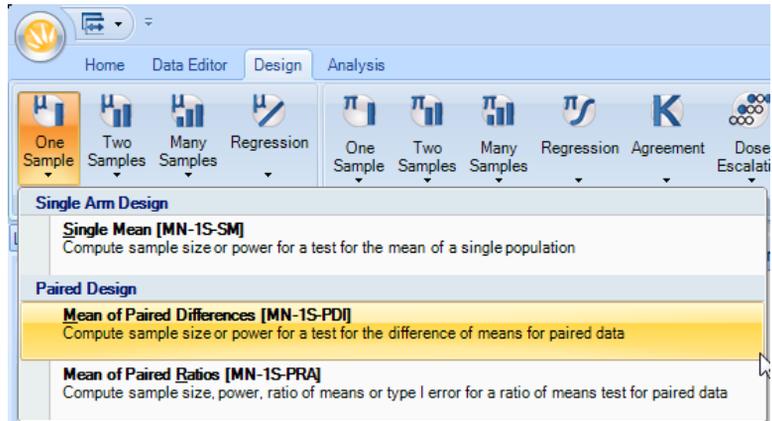
63.2 One Sample Test for the Mean of Paired Differences for Continuous Data

Consider the problem of comparing the means of two normal distributions when each observation in the random sample from one distribution is matched with a unique observation from the other distribution. Suppose it is required to estimate the sample size for obtaining a 99% two-sided confidence interval for the difference of means with a precision of 1.0 units, when the population standard deviation is known to be 3.4 units.

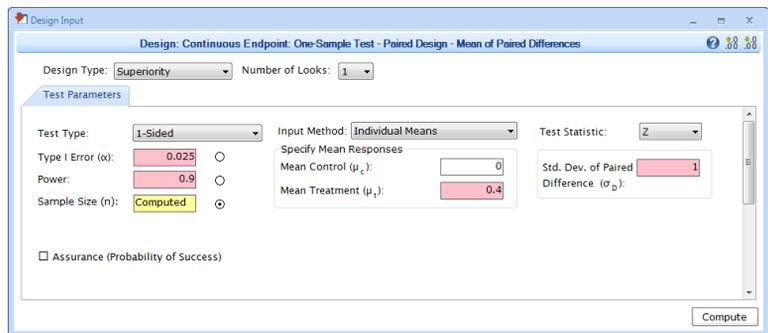
To illustrate this example, in East under the **Design** ribbon for **Continuous** data, click

63 Confidence Interval Based Design

One Sample and then click **Paired Design: Mean of Paired Differences** as shown:

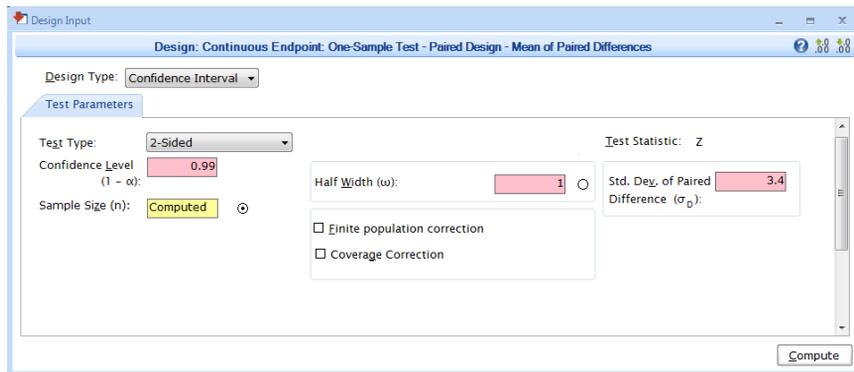


This will launch the following input window:

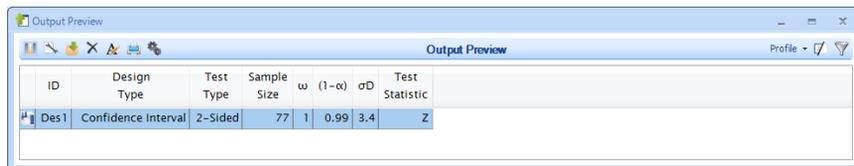


Choose **Confidence Interval** in the **Design Type** dropdown box and enter the following design parameters:

- Test Type: 2 sided
- Confidence Level ($1 - \alpha$): 0.99
- Sample Size (n): Computed (select radio button)
- Half Width (ω): 1.0
- Standard deviation of Paired Difference (σ_D): 3.4



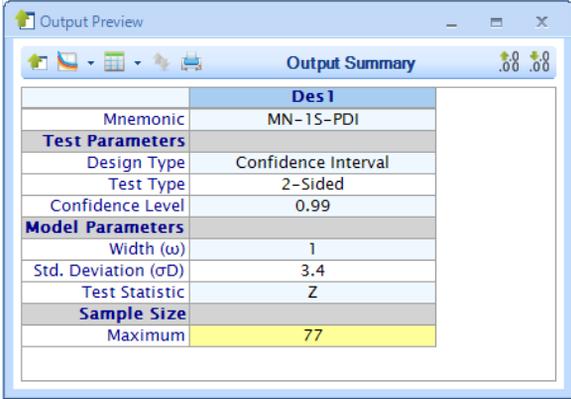
The Confidence Interval based design for this particular test also allows the user to specify whether or not a **Finite Population Correction** for a fixed **Population Size** is used. In addition, the user can also determine if a **Coverage Correction** is to be used for a given **Coverage Probability**. This coverage correction may become necessary when the population standard deviation is unknown and is to be estimated from the sample. For now leave these boxes unchecked and click **Compute**. The sample size for this design is calculated and the output is shown as a row in the **Output Preview** window:



As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview toolbar. The design details will be displayed, labeled **Output**

63 Confidence Interval Based Design

Summary.



| Des 1 | |
|-------------------------------|---------------------|
| Mnemonic | MN-1S-PDI |
| Test Parameters | |
| Design Type | Confidence Interval |
| Test Type | 2-Sided |
| Confidence Level | 0.99 |
| Model Parameters | |
| Width (ω) | 1 |
| Std. Deviation (σ_D) | 3.4 |
| Test Statistic | Z |
| Sample Size | |
| Maximum | 77 |

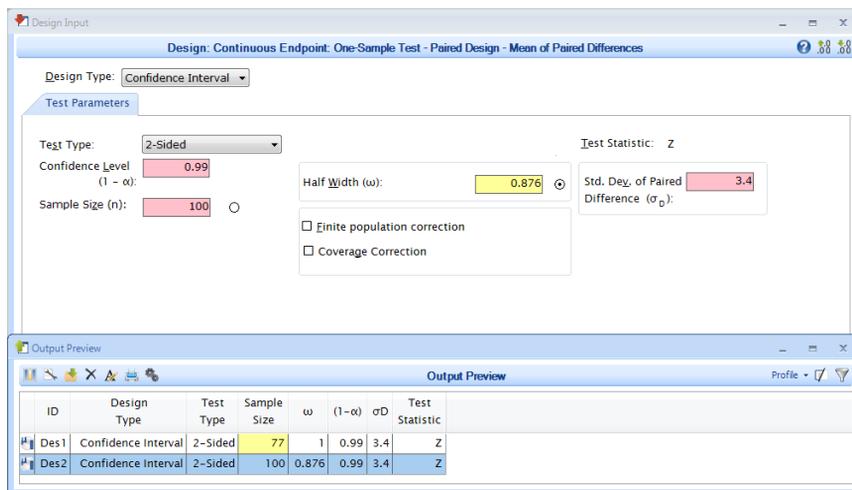
This design can be saved to the **Library** by selecting the **Des 1** in the **Output Preview** window and clicking the  icon. This test can easily be repeated for a **One-sided** confidence interval and with various values for ω and σ_D , as well as any desired differences in **Population Size** and **Coverage Probability**.

| | |
|--|-----------------------------------|
| <input checked="" type="checkbox"/> Finite Population Correction | |
| Population Size: | <input type="text" value="1000"/> |
| <input checked="" type="checkbox"/> Coverage Correction | |
| Coverage Probability: | <input type="text" value="0.9"/> |

Alternatively East can compute the precision level ω given a fixed sample size using a confidence interval based design. Following the example above, to determine the precision of the estimate of population parameter where the sample size is fixed, the user has to only enter the desired value, for example $n = 100$.

Enter the following in the **Design Input** screen and click **Compute**:

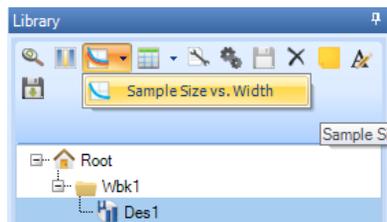
Test Type: 2 sided
 Confidence Level ($1 - \alpha$): 0.99
 Sample Size (n): 100
 Half Width (ω): Computed (select radio button)
 Standard deviation of Paired Difference (σ_D): 3.4



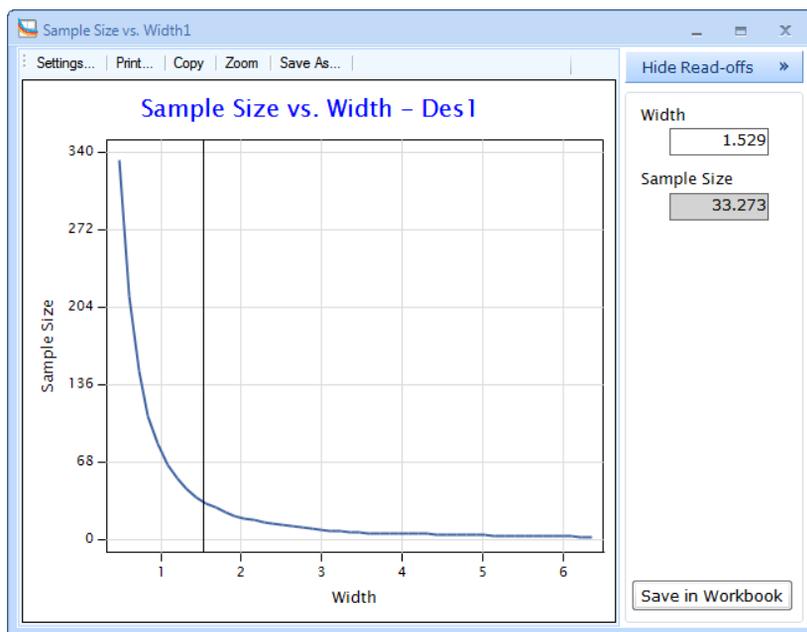
The precision parameter ω is calculated to be 0.876. As the sample size is increased the resulting estimate of precision increases, which is to say the precision limit decreases, providing a tighter confidence interval for the parameter of interest. The output for all parameters is again shown as a row in the **Output Preview** window and the design can be saved to the **Library** using the standard method as with all tests in East. From there, a summary of the design can be generated using the details  icon. East also provides a very useful **Sample Size vs. Width** plot. This dynamic visual can immediately assess how changing the sample size effects the resulting width of the confidence interval. From the **Library** choose **Sample Size vs. Width** from the

63 Confidence Interval Based Design

Plots menu.



Here, the user can move the cursor horizontally back and forth to change the interval width and immediately view the resulting sample size.



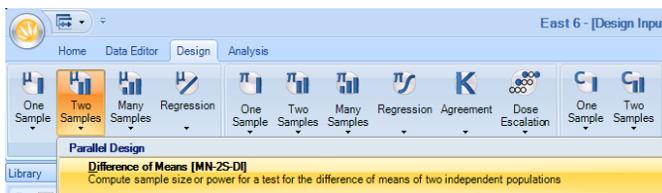
A table of **Sample Size vs. Width** values can be generated using the **Tables**  menu, also found in the **Library**. This feature allows the user to input a range of values to generate multiple confidence intervals and the corresponding sample sizes.

63.3 Two Sample Test for the Difference of Means for Continuous Data

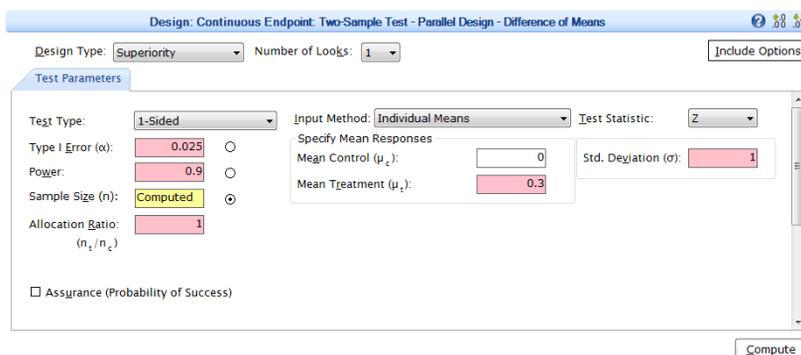
Consider the problem of comparing a new treatment to a standard protocol. It is often necessary to randomize subjects to the control and treatment arms, and then determine if the group-dependent means of the outcome variables are significantly different. The following example illustrates a confidence interval based design for such a trial when the outcomes from both groups follow a normal distribution.

Suppose it is required to estimate the sample size for obtaining a 95% two-sided confidence interval for the difference of two means with a precision of 3.0 units. Assume that the common standard deviation of the observations is 8.

In East under the **Design** ribbon for **Continuous** data, click **Two Sample** and then click **Parallel Design: Difference of Means** as shown:



This will launch the following input window:

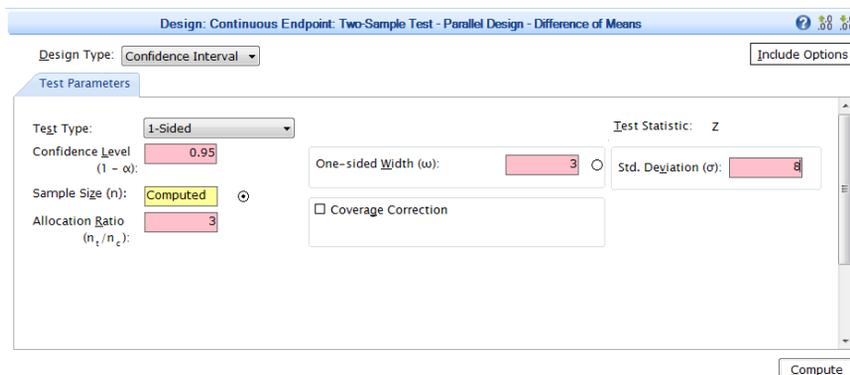


Choose **Confidence Interval** in the **Design Type** dropdown box. Consider a one sided test with 5% significance level, and an Allocation Ratio ($n_t : n_c$) of 3:1, that is, 75% of the patients are randomized to the treatment arm. Enter the following design

63 Confidence Interval Based Design

parameters:

Test Type: 1 sided
 Confidence Level ($1 - \alpha$): 0.95
 Sample Size (n): Computed (select radio button)
 Allocation Ratio: 3
 One-sided Width (ω): 3.0
 Standard Deviation (σ): 8



The Confidence Interval based design for this particular test also allows the user to specify whether or not a **Coverage Correction** is to be used for a given **Coverage Probability**. This coverage correction may become necessary when the population standard deviation is unknown and is to be estimated from the sample. For now leave this box unchecked and click **Compute**. The sample size for this design is calculated and the output is shown as a row in the **Output Preview** window:

| ID | Design Type | Test Type | Sample Size | ω | $(1-\alpha)$ | σ | nt/nc | Test Statistic |
|------|---------------------|-----------|-------------|----------|--------------|----------|-------|----------------|
| Des1 | Confidence Interval | 1-Sided | 103 | 3 | 0.95 | 8 | 3 | Z |

As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview toolbar. The design details will be displayed, labeled **Output**

Summary.

| Des 1 | |
|-----------------------------|---------------------|
| Mnemonic | MN-2S-DI |
| Test Parameters | |
| Design Type | Confidence Interval |
| Test Type | 1-Sided |
| Confidence Level | 0.95 |
| Model Parameters | |
| Width (ω) | 3 |
| Std. Deviation (σ) | 8 |
| Allocation Ratio (nt/nc) | 3 |
| Test Statistic | Z |
| Sample Size | |
| Maximum | 103 |

This design can be saved to the **Library** by selecting the **Des 1** in the **Output Preview** window and clicking the  icon. This test can easily be repeated for a **Two-sided** confidence interval and with various values for **Allocation Ratio**, ω and σ , as well as any desired differences in **Coverage Probability**.

Coverage Correction
 Coverage Probability:

Alternatively East can compute the precision level ω given a fixed sample size using a confidence interval based design. Following the example above, to determine the precision of the estimate of population parameter where the sample size is fixed, the user has to only enter the desired value, for example $n = 80$.

Enter the following in the **Design Input** screen and click **Compute**:

- Test Type: 1 sided
- Confidence Level ($1 - \alpha$): 0.95
- Sample Size (n): 80
- Allocation Ratio: 3
- One-sided Width (ω): Computed (select radio button)
- Standard Deviation (σ): 8

63 Confidence Interval Based Design

Design Input

Design: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means

Design Type: Confidence Interval

Test Parameters

Test Type: 1-Sided

Confidence Level (1 - α): 0.95

Sample Size (n): 80

Allocation Ratio (n_1/n_2): 3

One-sided Width (ω): 3.398

Std. Deviation (σ): 8

Test Statistic: Z

Coverage Correction

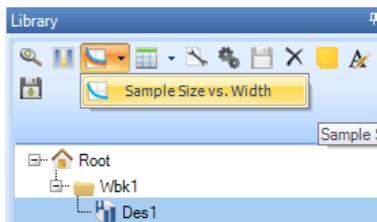
Output Preview

| ID | Design Type | Test Type | Sample Size | ω | (1- α) | σ | nt/nc | Test Statistic |
|------|---------------------|-----------|-------------|----------|----------------|----------|-------|----------------|
| Des1 | Confidence Interval | 1-Sided | 103 | 3 | 0.95 | 8 | 3 | Z |
| Des2 | Confidence Interval | 1-Sided | 80 | 3.398 | 0.95 | 8 | 3 | Z |

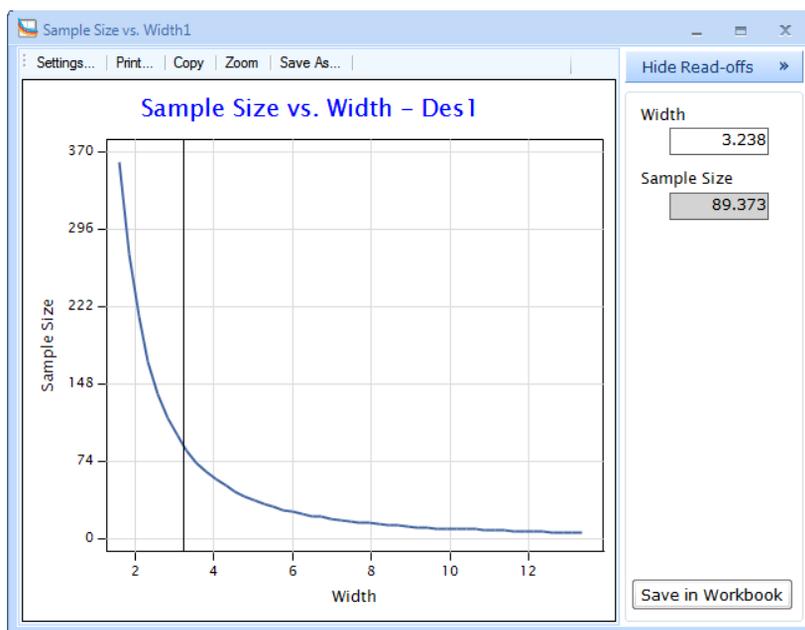
The precision parameter ω is calculated to be 3.398. As the sample size is decreased, the resulting value of ω increases. In other words, the precision limit increases, resulting in a wider confidence interval for the parameter of interest. The output for all parameters is again shown as a row in the **Output Preview** window and the design can be saved to the **Library** using the standard method as with all tests in East. From

there, a summary of the design can be generated using the details  icon. East also provides a very useful **Sample Size vs. Width** plot. This dynamic visual can immediately assess how changing the sample size effects the resulting width of the confidence interval. From the **Library** choose **Sample Size vs. Width** from the **Plots**

menu.



Here, the user can move the cursor horizontally back and forth to change the interval width and immediately view the resulting sample size.



A table of **Sample Size vs. Width** values can be generated using the **Tables**  menu, also found in the **Library**. This feature allows the user to input a range of values to generate multiple confidence intervals and the corresponding sample sizes.

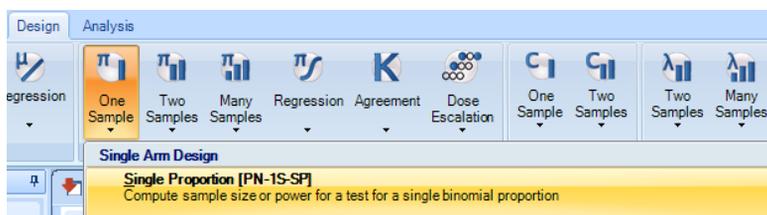
63 Confidence Interval Based Design

63.4 One Sample Test for a Single Binomial Proportion

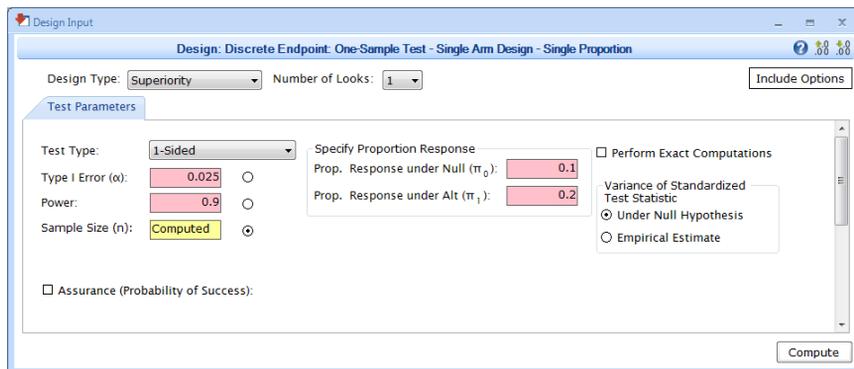
Consider the experimental situation in which an observed treatment response rate is compared to a fixed response rate derived from historical data, where the variable of interest has a binomial distribution. It is therefore of interest to determine whether the response rate π differs from a fixed value π_0 . The following example illustrates a confidence interval based design for a one arm trial having a binomial response rate, where a single binomial proportion is tested against a fixed value.

Suppose it is required to estimate the sample size to obtain a 95% two-sided confidence interval for π with a precision of 0.01 units. The sample size is determined for a specified value of π which is consistent with the alternative hypothesis, denoted π_1 . The design is a single-arm trial in which we wish to determine if the response rate of a new therapy is at least 15%. Thus, it is desired to test the null hypothesis $H_0: \pi = 0.15$ against the one-sided alternative hypothesis $H_1: \pi > 0.15$. Assume $\pi = \pi_1 = 0.25$ and a type one error rate of 0.05.

In East under the **Design** ribbon for **Discrete** data, click **One Sample** and then click **Single Arm Design: Single Proportion** as shown:

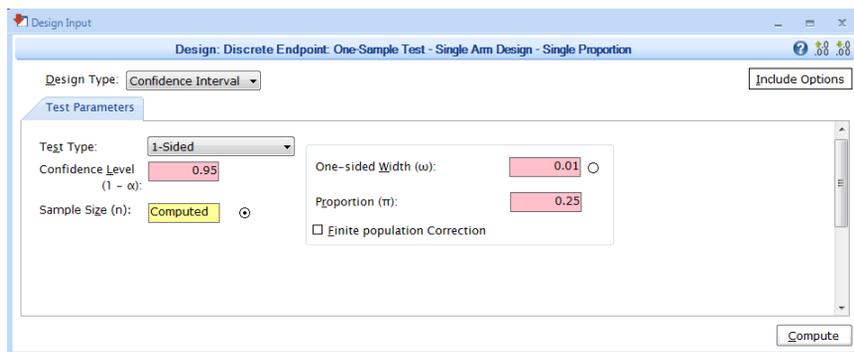


This will launch the following Design Input window:



Choose **Confidence Interval** in the **Design Type** dropdown box. Consider a one sided test with 5% significance level and fixed value of $\pi = 0.25$. Enter the following design parameters:

- Test Type: 1 sided
- Confidence Level $(1 - \alpha)$: 0.95
- Sample Size (n): Computed (select radio button)
- One-sided Width (ω): 0.01
- Proportion (π): 0.25



The Confidence Interval based design for this particular test also allows the user to specify whether or not a **Finite population Correction** is to be used for a given **Population Size**. For now leave this box unchecked and click **Compute**. The sample size for this design is calculated and the output is shown as a row in the **Output**

63 Confidence Interval Based Design

Preview window:

| ID | Design Type | Test Type | Sample Size | ω | $(1-\alpha)$ | Proportion (π) |
|------|---------------------|-----------|-------------|----------|--------------|----------------------|
| Des1 | Confidence Interval | 1-Sided | 5073 | 0.01 | 0.95 | 0.25 |

As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the icon in the Output Preview toolbar. The design details will be displayed, labeled **Output Summary**.

| Des 1 | |
|-------------------------|---------------------|
| Mnemonic | PN-1S-SP |
| Test Parameters | |
| Design Type | Confidence Interval |
| Test Type | 1-Sided |
| Confidence Level | 0.95 |
| Model Parameters | |
| Width (ω) | 0.01 |
| Proportion (π) | 0.25 |
| Sample Size | |
| Maximum | 5073 |

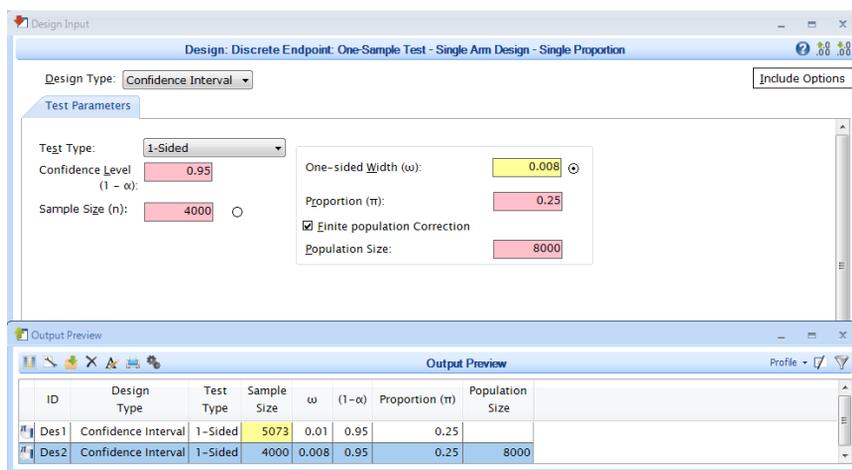
This design can be saved to the **Library** by selecting the **Des 1** in the **Output Preview** window and clicking the icon. This test can easily be repeated for a **Two-sided** confidence interval and with various values for ω and π , as well as any desired differences in **Finite Population Correction**.

| | | |
|--|-----------------------------------|-----------------------|
| Half Width (ω): | <input type="text" value="0.01"/> | <input type="radio"/> |
| Proportion (π): | <input type="text" value="0.25"/> | |
| <input checked="" type="checkbox"/> Finite population Correction | | |
| Population Size: | <input type="text" value="8000"/> | |

Alternatively East can compute the precision level ω given a fixed sample size using a confidence interval based design. Following the example above, to determine the precision of the estimate of population parameter where the sample size is fixed, the user has to only enter the desired value, for example $n = 4000$ with a finite population correction of size 8000.

Enter the following in the **Design Input** screen and click **Compute**:

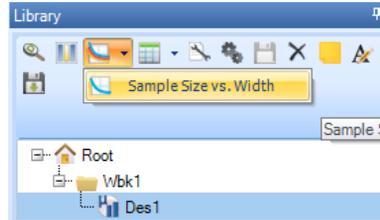
- Test Type: 1 sided
- Confidence Level ($1 - \alpha$): 0.95
- Sample Size (n): 4000
- One-sided Width (ω): Computed (select radio button)
- Proportion (π): 0.25
- Finite Population Correction: box checked
- Population Size: 8000



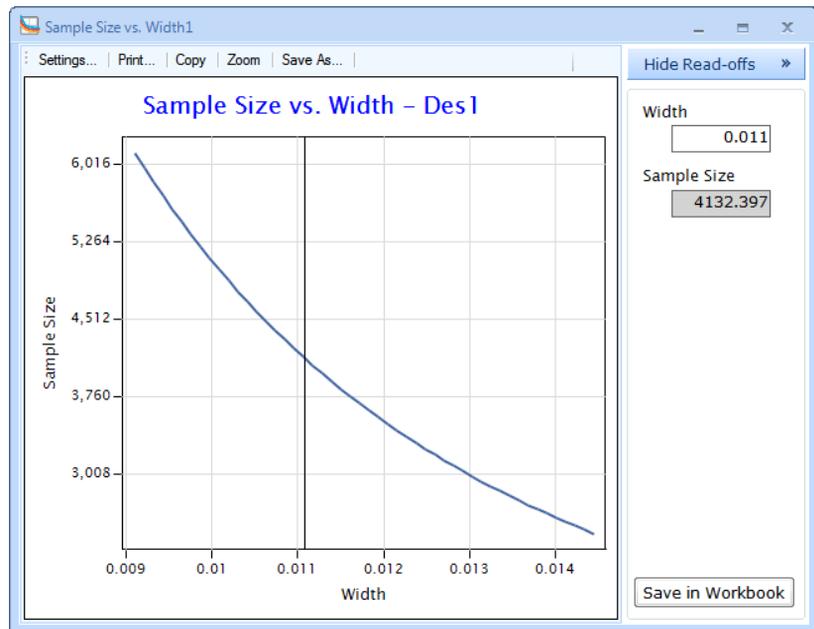
For a sample size of 4000 the precision parameter ω is calculated to be 0.008. As the sample size is decreased, the resulting value of ω decreases. For binomial data, this results in a wider confidence interval for the parameter of interest. The output for all parameters is again shown as a row in the **Output Preview** window and the design can be saved to the **Library** using the standard method as with all tests in East. From there, a summary of the design can be generated using the details  icon. East also provides a very useful **Sample Size vs. Width** plot. This dynamic visual can immediately assess how changing the sample size effects the resulting width of the

63 Confidence Interval Based Design

confidence interval. From the **Library** choose **Sample Size vs. Width** from the **Plots** menu.



Here, the user can move the cursor horizontally back and forth to change the interval width and immediately view the resulting sample size.



A table of **Sample Size vs. Width** values can be generated using the **Tables**  menu, also found in the **Library**. This feature allows the user to input a range of

values to generate multiple confidence intervals and the corresponding sample sizes.

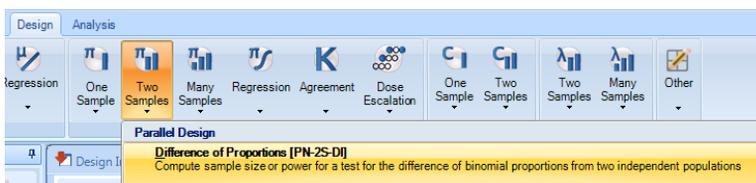
63.5 Two Sample Test for the Difference of Binomial Proportions

In medical research, outcomes dealing with the proportion of patients responding to a therapy, developing a certain side effect or requiring specialized care, are experiments based on binomial data designs. In these situations the goal is to compare independent samples from two populations in terms of the proportion of patients presenting the characteristic or outcome. East supports a Confidence Interval based approach to the design of clinical trials, independent of the power of the test, in which treatment comparison is based on the difference of such proportions.

For example, in a prospective randomized trial of placebo versus treatment for patients with a heart condition, the endpoint may be reduction in death or MI within a certain period of time after entering the study. It is of interest to detect a reduction in the event rate from 15% on the placebo arm to 10% on the treatment arm. In other words the goal is to test the null hypothesis that the treatment and placebo arms both have an event rate of 15%, versus the alternative that the treatment reduces the event rate by 5% (from 15% to 10%).

Let π_c and π_t denote the binomial probabilities for the control and treatment arms, respectively, and let $\delta = \pi_t - \pi_c$. The interest is therefore in testing the null hypothesis $H_0 : \delta = 0$, for a two-sided test with a type-1 error of 5%. Consider a confidence interval based design to estimate the sample size with a precision of $\omega = 0.05$.

In East under the **Design** ribbon for **Discrete** data, click **Two Samples** and then click **Parallel Design: Difference of Proportions** as shown:



63 Confidence Interval Based Design

This will launch the following input window:

Design Input
Design: Discrete Endpoint, Two-Sample Test - Parallel Design - Difference of Proportions

Design Type: Superiority Number of Looks: 1 [Include Options]

Test Parameters

Test Type: 1-Sided

Type I Error (α): 0.025

Power: 0.9

Sample Size (n): Computed

Allocation Ratio: 1
(n_t/n_c)

Specify Proportion Response
Prop. under Control (π_c): 0.1

Specify Alternative Hypothesis
Prop. under Treatment (π_t): 0.5
Diff. in Prop. ($\delta_1 = \pi_t - \pi_c$): 0.4

Specify Variance
 Pooled Estimate
 Unpooled Estimate

Perform Exact Computations
 Use Casagrande-Pike-Smith Correction (Ignored if alloc. ratio is not 1)

Assurance (Probability of Success):

[Compute]

Choose **Confidence Interval** in the **Design Type** dropdown box and enter the following design parameters:

Test Type: 2 sided
 Confidence Level ($1 - \alpha$): 0.95
 Sample Size (n): Computed (select radio button)
 Allocation Ratio (n_t/n_c): 1
 Prop. under Control (π_c): 0.15
 Prop. under Treatment (π_t): 0.10
 Diff. in Prop. ($\delta_1 = \pi_t - \pi_c$): -0.05 (this will be calculated)
 Half Width (ω): 0.05
 Specify Variance: Select **Unpooled Estimate** radio button

Design Input
Design: Discrete Endpoint, Two-Sample Test - Parallel Design - Difference of Proportions

Design Type: Confidence Interval [Include Options]

Test Parameters

Test Type: 2-Sided

Confidence Level ($1 - \alpha$): 0.95

Sample Size (n): Computed

Allocation Ratio: 1
(n_t/n_c)

Specify Proportion Response
Prop. under Control (π_c): 0.15

Specify Alternative Hypothesis
Prop. under Treatment (π_t): 0.1
Diff. in Prop. ($\delta_1 = \pi_t - \pi_c$): -0.05

Specify Variance
 Pooled Estimate
 Unpooled Estimate

Half Width (ω): 0.05

[Compute]

The Allocation Ratio ($n_t : n_c$) describes the ratio of patients to each arm. For example, an allocation ratio of 3:1 indicates that 75% of the patients are randomized to the treatment arm as opposed to 25% to the control.

In binomial designs, the variance of a random variable is dependent on its mean. The maximum sample size required for a study will be affected by how the differences of binomial response rates are standardized when computing the test statistic, regardless of the other design parameters. There are two options for determining how the test statistic will be standardized, using either the **Unpooled** or **Pooled** specification for variance. The difference becomes important when planning a binomial study with unbalanced randomization. In this case, both pooled and unpooled designs should be considered and the one that produces a tighter confidence interval (measure of ω) with fewer patients should be chosen. This will depend on the response rates of the control and treatment arms as well as the value of the fraction assigned to the treatment arm. More information on this can be found in Section 23.1.

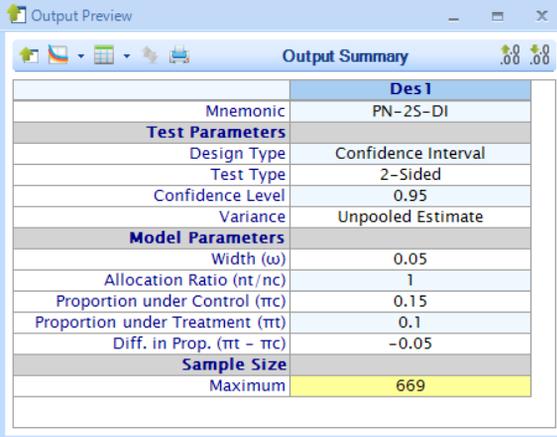
For this example, keep the default settings (Allocation Ratio = 1 and **Unpooled Estimate** selected) and click **Compute**. The sample size for this design is calculated and the output is shown as a row in the **Output Preview** window:

| ID | Design Type | Test Type | Sample Size | ω | (1- α) | n_t/n_c | π_c | Prop. Treatment (Alt.) | δ_1 | Variance |
|------|---------------------|-----------|-------------|----------|----------------|-----------|---------|------------------------|------------|-------------------|
| Des1 | Confidence Interval | 2-Sided | 669 | 0.05 | 0.95 | 1 | 0.15 | 0.1 | -0.05 | Unpooled Estimate |

As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview toolbar. The design details will be displayed, labeled **Output**

63 Confidence Interval Based Design

Summary.



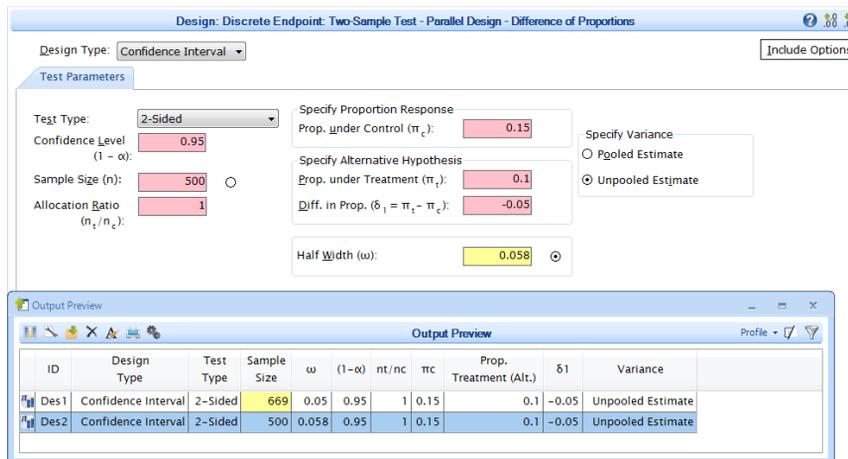
| Des 1 | |
|--|---------------------|
| Mnemonic | PN-2S-DI |
| Test Parameters | |
| Design Type | Confidence Interval |
| Test Type | 2-Sided |
| Confidence Level | 0.95 |
| Variance | Unpooled Estimate |
| Model Parameters | |
| Width (ω) | 0.05 |
| Allocation Ratio (n_t/n_c) | 1 |
| Proportion under Control (π_c) | 0.15 |
| Proportion under Treatment (π_t) | 0.1 |
| Diff. in Prop. ($\pi_t - \pi_c$) | -0.05 |
| Sample Size | |
| Maximum | 669 |

This design can be saved to the **Library** by selecting the **Des 1** in the **Output Preview** window and clicking the  icon. This test can easily be repeated for a **One-sided** confidence interval and with various values for ω , proportions of responses for treatment and control groups (π_t and π_c), and different specifications for variance estimates.

Alternatively East can compute the precision level ω given a fixed sample size using a confidence interval based design. Following the example above, to determine the precision of the estimate of population parameter where the sample size is fixed, the user has to only enter the desired value, for example $n = 500$.

Enter the following in the **Design Input** screen and click **Compute**:

Test Type: 2 sided
 Confidence Level ($1 - \alpha$): 0.95
 Sample Size (n): 500
 Allocation Ratio (n_t/n_c): 1
 Prop. under Control (π_c): 0.15
 Prop. under Treatment (π_t): 0.10
 Diff. in Prop. ($\delta_1 = \pi_t - \pi_c$): -0.05 (this will be calculated)
 Half Width (ω): Computed (select radio button)
 Specify Variance: Select **Unpooled Estimate** radio button



For a sample size of 500 the precision parameter ω is calculated to be 0.058. As the sample size is decreased, the resulting value of ω slightly increases. For binomial data, this results in a wider confidence interval for the parameter of interest. The output for all parameters is again shown as a row in the **Output Preview** window and the design can be saved to the **Library** using the standard method as with all tests in East. From there, a summary of the design can be generated using the details  icon. East also provides a very useful **Sample Size vs. Width** plot, found in the plots  menu. This dynamic visual can immediately assess how changing the sample size effects the resulting width of the confidence interval. A table of **Sample Size vs. Width** values can be generated using the **Tables**  menu, also found in the **Library**. This feature allows the user to input a range of values to generate multiple confidence intervals and the corresponding sample sizes.

63.6 Two Sample Test for the Ratio of Binomial Proportions

In experiments based on binomial data, independent samples from different populations are compared in terms of the proportion of participants presenting a particular trait or outcome of interest. For example, outcomes such as the proportion of

63 Confidence Interval Based Design

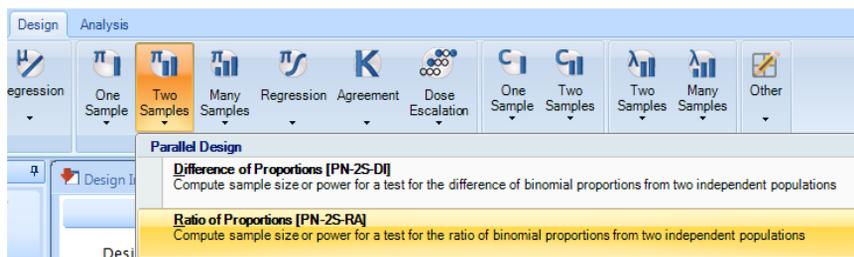
patients responding to a treatment, developing an adverse reaction, or requiring specialized care could be of interest in medical research. East supports a Confidence Interval based approach to the design of clinical trials in which this comparison is based on the ratio of proportions.

For example, consider a prospective randomized trial of a standard treatment (control arm) versus a new combination treatment (therapy arm) for patients with a heart condition, where the endpoint is either death or MI within a certain period of time after randomization. Suppose it is of interest to determine the sample size required for a trial to detect a 25% decline in the rate of such outcomes. It can be assumed that the control arm has a 30% event rate.

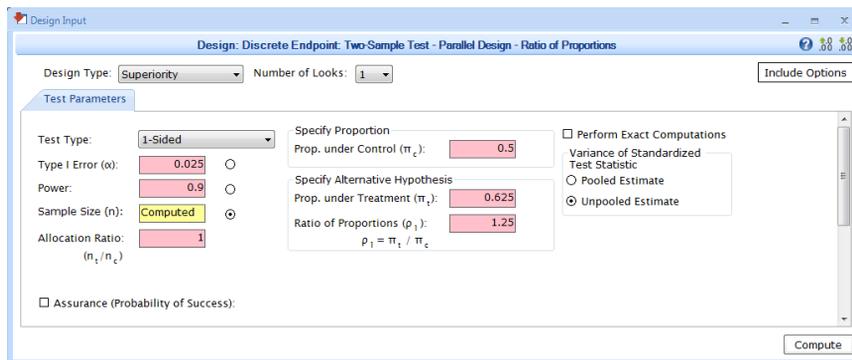
Let π_c and π_t denote the binomial probabilities for the control and treatment arms, respectively, and let $\rho = \pi_t/\pi_c$. Under H_0 , $\pi_t = \pi_c = 0.3$. A 25% decline in the event rate is thus $\rho = \pi_t/\pi_c = 0.75$. It is of interest to test the null hypothesis that $\rho = 1$ against one or two-sided alternatives. When dealing with ratios, it is mathematically more convenient to express this hypothesis in terms of the difference of the (natural) logarithms. Defining $\delta = \ln(\pi_t) - \ln(\pi_c)$ leads to the equivalent of testing $H_0: \delta = 0$. More information on this design can be found in Section 23.2

Consider a confidence interval based design of a two-arm study that compares the control arm to the combination therapy arm, where the sample size required for obtaining a 95% two-sided confidence interval for the ratio of proportions with a precision (width) of $\omega = 0.35$ must be determined.

In East under the **Design** ribbon for **Discrete** data, click **Two Samples** and then click **Parallel Design: Ratio of Proportions** as shown:

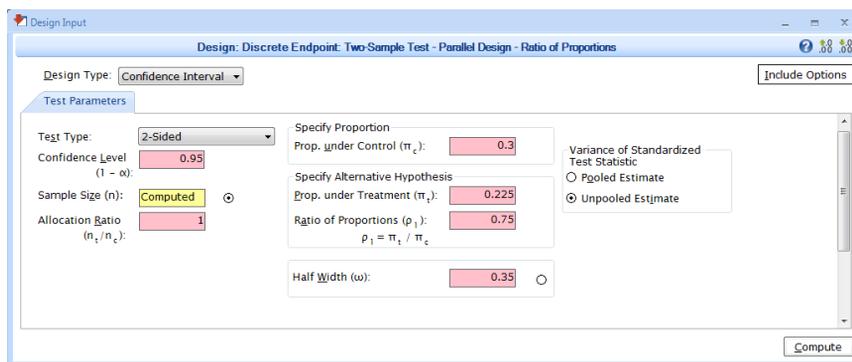


This will launch the following input window:



Choose **Confidence Interval** in the **Design Type** dropdown box and enter the following design parameters:

- Test Type: 2 sided
- Confidence Level $(1 - \alpha)$: 0.95
- Sample Size (n): Computed (select radio button)
- Allocation Ratio (n_t/n_c) : 1
- Prop. under Control (π_c) : 0.3
- Ratio of Proportions $(\rho_1 = \pi_t/\pi_c)$: 0.75
- Prop. under Treatment (π_t) : 0.225 (this will be calculated)
- Half-Width (ω) : 0.35
- Variance of Standardized Test Statistic: Select **Unpooled Estimate** radio button

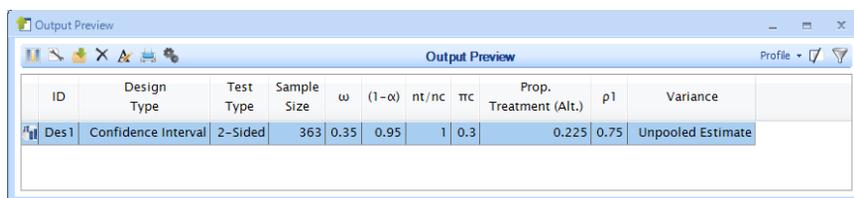


63 Confidence Interval Based Design

The Allocation Ratio ($n_t : n_c$) describes the ratio of patients to each arm. For example, an allocation ratio of 3:1 indicates that 75% of the patients are randomized to the treatment arm as opposed to 25% to the control.

In binomial designs, the variance of a random variable is dependent on its mean. The maximum sample size required for a study will be affected by how the differences of binomial response rates are standardized when computing the test statistic, regardless of the other design parameters. There are two options for determining how the test statistic will be standardized, using either the **Unpooled** or **Pooled** specification for variance. The difference becomes important when planning a binomial study with unbalanced randomization. In this case, both pooled and unpooled designs should be considered and the one that produces a tighter confidence interval (measure of ω) with fewer patients should be chosen. This will depend on the response rates of the control and treatment arms as well as the value of the fraction assigned to the treatment arm. More information on this can be found in Section 23.1.

For this example, keep the default settings (Allocation Ratio = 1 and **Unpooled Estimate** selected) and click **Compute**. The sample size for this design is calculated and the output is shown as a row in the **Output Preview** window:



| ID | Design Type | Test Type | Sample Size | ω | (1- α) | nt/nc | π_c | Prop. Treatment (Alt.) | ρ_1 | Variance |
|------|---------------------|-----------|-------------|----------|----------------|-------|---------|------------------------|----------|-------------------|
| Des1 | Confidence Interval | 2-Sided | 363 | 0.35 | 0.95 | 1 | 0.3 | 0.225 | 0.75 | Unpooled Estimate |

As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview toolbar. The design details will be displayed, labeled **Output**

Summary.

| Output Summary | |
|--|---------------------|
| | Des 1 |
| Mnemonic | PN-2S-RA |
| Test Parameters | |
| Design Type | Confidence Interval |
| Test Type | 2-Sided |
| Confidence Level | 0.95 |
| Model Parameters | |
| Width (ω) | 0.35 |
| Allocation Ratio (n_t/n_c) | 1 |
| Proportion under Control (π_c) | 0.3 |
| Proportion under Treatment (π_t) | 0.225 |
| Ratio of Proportions (π_t / π_c) | 0.75 |
| Variance | Unpooled Estimate |
| Sample Size | |
| Maximum | 363 |

This design can be saved to the **Library** by selecting the **Des 1** in the **Output Preview** window and clicking the  icon. This test can easily be repeated for a **One-sided** confidence interval and with various values for ω , proportions of responses for treatment and control groups (π_t and π_c), and different specifications for variance estimates.

Alternatively East can compute the precision level ω given a fixed sample size using a confidence interval based design. Following the example above, to determine the precision of the estimate of population parameter where the sample size is fixed, the user has to only enter the desired value, for example $n = 500$.

Enter the following in the **Design Input** screen and click **Compute**:

- Test Type: 2 sided
- Confidence Level ($1 - \alpha$): 0.95
- Sample Size (n): 500
- Allocation Ratio (n_t/n_c): 1
- Prop. under Control (π_c): 0.3
- Prop. under Treatment (π_t): 0.225(this will be calculated)
- Ratio of Proportions ($\rho_1 = \pi_t/\pi_c$): 0.75
- Half-Width (ω): Computed (select radio button)
- Variance of Standardized Test Statistic: Select **Unpooled Estimate** radio button

63 Confidence Interval Based Design

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Ratio of Proportions

Design Type: Confidence Interval

Test Parameters

Test Type: 2-Sided

Confidence Level (1 - α): 0.95

Sample Size (n): 500

Allocation Ratio (n_1/n_2): 1

Specify Proportion

Prop. under Control (π_c): 0.3

Specify Alternative Hypothesis

Prop. under Treatment (π_t): 0.225

Ratio of Proportions (ρ_1): 0.75
 $\rho_1 = \pi_t / \pi_c$

Half Width (ω): 0.298

Variance of Standardized Test Statistic

Pooled Estimate

Unpooled Estimate

Output Preview

| ID | Design Type | Test Type | Sample Size | ω | (1 - α) | nt/nc | π_c | Prop. Treatment (Alt.) | ρ_1 | Variance |
|------|---------------------|-----------|-------------|----------|-----------------|-------|---------|------------------------|----------|-------------------|
| Des1 | Confidence Interval | 2-Sided | 363 | 0.35 | 0.95 | 1 | 0.3 | 0.225 | 0.75 | Unpooled Estimate |
| Des2 | Confidence Interval | 2-Sided | 500 | 0.298 | 0.95 | 1 | 0.3 | 0.225 | 0.75 | Unpooled Estimate |

For a sample size of 500 the precision parameter ω is calculated to be 0.298. As the sample size is increased, the resulting value of ω slightly decreases. For binomial data, this results in a tighter confidence interval for the parameter of interest. The output for all parameters is again shown as a row in the **Output Preview** window and the design can be saved to the **Library** using the standard method as with all tests in East. From there, a summary of the design can be generated using the details icon. East also provides a very useful **Sample Size vs. Width** plot, found in the plots menu. This dynamic visual can immediately assess how changing the sample size effects the resulting width of the confidence interval. A table of **Sample Size vs. Width** values can be generated using the **Tables** menu, also found in the **Library**. This feature allows the user to input a range of values to generate multiple confidence intervals and the corresponding sample sizes.

63.7 Two Sample Test for the Odds Ratio of Proportions

It is often of interest to compare two independent samples from different populations in terms of the proportion of participants presenting a particular response. For example, outcomes such as the proportion of patients responding to a therapy, developing a certain side effect, or requiring specialized care are common in clinical research. East supports a Confidence Interval based approach to the design of clinical

trials for such experiments based on binomial data, in which the relationship between the odds ratio of the two populations is to be investigated.

For example, consider a prospective randomized trial where the hope is that a new experimental treatment can triple the odds ratio of exhibiting a positive outcome. The standard treatment (control arm) is compared to the new treatment (therapy arm). Suppose the goal is for the 10% response rate of the standard treatment (control) to increase to 25% for the new therapy arm.

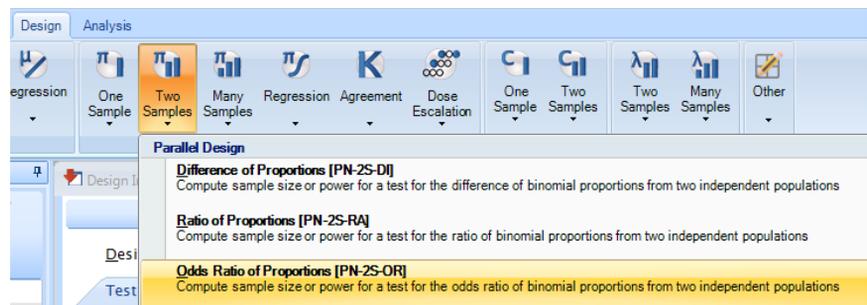
Let π_t and π_c denote the two binomial probabilities associated with the treatment and the control, respectively. The odds ratio is defined as:

$$\psi = \frac{\pi_t / (1 - \pi_t)}{\pi_c / (1 - \pi_c)} = \frac{\pi_t (1 - \pi_c)}{\pi_c (1 - \pi_t)}. \tag{63.1}$$

The problem reduces to testing $H_0: \psi = 1$ against the two-sided alternative $H_1: \psi \neq 1$ or against a one-sided alternative $H_1: \psi < 1$ or $H_1: \psi > 1$. Similar to tests dealing with the ratio of proportions, it is mathematically convenient to express the hypothesis testing of odds ratios in terms of the (natural) logarithm of ψ . Information regarding the specific details of parameter estimation for this test can be found in section 23.3

Consider a confidence interval based design for a study that compares the odds ratio of proportions between the control and experimental therapy arms. Use a two-sided test to determine the sample size required given $\pi_c = 0.1$ and $\psi_1 = 3$ with a precision parameter (width) of $\omega = 0.35$.

In East under the **Design** ribbon for **Discrete** data, click **Two Samples** and then click **Parallel Design: Odds Ratio of Proportions** as shown:



63 Confidence Interval Based Design

This will launch the following input window:

Choose **Confidence Interval** in the **Design Type** dropdown box and enter the following design parameters:

- Test Type: 2 sided
- Confidence Level ($1 - \alpha$): 0.95
- Sample Size (n): Computed (select radio button)
- Allocation Ratio (n_t/n_c): 1
- Prop. under Control (π_c): 0.1
- Prop. under Treatment (π_t): 0.25(this will be calculated)
- Odds Ratio of Proportions (ψ_1): 3
- Half Width (ω): 0.5

The Allocation Ratio ($n_t : n_c$) describes the ratio of patients to each arm. For example,

an allocation ratio of 3:1 indicates that 75% of the patients are randomized to the treatment arm as opposed to 25% to the control. For this example, keep the default Allocation Ratio = 1 and click **Compute**. The sample size for this design is calculated and the output is shown as a row in the **Output Preview** window:

| ID | Design Type | Test Type | Sample Size | ω | $(1-\alpha)$ | nt/nc | π_c | Prop. Treatment (Alt.) | ψ_1 |
|-------|---------------------|-----------|-------------|----------|--------------|-------|---------|------------------------|----------|
| Des 1 | Confidence Interval | 2-Sided | 506 | 0.5 | 0.95 | 1 | 0.1 | 0.25 | 3 |

As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview toolbar. The design details will be displayed, labeled **Output Summary**.

| Des 1 | |
|--|---------------------|
| Mnemonic | PN-2S-OR |
| Test Parameters | |
| Design Type | Confidence Interval |
| Test Type | 2-Sided |
| Confidence Level | 0.95 |
| Model Parameters | |
| Width (ω) | 0.5 |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.1 |
| Proportion under Treatment (π_t) | 0.25 |
| Odds Ratio of Proportions (ψ_1) | 3 |
| Sample Size | |
| Maximum | 506 |

This design can be saved to the **Library** by selecting the **Des 1** in the **Output Preview** window and clicking the  icon. This test can easily be repeated for a **One-sided** confidence interval and with various values for ω , different proportions of responses for treatment and control groups (π_t and π_c), and desired odds ratios of proportions (ψ_1).

Alternatively East can compute the precision level ω given a fixed sample size using a confidence interval based design. Following the example above, to determine the precision of the estimate of population parameter where the sample size is fixed, the

63 Confidence Interval Based Design

user has to only enter the desired value, for example $n = 300$.

Enter the following in the **Design Input** screen and click **Compute**:

- Test Type: 2 sided
- Confidence Level ($1 - \alpha$): 0.95
- Sample Size (n): 300
- Allocation Ratio (n_t/n_c): 1
- Prop. under Control (π_c): 0.1
- Prop. under Treatment (π_t): 0.25(this will be calculated)
- Odds Ratio of Proportions (ψ_1): 3
- Half Width (ω): Computed (select radio button)

| ID | Design Type | Test Type | Sample Size | ω | (1- α) | n_t/n_c | π_c | Prop. Treatment (Alt.) | ψ_1 |
|------|---------------------|-----------|-------------|----------|----------------|-----------|---------|------------------------|----------|
| Des1 | Confidence Interval | 2-Sided | 506 | 0.5 | 0.95 | 1 | 0.1 | 0.25 | 3 |
| Des2 | Confidence Interval | 2-Sided | 300 | 0.649 | 0.95 | 1 | 0.1 | 0.25 | 3 |

For a sample size of 300 the precision parameter ω is calculated to be 0.649. As the sample size is decreased, the resulting value of ω increases. For binomial data, this results in a wider confidence interval for the parameter of interest. The output for all parameters is again shown as a row in the **Output Preview** window and the design can be saved to the **Library** using the standard method as with all tests in East. From

there, a summary of the design can be generated using the details  icon. East

also provides a very useful **Sample Size vs. Width** plot, found in the plots  menu. This dynamic visual can immediately assess how changing the sample size effects the resulting width of the confidence interval. A table of **Sample Size vs.**

Width values can be generated using the **Tables**  menu, also found in the

Library. This feature allows the user to input a range of values to generate multiple confidence intervals and the corresponding sample sizes.

63.8 One Sample Test for McNemar’s Test for Comparing Matched Pairs

Often two binary response measurements are made on each subject, from either two different treatments or from two different time points. For example, in a comparative clinical trial, subjects are matched on baseline demographics and disease characteristics and then randomized with one subject in the pair receiving the experimental treatment and the other subject receiving the control. Another example is the cross over clinical trial in which each subject receives both treatments. By random assignment, some subjects receive the experimental treatment followed by the control while others receive the control followed by the experimental treatment. **McNemar’s Test** is used in experimental situations where such paired comparisons are observed. More specific theoretical detail about this method with examples can be found in section 22.2

The probability parameters for McNemar’s test are displayed in the following table where π_c and π_t denote the response probabilities for the control and experimental treatments, respectively.

Table 63.1: A 2 x 2 Table of Probabilities for McNemar’s Test

| Control | Experimental | | Total Probability |
|-------------------|--------------|------------|-------------------|
| | No Response | Response | |
| No Response | π_{00} | π_{01} | $1 - \pi_c$ |
| Response | π_{10} | π_{11} | π_c |
| Total Probability | $1 - \pi_t$ | π_t | 1 |

The following example taken from Section 22.2 illustrates how a confidence interval based approach to the trial design can be applied to McNemar’s test for comparing matched pairs of binomial responses. Consider a trial in which we wish to determine whether a transdermal delivery system (TDS) can be improved with a new adhesive. Subjects are to wear the old TDS (control) and new TDS (experimental) in the same area of the body for one week each. A response is said to occur if the TDS remains on for the entire one week observation period. From historical data, it is known that control has a response rate of 85% ($\pi_c = 0.85$). It is hoped that the new adhesive will increase this to 95% ($\pi_t = 0.95$). Furthermore, of the 15% of the subjects who did not respond on the control, it is hoped that 87% will respond on the experimental system.

63 Confidence Interval Based Design

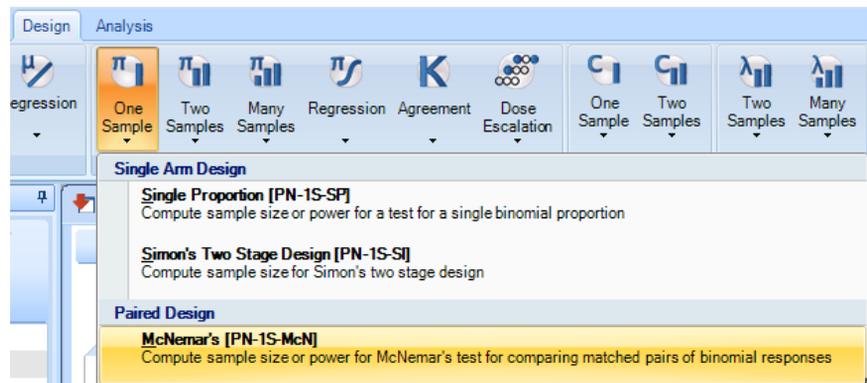
That is, $\pi_{01} = 0.87 \times 0.15 = 0.13$. Based on these data, we can fill in all the entries of Table 63.1 as follows:

Table 63.2: McNemar Probabilities for the TDS Trial

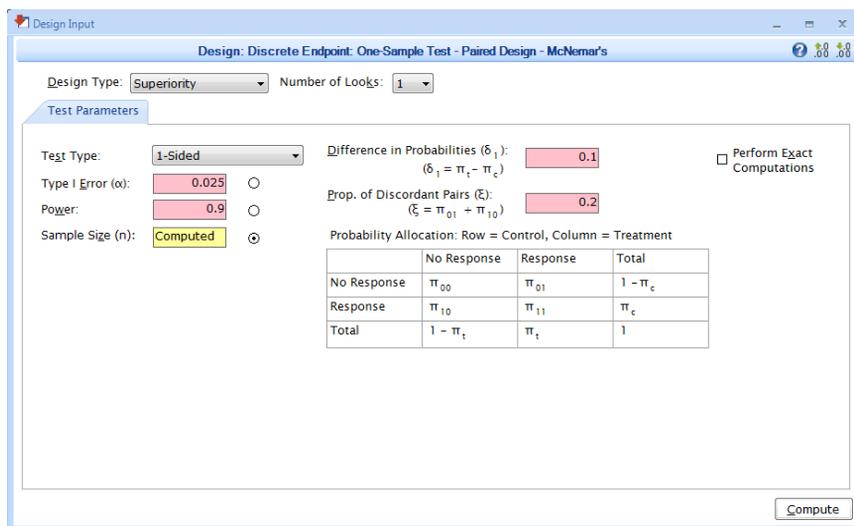
| Control | Experimental | | Total Probability |
|-------------------|--------------|----------|-------------------|
| | No Response | Response | |
| No Response | 0.02 | 0.13 | 0.15 |
| Response | 0.03 | 0.82 | 0.85 |
| Total Probability | 0.05 | 0.95 | 1 |

Although it is expected that the new adhesive will increase the adherence rate, the comparison is posed as a two-sided testing problem, testing $H_0: \pi_c = \pi_t$ against $H_1: \pi_c \neq \pi_t$ at the 0.05 level. We wish to determine the sample size for the values displayed in the above table using a Confidence Interval based design.

In East under the **Design** ribbon for **Discrete** data, click **One Sample** and then click **Paired Design: McNemar's** as shown:



This will launch the following input window:



Choose **Confidence Interval** in the **Design Type** dropdown box. Consider a two sided test with 5% significance level and specify $\delta_1 = \pi_t - \pi_c = 0.1$ and $\xi = \pi_{01} + \pi_{10} = 0.16$ with a precision (width) of 0.5 units.

Enter the following design parameters:

- Test Type: 2 sided
- Confidence Level $(1 - \alpha)$: 0.95
- Sample Size (n): Computed (select radio button)
- Half Width (ω) : 0.5
- Difference in Probabilities (δ_1) : 0.1
- Proportion of Discordant Pairs ξ : 0.16

63 Confidence Interval Based Design

Design Input
Design: Discrete Endpoint: One-Sample Test - Paired Design - McNemar's

Design Type: Confidence Interval

Test Parameters

Test Type: 2-Sided

Confidence Level (1 - α): 0.95

Sample Size (n): Computed

Half Width (ω): 0.5

Difference in Probabilities (δ_1): ($\delta_1 = \pi_t - \pi_c$): 0.1

Prop. of Discordant Pairs (ξ): ($\xi = \pi_{01} + \pi_{10}$): 0.16

Probability Allocation: Row = Control, Column = Treatment

| | No Response | Response | Total |
|-------------|-------------|------------|-------------|
| No Response | π_{00} | π_{01} | $1 - \pi_c$ |
| Response | π_{10} | π_{11} | π_c |
| Total | $1 - \pi_t$ | π_t | 1 |

Compute

Click **Compute**. The sample size for this design is calculated and the output is shown as a row in the **Output Preview** window:

Output Preview

| ID | Design Type | Test Type | Sample Size | ω | (1 - α) | δ_1 | ξ |
|------|---------------------|-----------|-------------|----------|-----------------|------------|-------|
| Des1 | Confidence Interval | 2-Sided | 631 | 0.5 | 0.95 | 0.1 | 0.16 |

As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview toolbar. The design details will be displayed, labeled **Output**

Summary.

| Des 1 | |
|--|---------------------|
| Mnemonic | PN-15-McN |
| Test Parameters | |
| Design Type | Confidence Interval |
| Test Type | 2-Sided |
| Confidence Level | 0.95 |
| Model Parameters | |
| Width (ω) | 0.5 |
| Difference in Probabilities (δ_1) | 0.1 |
| Prop. of Discordant Pairs (ξ) | 0.16 |
| Sample Size | |
| Maximum | 631 |

This design can be saved to the **Library** by selecting the **Des 1** in the **Output Preview** window and clicking the  icon. From there, a summary of the design can be generated using the details  icon. East also provides a very useful **Sample Size vs. Width** plot, found in the plots  menu. This dynamic visual can immediately assess how changing the sample size effects the resulting width of the confidence interval. A table of **Sample Size vs. Width** values can be generated using the **Tables**  menu, also found in the **Library**. This feature allows the user to input a range of values to generate multiple confidence intervals and the corresponding sample sizes.

This test can easily be repeated for a **one-sided** confidence interval and with various values for ω and difference in probabilities (δ_1) or proportion of discordant pairs (ξ). East can also compute the precision level ω for a given fixed sample size using a confidence interval based design for McNemar’s test. Following the example above, the precision of the estimate (ω) of population parameter can easily be determined.

63.9 Many Sample Test - One Way ANOVA

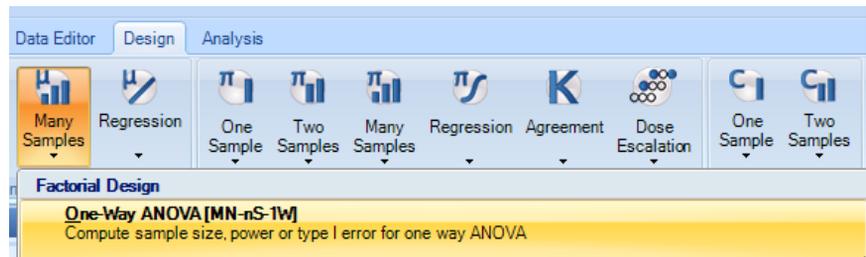
East offers the capability to design trials comparing more than two continuous means. A **One-Way ANOVA** tests the equality of means across R independent groups. The two sample difference of means test for independent data is a one-way ANOVA test for 2 groups. More information, including the following example which is modified here to illustrate a confidence interval based approach to the trial design, can be found in Section 22.2.

63 Confidence Interval Based Design

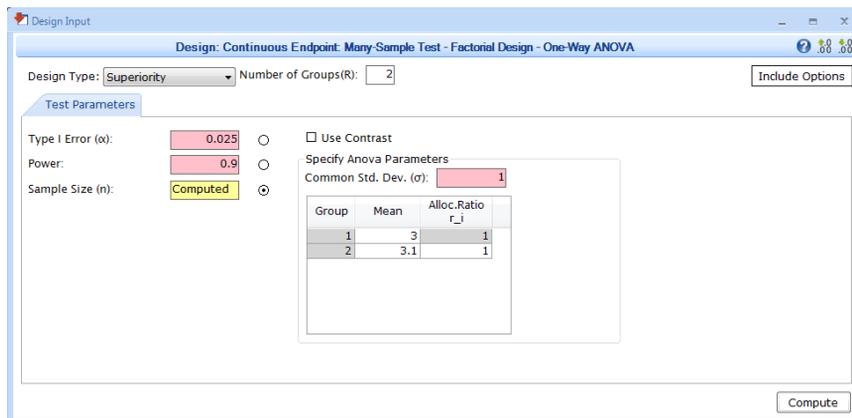
Suppose n patients have been allocated randomly to R treatments. We assume that the data of the R treatment groups comes from R normally distributed populations with the same variance σ^2 , and with population means $\mu_1, \mu_2, \dots, \mu_R$. The null hypothesis $H_0 : \mu_1 = \mu_2 = \dots = \mu_R$ is tested against the alternative hypothesis H_1 : for at least one pair (i, j) , $\mu_i \neq \mu_j$, where $i, j = 1, 2, \dots, R$.

Consider a clinical trial with four groups of patients where the goal is to study the efficacy of a treatment protocol. Three different doses of a drug are being compared against placebo in patients with Alzheimer’s disease. Suppose, based on historical data, the expected mean responses are 0, 1.5, 2.5, and 2, for Groups 1 to 4, respectively. The common standard deviation within each group is $\sigma = 3.5$. We wish to compute the required sample size using a confidence interval based design with a type-1 error of 5% and precision estimate of $\omega = 2$.

In East under the **Design** ribbon for **Continuous** data, click **Many Samples** and then click **Factorial Design: One Way ANOVA** as shown:



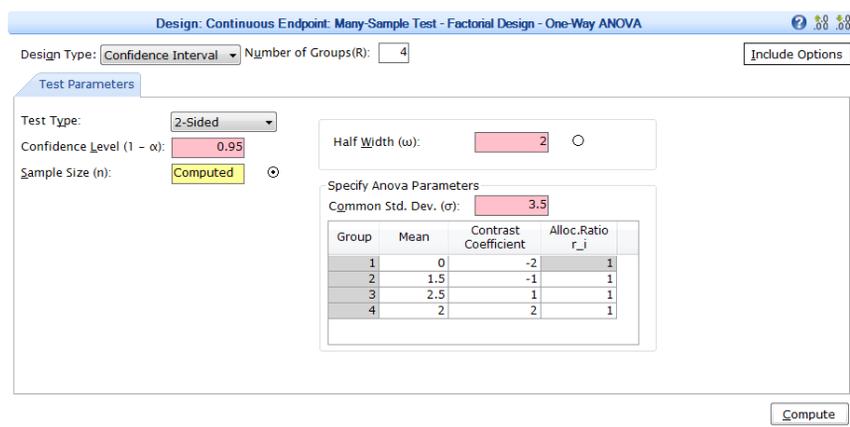
This will launch the following input window:



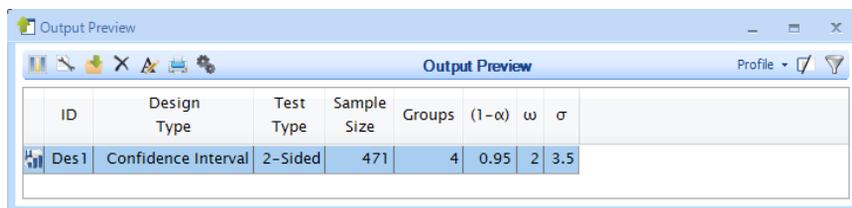
Choose **Confidence Interval** in the **Design Type** dropdown box and enter the following design parameters:

- Number of Groups(R): 4
- Test type: 2 sided
- Confidence level $(1 - \alpha)$: 0.95
- Sample size (n): Computed (select radio button)
- One-sided Width(ω): 2
- Common Standard Deviation (σ): 3.5
- Group 1: Mean= 0
- Group 2: Mean= 1.5
- Group 3: Mean= 2.5
- Group 4: Mean= 2

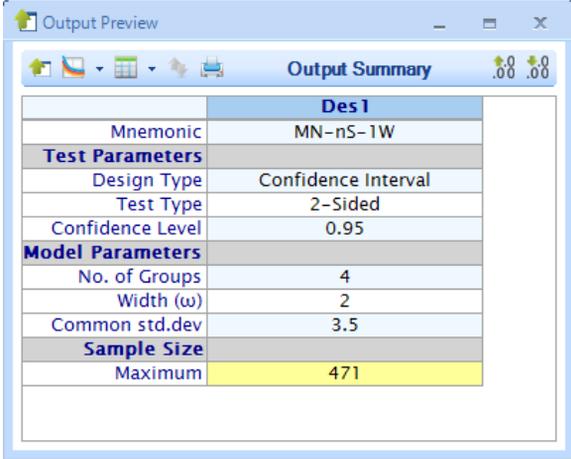
63 Confidence Interval Based Design



Leave all other Group values (Contrast Coefficients and Allocation Ratios) as defaults and click **Compute**. The sample size for this design is calculated and the output is shown as a row in the **Output Preview** window:



As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview toolbar. The design details will be displayed, labeled **Output**

Summary.


| Des 1 | |
|-------------------------|---------------------|
| Mnemonic | MN-nS-1W |
| Test Parameters | |
| Design Type | Confidence Interval |
| Test Type | 2-Sided |
| Confidence Level | 0.95 |
| Model Parameters | |
| No. of Groups | 4 |
| Width (w) | 2 |
| Common std.dev | 3.5 |
| Sample Size | |
| Maximum | 471 |

This design can be saved to the **Library** by selecting the **Des 1** in the **Output Preview** window and clicking the  icon. This test can easily be repeated for a **One-sided** confidence interval and with various values for w and σ , as well as any desired differences in group means, contrast coefficients or group allocation ratios.

Alternatively East can compute the precision level w given a fixed sample size using a confidence interval based design. Following the example above, to determine the precision of the estimate of population parameter where the sample size is fixed, the user has to only enter the desired value, for example $n = 300$.

In the **Design Window** the parameters now become:

Number of Groups(R): 4
 Test type: 2 sided
 Confidence level ($1 - \alpha$): 0.95
 Sample size (n): 300
 Half Width(w): Computed (select radio button)
 Common Standard Deviation (σ): 3.5
 Group 1: Mean= 0
 Group 2: Mean= 1.5
 Group 3: Mean= 2.5
 Group 4: Mean= 2

63 Confidence Interval Based Design

Enter the above in the **Design Input** screen and click **Compute**:

The screenshot shows the 'Design Input' screen for a 'Confidence Interval' based design. The 'Design Type' is set to 'Confidence Interval' and the 'Number of Groups (R)' is 4. The 'Test Parameters' section shows a '2-Sided' test with a 'Confidence Level (1 - alpha)' of 0.95 and a 'Sample Size (n)' of 300. The 'Half Width (omega)' is 2.505. The 'Specify Anova Parameters' section shows a 'Common Std. Dev. (sigma)' of 3.5 and a table of Contrast Coefficients and Alloc. Ratio for four groups.

| Group | Mean | Contrast Coefficient | Alloc. Ratio r _j |
|-------|------|----------------------|-----------------------------|
| 1 | 0 | -2 | 1 |
| 2 | 1.5 | -1 | 1 |
| 3 | 2.5 | 1 | 1 |
| 4 | 2 | 2 | 1 |

The 'Output Preview' window shows the following table:

| ID | Design Type | Test Type | Sample Size | Groups | (1-α) | ω | σ |
|------|---------------------|-----------|-------------|--------|-------|-------|-----|
| Des1 | Confidence Interval | 2-Sided | 471 | 4 | 0.95 | 2 | 3.5 |
| Des2 | Confidence Interval | 2-Sided | 300 | 4 | 0.95 | 2.505 | 3.5 |

The precision parameter ω is calculated to be 2.505. As the sample size is decreased, the resulting value of ω increases. In other words, the precision limit increases, resulting in a wider confidence interval for the parameter of interest. The output for all parameters is again shown as a row in the **Output Preview** window and the design can be saved to the **Library** using the standard method as with all tests in East. From

there, a summary of the design can be generated using the details  icon. East

also provides a very useful **Sample Size vs. Width** plot, found in the plots  menu. This dynamic visual can immediately assess how changing the sample size effects the resulting width of the confidence interval. A table of **Sample Size vs.**

Width values can be generated using the **Tables**  menu, also found in the **Library**. This feature allows the user to input a range of values to generate multiple confidence intervals and the corresponding sample sizes.

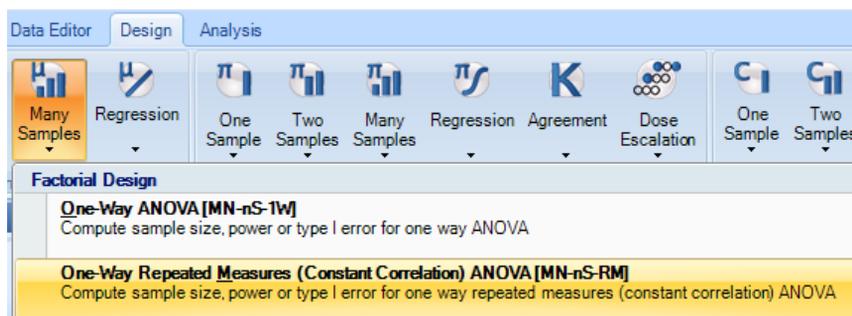
63.10 Many Sample Test - One Way Repeated Measures

The **One Way Repeated ANOVA** tests for equality of means in a repeated measures setting. As the patient population is exposed to each treatment, the measurement of the dependent variable is repeated, resulting in correlation between observations from the same patient. Constant correlation assumes that the correlation between observations from the same patient is constant for all patients. This correlation parameter (ρ) needs

to be specified in the one way repeated measures study design.

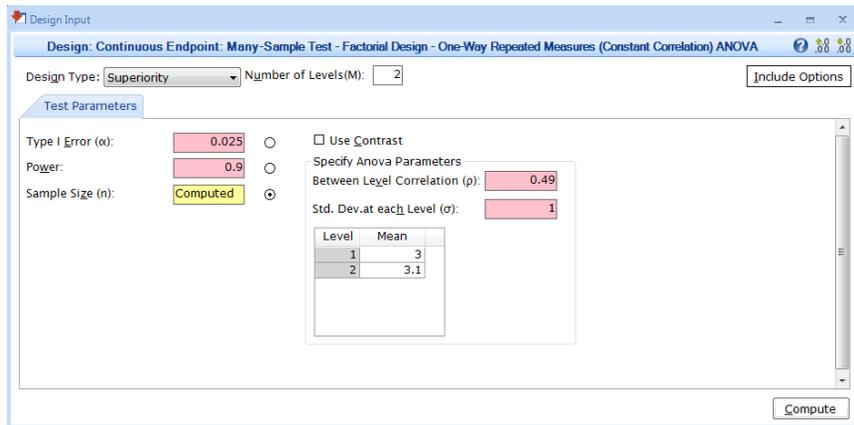
Consider a hypothetical longitudinal study that investigates the effect of a dietary intervention on weight loss, where the endpoint is decrease in weight (in kilograms) from baseline. Data is collected at four time points: baseline, 4 weeks, 8 weeks, and 12 weeks and are measured to be 0 kg, 10.5 kg, 25kg, and 20kg respectively. Assume the common standard deviation within each group (i.e. at each level) is $\sigma = 3.5$ and the constant correlation (between level) $\rho = 0.2$. We wish to compute the required sample size for this study, using a two-sided confidence interval based design with a type-1 error of 5% and precision estimate (width) of $\omega = 2$.

In East under the **Design** ribbon for **Continuous** data, click **Many Samples** and then click **Factorial Design: One Way Repeated Measures (Constant Correlation) ANOVA** as shown:



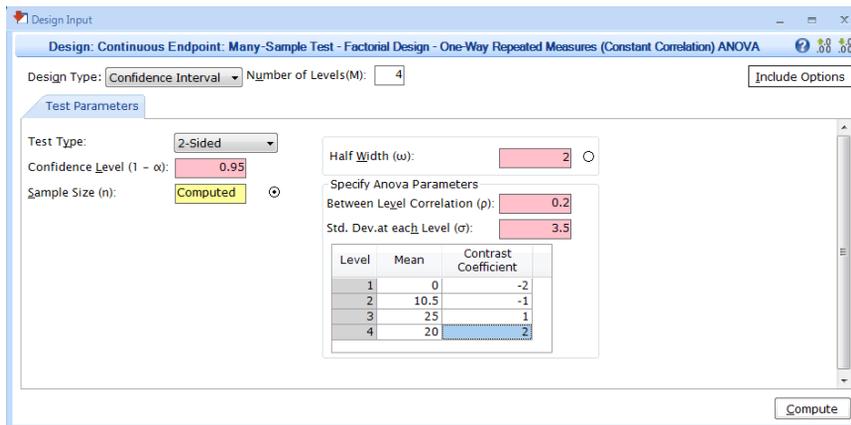
63 Confidence Interval Based Design

This will launch the following input window:

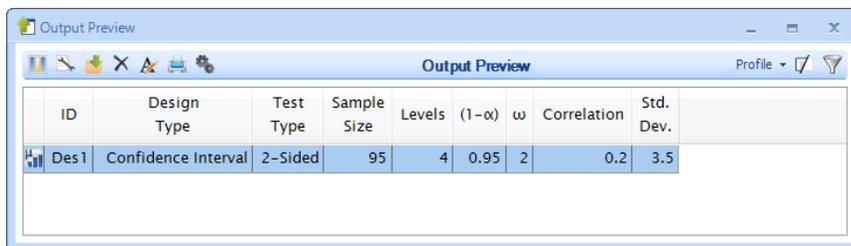


Choose **Confidence Interval** in the **Design Type** dropdown box and enter the following design parameters:

- Number of Levels (M): 4
- Test type: 2 sided
- Confidence level ($1 - \alpha$): 0.95
- Sample size (n): Computed (select radio button)
- Half Width(ω): 2
- Between Level Correlation (ρ): 0.2
- Standard Deviation at each Level (σ): 3.5
- Group 1: Mean= 0
- Group 2: Mean= 10.5
- Group 3: Mean= 25
- Group 4: Mean= 20



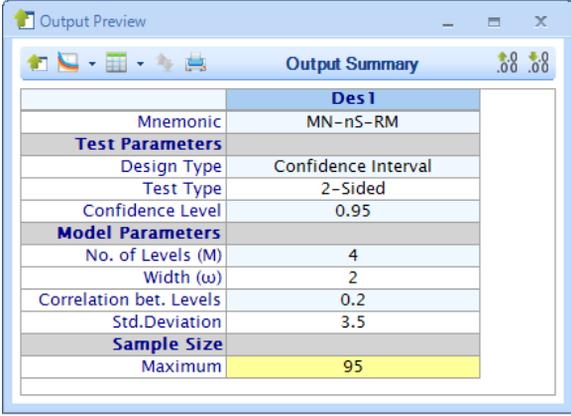
Leave all other Group level values (Contrast coefficients) as defaults and click **Compute**. The sample size for this design is calculated and the output is shown as a row in the **Output Preview** window:



As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview toolbar. The design details will be displayed, labeled **Output**

63 Confidence Interval Based Design

Summary.



| Des 1 | |
|-------------------------|---------------------|
| Mnemonic | MN-nS-RM |
| Test Parameters | |
| Design Type | Confidence Interval |
| Test Type | 2-Sided |
| Confidence Level | 0.95 |
| Model Parameters | |
| No. of Levels (M) | 4 |
| Width (ω) | 2 |
| Correlation bet. Levels | 0.2 |
| Std.Deviation | 3.5 |
| Sample Size | |
| Maximum | 95 |

This design can be saved to the **Library** by selecting the **Des 1** in the **Output Preview** window and clicking the  icon. This test can easily be repeated for a **One-sided** confidence interval and with various values for ω , ρ and σ , as well as any desired differences in group information.

Alternatively East can compute the precision level ω given a fixed sample size using a confidence interval based design. Following the example above, to determine the precision of the estimate of population parameter where the sample size is fixed, the user has to only enter the desired value, for example increase n from 95 to $n = 200$.

In the **Design Window** the parameters now become:

Number of Levels(M): 4
 Test type: 2 sided
 Confidence level ($1 - \alpha$): 0.95
 Sample size (n): 200
 Half Width(ω): Computed (radio button selected)
 Between Level Correlation (ρ): 0.2
 Standard Deviation at each Level (σ): 3.5
 Group 1: Mean= 0
 Group 2: Mean= 10.5
 Group 3: Mean= 25
 Group 4: Mean= 20

Enter the above in the **Design Input** screen and click **Compute**:

The screenshot shows the 'Design: Continuous Endpoint - Many-Sample Test - Factorial Design - One-Way Repeated Measures (Constant Correlation) ANOVA' interface. The 'Design Type' is 'Confidence Interval' and 'Number of Levels(M)' is 4. Under 'Test Parameters', 'Test Type' is '2-Sided', 'Confidence Level (1 - α)' is 0.95, and 'Sample Size (n)' is 200. Under 'Specify Anova Parameters', 'Half Width (ω)' is 1.372, 'Between Level Correlation (ρ)' is 0.2, and 'Std. Dev. at each Level (σ)' is 3.5. A table below shows levels 1-4 with means 0, 10.5, 25, 20 and contrast coefficients -2, -1, 1, 2. The 'Output Preview' window at the bottom shows a table with two rows: Des1 (Sample Size 95, Levels 4, (1-α) 0.95, ω 2, Correlation 0.2, Std. Dev. 3.5) and Des2 (Sample Size 200, Levels 4, (1-α) 0.95, ω 1.372, Correlation 0.2, Std. Dev. 3.5).

The precision parameter ω is calculated to be 1.372. As the sample size is increased, the resulting value of ω decreases. In other words, the precision limit decreases, resulting in a tighter confidence interval for the parameter of interest. The output for all parameters is again shown as a row in the **Output Preview** window and the design can be saved to the **Library** using the standard method as with all tests in East. From there, a summary of the design can be generated using the details icon. East also provides a very useful **Sample Size vs. Width** plot, found in the plots menu. This dynamic visual can immediately assess how changing the sample size effects the resulting width of the confidence interval. A table of **Sample Size vs. Width** values can be generated using the **Tables** menu, also found in the **Library**. This feature allows the user to input a range of values to generate multiple confidence intervals and the corresponding sample sizes.

63.11 Normal Test for Linear Regression - Single Slope

Regression models are often used to examine the relationship between a response and one or more explanatory variables. A simple linear regression model tests a single slope for one continuous covariate when the relationship with response is linear. The assumption is that the observed value of a response variable Y is a linear function of the explanatory variable X , plus some random noise.

63 Confidence Interval Based Design

For $i = 1, \dots, n$ subjects in a study the model can be written as:

$$Y_i = \gamma + \theta X_i + \epsilon_i$$

where each ϵ_i is an independent normal random variable with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma_\epsilon^2$. X_i (subject i) is a random variable with a variance σ_x^2 . More information on simple linear regression models, including distinctions between different types of studies and details on the calculation of the test statistic can be found in Section 19.1.

A dose-response relationship describes the effect of an exposure on an outcome (positive or negative) and is a crucial consideration in the development of a drug or other treatment. The relationship is often determined by estimating the slope of a regression model such as the one above, where Y is the appropriate response variable and the explanatory variable X is a set of specified doses. Consider a hypothetical clinical trial involving different doses of a medication under study. Assume that the doses and randomization of subjects across the doses have been chosen so that the standard deviation $\sigma_x = 9$. Based on information gained from prior studies, it can be assumed that $\sigma_\epsilon = 15$.

When the slope of the linear regression model is 0, the relationship between the outcome and covariate is flat. In other words, there is no evidence of a dose-response relationship. It therefore of interest to test the null hypothesis $H_0: \theta = 0$ against a two-sided alternative $H_1: \theta \neq 0$.

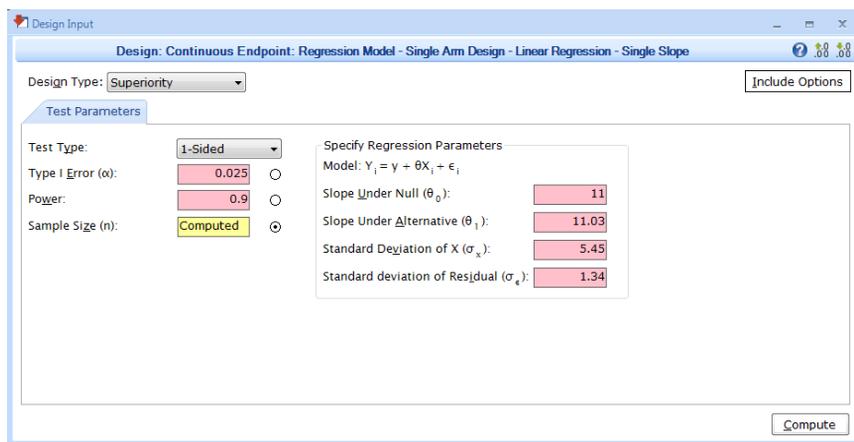
Consider a confidence interval based design for the above study to determine if a dose-response relationship exists between the patient outcome and dose level of a drug. Use a two-sided test with a type-1 error rate of 5% to compute the sample size required using a precision parameter (width) of $\omega = 0.15$.

To illustrate this example, in East under the **Design** ribbon for **Continuous** data, click **Regression** and then click **Single Arm Design: Linear Regression - Single Slope** as

shown:



This will launch the following input window:



Choose **Confidence Interval** in the **Design Type** dropdown box, enter the following design parameters and click **Compute**:

- Test type: 2 sided
- Confidence level $(1 - \alpha)$: 0.95
- Sample size (n): Computed (select radio button)
- Half Width (ω): 0.15
- Standard Deviation of $X(\sigma_X)$: 9
- Standard Deviation of Residuals $X(\sigma_\epsilon)$: 15

63 Confidence Interval Based Design

Design: Continuous Endpoint: Regression Model - Single Arm Design - Linear Regression - Single Slope

Design Type: Confidence Interval

Test Parameters

Test Type: 2-Sided

Confidence Level (1 - α): 0.95

Sample Size (n): Computed

Half Width (ω): 0.15

Standard Deviation of X (σ_x): 9

Standard Deviation of Residuals (σ_e): 15

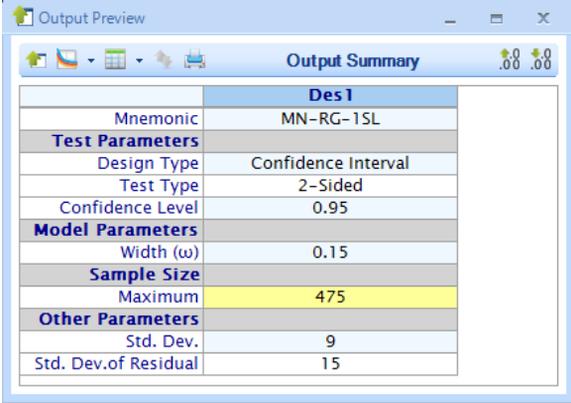
Compute

The sample size for this design is calculated and the output is shown as a row in the **Output Preview** window:

| ID | Design Type | Test Type | Sample Size | (1- α) | ω | Std. Dev. | Std. Dev. of Residual |
|-------|---------------------|-----------|-------------|----------------|----------|-----------|-----------------------|
| Des 1 | Confidence Interval | 2-Sided | 475 | 0.95 | 0.15 | 9 | 15 |

As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview toolbar. The design details will be displayed, labeled **Output**

Summary.



| Des 1 | |
|-------------------------|---------------------|
| Mnemonic | MN-RG-1SL |
| Test Parameters | |
| Design Type | Confidence Interval |
| Test Type | 2-Sided |
| Confidence Level | 0.95 |
| Model Parameters | |
| Width (ω) | 0.15 |
| Sample Size | |
| Maximum | 475 |
| Other Parameters | |
| Std. Dev. | 9 |
| Std. Dev. of Residual | 15 |

This design can be saved to the **Library** by selecting the **Des 1** in the **Output Preview** window and clicking the  icon. This test can easily be repeated for a **One-sided** confidence interval and with various values for ω , σ_X , and σ_ϵ .

Alternatively East can compute the precision level ω given a fixed sample size using a confidence interval based design. Following the example above, to determine the precision of the estimate of population parameter where the sample size is fixed, the user has to only enter the desired value, for example $n = 200$.

In the **Design Window** the parameters now become:

Test type: 2 sided
 Confidence level ($1 - \alpha$): 0.95
 Sample size (n): 200
 Half Width (ω): Computed (select radio button)
 Standard Deviation of $X(\sigma_X)$: 9
 Standard Deviation of Residuals $X(\sigma_\epsilon)$: 15

63 Confidence Interval Based Design

Enter the following in the **Design Input** screen and click **Compute**:

Design: Continuous Endpoint: Regression Model - Single Arm Design - Linear Regression - Single Slope

Design Type: Confidence Interval

Test Parameters

Test Type: 2-Sided

Confidence Level (1 - α): 0.95

Sample Size (n): 200

Half Width (ω): 0.231

Standard Deviation of X (σ_x): 9

Standard Deviation of Residuals (σ_e): 15

Output Preview

| ID | Design Type | Test Type | Sample Size | (1 - α) | ω | Std. Dev. | Std. Dev. of Residual |
|------|---------------------|-----------|-------------|-----------------|----------|-----------|-----------------------|
| Des1 | Confidence Interval | 2-Sided | 475 | 0.95 | 0.15 | 9 | 15 |
| Des2 | Confidence Interval | 2-Sided | 200 | 0.95 | 0.231 | 9 | 15 |

Compute

The precision parameter ω is calculated to be 0.231. When the sample size is decreased the estimate of the precision limit increases leading to a wider confidence interval for the parameter of interest. The output for all parameters is again shown as a row in the **Output Preview** window and the design can be saved to the **Library** using the standard method as with all tests in East. From there, a summary of the design can be generated using the details icon. East also provides a very useful **Sample Size vs. Width** plot, found in the plots menu. This dynamic visual can immediately assess how changing the sample size effects the resulting width of the confidence interval. A table of **Sample Size vs. Width** values can be generated using the **Tables** menu, also found in the **Library**. This feature allows the user to input a range of values to generate multiple confidence intervals and the corresponding sample sizes.

63.12 Normal Test for Linear Regression - Difference of Slopes

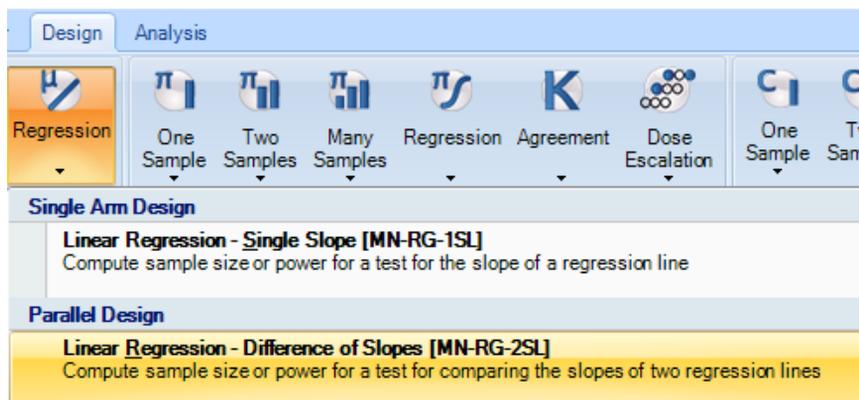
Linear regression models are used to examine the relationship between a response variable and one or more explanatory variables assuming that the relationship is linear. One type of linear regression tests the equality of two slopes in a model with only one observation per subject. In such experimental situations, it is of interest to compare the slopes of two regression lines.

The regression model relates the response variable Y to the explanatory variable X using the model $Y_{il} = \gamma + \theta_i X_{il} + \epsilon_{il}$, where the error ϵ_{il} has a normal distribution with mean zero and an unknown variance σ_ϵ^2 for Subject l in Treatment i , $i = c, t$ and $l = 1, \dots, n_i$. Let σ_{xc}^2 and σ_{xt}^2 denote the variance of the explanatory variable X for control (c) and treatment (t), respectively. More information on linear regression models for comparing two slopes and details on the calculation of the test statistic can be found in Section 19.2.

Suppose a treatment response depends on the level of a certain laboratory parameter. A new formulation is to be developed to decrease this interaction between the response and the level. The explanatory variable is the baseline value of the laboratory parameter. The study is designed with $\sigma_{xc} = \sigma_{xt} = 6$ and $\sigma_\epsilon = 10$. It is of interest to test the equality of the slopes θ_c and θ_t under the null hypothesis $H_0: \theta_t = \theta_c$ against the two-sided alternative $H_1: \theta_t \neq \theta_c$.

Consider a confidence interval based design for the above study to determine if there exists a difference between the slopes of the two regression lines. Use a two-sided test with a type-1 error rate of 5% to compute the sample size required using a precision parameter (width) of $\omega = 0.5$.

To illustrate this example, in East under the **Design** ribbon for **Continuous** data, click **Regression** and then click **Parallel Design: Linear Regression - Difference of Slopes** as shown:



63 Confidence Interval Based Design

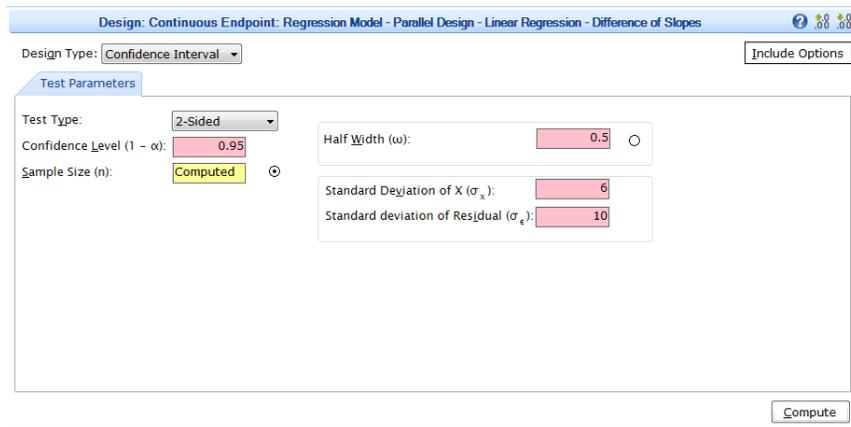
This will launch the following input window:

Choose **Confidence Interval** in the **Design Type** dropdown box, enter the following design parameters and click **Compute**:

- Test type: 2 sided
- Confidence level $(1 - \alpha)$: 0.95
- Sample size (n): Computed (select radio button)
- Allocation Ratio (n_t/n_c) : 1
- Half Width (ω) : 0.5
- Standard Deviation of $X(\sigma_x)$: 6
- Standard Deviation of Residual $X(\sigma_\epsilon)$: 10

The Allocation Ratio $(n_t : n_c)$ describes the ratio of patients to each arm. For example, an allocation ratio of 3:1 indicates that 75% of the patients are randomized to the treatment arm as opposed to 25% to the control. Keep the default allocation

ratio(n_t/n_c)= 1.



The sample size for this design is calculated and the output is shown as a row in the **Output Preview** window:

| ID | Design Type | Test Type | Sample Size | (1-α) | ω | Std. Dev. of X | Std. Dev. of Residual |
|------|---------------------|-----------|-------------|-------|-----|----------------|-----------------------|
| Des1 | Confidence Interval | 2-Sided | 171 | 0.95 | 0.5 | 6 | 10 |

As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview toolbar. The design details will be displayed, labeled **Output**

63 Confidence Interval Based Design

Summary.

| Des 1 | |
|-------------------------|---------------------|
| Mnemonic | MN-RG-2SL |
| Test Parameters | |
| Design Type | Confidence Interval |
| Test Type | 2-Sided |
| Confidence Level | 0.95 |
| Model Parameters | |
| Width (ω) | 0.5 |
| Sample Size | |
| Maximum | 171 |
| Other Parameters | |
| Std. Dev. | 6 |
| Std. Dev. of Residual | 10 |

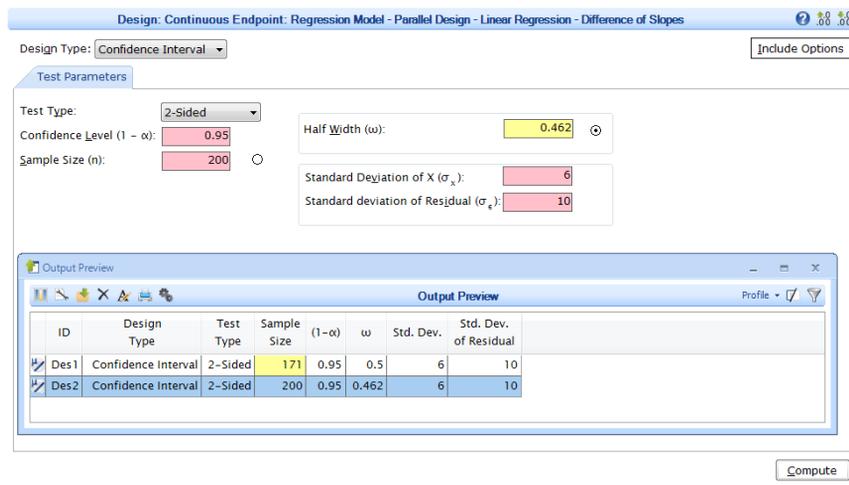
This design can be saved to the **Library** by selecting the **Des 1** in the **Output Preview** window and clicking the  icon. This test can easily be repeated for a **One-sided** confidence interval and with various values for ω , σ_x , and σ_ϵ .

Alternatively, East can compute the precision level ω given a fixed sample size using a confidence interval based design. Following the example above, to determine the precision of the estimate of population parameter where the sample size is fixed, the user has to only enter the desired value, for example $n = 200$.

In the **Design Window** the parameters now become:

Test type: 2 sided
 Confidence level ($1 - \alpha$): 0.95
 Sample size (n): 200
 Allocation Ratio (n_t/n_c): 1
 Half Width (ω): Computed (select radio button)
 Standard Deviation of X (σ_x): 6
 Standard Deviation of Residual X (σ_ϵ): 10

Enter the following in the **Design Input** screen and click **Compute**:



The precision parameter ω is calculated to be 0.462. As the sample size increases the precision limit decreases, providing a tighter confidence interval for the parameter of interest. The output for all parameters is again shown as a row in the **Output Preview** window and the design can be saved to the **Library** using the standard method as with all tests in East. From there, a summary of the design can be generated using the details  icon. East also provides a very useful **Sample Size vs. Width** plot, found in the plots  menu. This dynamic visual can immediately assess how changing the sample size effects the resulting width of the confidence interval. A table of **Sample Size vs. Width** values can be generated using the **Tables**  menu, also found in the **Library**. This feature allows the user to input a range of values to generate multiple confidence intervals and the corresponding sample sizes.

64 Simulation in East

East lets you simulate studies that were created by its design module. This chapter describes the simulations that are available in East. Through these simulation capabilities, you can repeatedly generate the entire path traced out by a test statistic under user-specified assumptions about treatment effects. Thereby you can verify various operating characteristics of your designs.

64.1 Normal Studies

To begin let us design a study. Click **Continuous: Two Samples** on the **Design** tab and then click **Parallel Design: Difference of Means**. Enter the design parameters as shown below.

Use the default boundary information and click **Compute** to create the design. The output summary is shown below.

| | Des1 |
|---|---------------------|
| Mnemonic | MN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 5 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.909 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Input Method | Difference of Means |
| Diff. in Means ($\delta = \mu_1 - \mu_2$) | 60 |
| Std. Deviation (σ) | 45 |
| Test Statistic | Z |
| Boundary Parameters | |
| Spacing of Looks | Unequal |
| Efficacy Boundary | LD (OF) |
| Sample Size | |
| Maximum | 25 |
| Expected Under H0 | 24.836 |
| Expected Under H1 | 18.347 |

The study is designed for up to 5 looks with the $LD(OF)$ spending function, and a

two-sided α of 0.05. At most 25 patients are needed in order to achieve 90% power with this large standardized treatment effect of $60/45 = 1.333$. Save the design to the workbook and click the **S** icon. You will be taken to the following simulation worksheet.

Number of Looks: 5

Test Parameters Response Generation Simulation Controls

Trial Type: Superiority Test Statistic: t

Test Type: 2-Sided Variance: Equal

Sample Size (n): 25

| Look # | Info. Fraction | Cum. α Spent | | Efficacy Z | |
|--------|----------------|---------------------|-------|------------|--------|
| | | Upper | Lower | Upper | Lower |
| 1 | 0.200 | 0.000 | 0.000 | 4.877 | -4.877 |
| 2 | 0.400 | 0.000 | 0.000 | 3.357 | -3.357 |
| 3 | 0.600 | 0.004 | 0.004 | 2.680 | -2.680 |
| 4 | 0.800 | 0.012 | 0.012 | 2.290 | -2.290 |
| 5 | 1.000 | 0.025 | 0.025 | 2.031 | -2.031 |

Restore Original Design

Notice the **Test Statistic** option on the right. For normally distributed data with known variance one would select the **z-test** option from the drop down menu. These simulations are accurate regardless of sample size.

However, for normally distributed responses with unknown variance, selecting the **z-test** option will result in simulations that may not be valid for designs with small sample sizes. This is because with small samples, the Wald test statistic at any monitoring time-point has a student-t distribution rather than a standard normal distribution. The stopping boundaries in East exhibit type-1 error and power exactly as specified only if the sequentially computed test statistic is normally distributed and has independent increments. To the extent that the test statistic relies on large sample theory for its distributional behavior, there may be some loss of accuracy in the operating characteristics of the stopping boundaries. For sample sizes exceeding 100, the loss of accuracy is scarcely noticeable. However when the sample size is of the order of 20, there is indeed a noticeable loss of accuracy and the study must be re-designed and simulated repeatedly until, by trial and error, it possesses the required type-1 error and power.

Let us now illustrate this with an example. At the interim monitoring stage we will be

64 Simulation in East

tracking the Wald statistic. Thus we should simulate the behavior of this statistic ahead of time and verify that the type-1 error and power of the study are indeed as specified. Suppose we have accrued a total of n_j subjects by the j th look. We have two choices for computing the test statistic and checking if it has crossed a stopping boundary.

1. Use the value of $\sigma = 45$ specified at the design stage and compute

$$Z_j = \frac{\bar{X}_{tj} - \bar{X}_{cj}}{\sqrt{\frac{4\sigma^2}{n_j}}} . \tag{64.1}$$

This statistic is normally distributed with variance 1 and a known correlation structure across different values of j . Consequently, it should produce the precise type 1 error and power specified in the study design even though the maximum sample size is only 24. To do this select **z-test** from the drop down menu next to **Test Statistic**. Next, click on the **Response Generation Info** tab and enter the parameters as shown below.

Next, click the **Simulate** button. The simulation intermediate window will appear as shown below:

| Look # | Look Position | H0+ | H0- | H1+ | H1- | Latest Simul ... Test S... | Average Info. | Average Sample Size | Cum. # Rejecting... Up(H0+) | Cum. # Rejecting... Low(H0-) | Cum. # Unable... Reject... | Total Simulati... Count (...) | Total Simulati... % (Cum.) |
|--------|---------------|-------|--------|-----|-----|----------------------------|---------------|---------------------|-----------------------------|------------------------------|----------------------------|-------------------------------|----------------------------|
| 1 | 5 | 4.877 | -4.877 | | | -0.293 | | 5 | 0 | 0 | 0 | 0 | 0 |
| 2 | 10 | 3.357 | -3.357 | | | 0.681 | | 10 | 0 | 0 | 0 | 0 | 0 |
| 3 | 15 | 2.68 | -2.68 | | | -0.361 | | 15 | 6 | 4 | 0 | 10 | 1 |
| 4 | 20 | 2.29 | -2.29 | | | -0.324 | | 20 | 13 | 11 | 0 | 24 | 2.4 |
| 5 | 25 | 2.031 | -2.031 | | | -0.6 | | 25 | 10 | 17 | 939 | 966 | 96.6 |
| Total | | | | | | | | 24.78 | | | | | |
| % | | | | | | | | | 2.9 | 3.2 | 93.9 | | |

In the actual trial we would have to know the value of σ^2 in order to compute Z_j . If the estimate $\sigma = 45$ is incorrect the power and type-1 error of the trial will not match the simulation results. Even worse, we have no way of knowing if the simulation results are correct or not, since it is difficult to verify the value of σ from a small data set. Thus it might be preferable to use an estimate of σ in the definition of the test statistic. This is discussed next.

2. Estimate σ^2 by s_j^2 from the interim data and compute

$$T_j = \frac{\bar{X}_{tj} - \bar{X}_{cj}}{\sqrt{\frac{4s_j^2}{n_j}}} . \tag{64.2}$$

It is more common to monitor a group sequential, normal endpoints trial with the test statistic T_j given by equation (64.2) than with the test statistic Z_j given (64.1). If we use (64.1) in the interim monitoring phase of a trial, we imply that we know the value of σ^2 with certainty, since it is needed in the computation. But the value of σ^2 that we use for this purpose may only be an informed guess with no data to back it up. At the interim monitoring stage, we have the opportunity to actually estimate σ^2 from the data and use the estimate, s_j^2 say, in the computation of the test statistic (64.2). This might be a more reliable approach than making a strong assumption that σ^2 is known with certainty. Now the distribution of T_j is only asymptotically normal. In small samples T_j has a student-t distribution under the null hypothesis. Thus use of T_j does not by itself ensure that the study will have the power and type-1 error that were implied by the sample size and stopping boundaries specified in the study design. This is where the simulations can help. Since the test statistic T_j is computed entirely from the data, and contains no unknown nuisance parameters, we can obtain the true power and type-1 error of any design that uses T_j for the interim monitoring, by means of simulation.

To do this open the simulation worksheet and select the following options:

Number of Looks: 5

Test Parameters | Response Generation | Simulation Controls

Trial Type: Superiority | Test Statistic: t

Test Type: 2-Sided | Variance: Equal

Sample Size (n): 25

| Look # | Info. Fraction | Cum. α Spent | | Efficacy Z | |
|--------|----------------|---------------------|-------|------------|--------|
| | | Upper | Lower | Upper | Lower |
| 1 | 0.200 | 0.000 | 0.000 | 4.877 | -4.877 |
| 2 | 0.400 | 0.000 | 0.000 | 3.357 | -3.357 |
| 3 | 0.600 | 0.004 | 0.004 | 2.680 | -2.680 |
| 4 | 0.800 | 0.012 | 0.012 | 2.290 | -2.290 |
| 5 | 1.000 | 0.025 | 0.025 | 2.031 | -2.031 |

Next, click on the **Response Generation Info** tab. You will notice that we do not have the option to select how the data are generated. We must use **Individual Means**. This is because when East calculates the t-statistic it needs to estimate the variance in each group. This is not possible if East generates the data using the **Difference of Means** option since East is only simulating differences of the means and not the actual means themselves. Thus, East cannot estimate the variance in each group. It is for this reason that the **Individual**

64 Simulation in East

Means option is selected.

Again we wish to simulated under the null hypothesis, $\mu_t - \mu_c = 0$. Enter the parameters in this tab as shown below.

Number of Looks: 5

| Test Parameters | Response Generation | Simulation Controls |
|---------------------------------------|---------------------|---|
| Generate Data Using: Individual Means | | <input checked="" type="checkbox"/> Common Standard Deviation |
| Mean Control (μ_c): 60 | | SD Control (σ_c): 45 |
| Mean Treatment (μ_t): 60 | | SD Treatment (σ_t): 45 |

Next, click the **Simulate** button. The simulation results will appear in the output preview window. Save this simulation to the workbook and then double click on Sim2 in the Library. A portion of the results are shown below.

Simulation Boundaries and Cumulative Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|---------|-----------------|------------|--------|----------------|----------------|-------------------|----------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 5 | 4.877 | -4.877 | 96 | 76 | 172 | 1.720% |
| 2 | 10 | 3.357 | -3.357 | 129 | 126 | 255 | 2.550% |
| 3 | 15 | 2.68 | -2.68 | 194 | 207 | 401 | 4.010% |
| 4 | 20 | 2.29 | -2.29 | 320 | 313 | 633 | 6.330% |
| 5 | 25 | 2.031 | -2.031 | 451 | 448 | 10000 | 100.000% |
| Total % | | | | 4.510% | 4.480% | | |

We observe that this small study does not preserve the type-1 error.

64.2 Binomial Studies

When computing a design for binomially distributed responses, East relies on the normal approximation to the binomial distribution. Thus, these designs may not be as accurate for small sample sizes. As in the previous section, the study should be re-designed and simulated repeatedly until, by trial and error, it possesses the required type-1 error and power. The simulations in East, as opposed to the designs, generate data from the actual binomial model specified instead of relying on a normal approximation. Thus, the simulations might provide a more realistic assessment of power and type-1 error for designs involving binomial endpoints and small sample sizes.

To illustrate, consider the following binomial design. Click **Discrete: Two Samples** on the **Design** tab and then click **Parallel Design: Difference of Proportions** and enter

the design parameters as shown below.

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

Design Type: Superiority Number of Looks: 5

Test Parameters Boundary

Test Type: 2-Sided

Type I Error (α): 0.05

Power: 0.9

Sample Size (n): Computed

Allocation Ratio: 1 (n_1/n_2)

Specify Proportion Response

Prop. under Control (π_c): 0.1

Specify Alternative Hypothesis

Prop. under Treatment (π_t): 0.65

Diff. in Prop. ($\delta_1 = \pi_t - \pi_c$): 0.55

Specify Variance

Pooled Estimate

Unpooled Estimate

Use Casagrande-Pike-Smith Correction (Ignored if alloc. ratio is not 1)

Next click on the **Boundary Info** tab and enter the parameters as shown below.

Design Type: Superiority Number of Looks: 5

Test Parameters Boundary

Efficacy

Boundary Family: Spending Functions

Spending Function: Gamma Family

Parameter (γ): 1

Type I Error (α): 0.05

Futility

Boundary Family: None

Spacing of Looks Equal Unequal

Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.200 | 0.014 | 2.449 | -2.449 |
| 2 | 0.400 | 0.026 | 2.419 | -2.419 |
| 3 | 0.600 | 0.036 | 2.398 | -2.398 |
| 4 | 0.800 | 0.044 | 2.391 | -2.391 |
| 5 | 1.000 | 0.050 | 2.395 | -2.395 |

64 Simulation in East

Click **Compute** to create the design. The output summary is shown below.

| Des 1 | |
|--|-----------------|
| Mnemonic | PN-2S-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 5 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.908 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.1 |
| Proportion under Treatment (π_t) | 0.65 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0.55 |
| Variance | Pooled Estimate |
| Boundary Parameters | |
| Spacing of Looks | Unequal |
| Efficacy Boundary | Gm (1) |
| Sample Size | |
| Maximum | 36 |
| Expected Under H0 | 35.139 |
| Expected Under H1 | 21.832 |

The study is designed for up to 5 looks and a two-sided α of 0.05. At most 36 patients are needed in order to achieve 90% power. Save the design to the workbook and click the **S** icon. You will be taken to the following simulation worksheet.

Number of Looks:

Test Parameters

Trial Type:

Test Type:

Sample Size (n):

Simulation Controls

Specify Variance

Pooled Estimate

Unpooled Estimate

| Look # | Info. Fraction | Cum. α Spent | | Efficacy Z | |
|--------|----------------|---------------------|-------|------------|--------|
| | | Upper | Lower | Upper | Lower |
| 1 | 0.194 | 0.007 | 0.007 | 2.458 | -2.458 |
| 2 | 0.389 | 0.013 | 0.013 | 2.427 | -2.427 |
| 3 | 0.611 | 0.018 | 0.018 | 2.382 | -2.382 |
| 4 | 0.806 | 0.022 | 0.022 | 2.393 | -2.393 |
| 5 | 1.000 | 0.025 | 0.025 | 2.398 | -2.398 |

Next, click **Simulate**. A portion of the results is shown below.

Simulation Boundaries and Cumulative Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|---------|-----------------|------------|--------|----------------|----------------|-------------------|----------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 7 | 2.458 | -2.458 | 0 | 0 | 0 | 0.000% |
| 2 | 14 | 2.427 | -2.427 | 118 | 0 | 118 | 11.800% |
| 3 | 22 | 2.382 | -2.382 | 589 | 0 | 589 | 58.900% |
| 4 | 29 | 2.393 | -2.393 | 795 | 0 | 795 | 79.500% |
| 5 | 36 | 2.398 | -2.398 | 898 | 0 | 1000 | 100.000% |
| Total % | | | | 89.800% | 0.000% | | |

Notice that the simulated power this design is (up to Monte Carlo accuracy) and slightly lower than 90%.

It is worth noting that by the time the sample size exceeds 100, the normal approximation should be sufficiently accurate. To see this, create a new design Des2 by editing Des1 and changing π_t value to 0.35. The Des2 summary will be as follows

| Des2 | |
|--|-----------------|
| Mnemonic | PN-25-DI |
| Test Parameters | |
| Design Type | Superiority |
| No. of Looks | 5 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.901 |
| Model Parameters | |
| Allocation Ratio (nt/nc) | 1 |
| Proportion under Control (π_c) | 0.1 |
| Proportion under Treatment (π_t) | 0.35 |
| Diff. in Prop. ($\pi_t - \pi_c$) | 0.25 |
| Variance | Pooled Estimate |
| Boundary Parameters | |
| Spacing of Looks | Unequal |
| Efficacy Boundary | Gm (1) |
| Sample Size | |
| Maximum | 137 |
| Expected Under H0 | 133.721 |
| Expected Under H1 | 79.609 |

Save this design to the workbook and open the simulation worksheet by clicking on the  icon in the Library. Again, click **Simulate**. A portion of the results is shown

64 Simulation in East

below.

Simulation Boundaries and Cumulative Boundary Crossing Probabilities

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|---------|-----------------|------------|--------|----------------|----------------|-------------------|----------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 27 | 2.453 | -2.453 | 85 | 0 | 85 | 8.500% |
| 2 | 55 | 2.415 | -2.415 | 402 | 0 | 402 | 40.200% |
| 3 | 82 | 2.4 | -2.4 | 631 | 0 | 631 | 63.100% |
| 4 | 110 | 2.389 | -2.389 | 808 | 0 | 808 | 80.800% |
| 5 | 137 | 2.396 | -2.396 | 898 | 0 | 1000 | 100.000% |
| Total % | | | | 89.800% | 0.000% | | |

This confirms that the power is indeed preserved (up to Monte Carlo accuracy) for group sequential designs based on the normal approximation to the binomial with large sample sizes. In general, whenever a small binomial study is contemplated it is a good idea to verify its operating characteristics through simulations.

64.3 Description of Simulation Output Columns

Following are the output quantities computed while simulating ‘Subject Data’.

| Column Name | Description | Applicability |
|---------------------|---|--|
| Scenario ID | Identification number of scenarios when multiple values are provided for a parameter(s) | All Simulations |
| Simulation ID | Identification number of simulations | All Simulations |
| Subject ID | Identification number of subjects | All Simulations |
| Arrival Time | Arrival Time of a particular subject | All Simulations |
| Treatment ID | Treatment given to a particular subject | All Simulations |
| Survival Time | Survival Time of a particular subject | SU-2S-LRAR, SU-2S-LRSD |
| DPN-2S-RAOut Time | Time when a particular subject dPN-2S-RAs out from the study | All Simulations |
| Stratum_Var_< i > | Variable which stratifies the subjects into different levels | SU-2S-LRAR, SU-2S-LRSD |
| CensorInd | Indicator variable (flag) denoting whether a particular subject is censored or not | Enhanced Simulations |
| Response | Response corresponding to a particular subject after given a particular treatment | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| Endpoint_< i > | Response of endpoint_< i > | All Simulations |
| Survival Time_Weeks | Survival Time of a particular subject in time unit weeks | Survival Designs |
| Site ID | Identification number of sites | MN-2S-DI,PN-2S-DI, PN-2S-RA,SU-2S-LRAR, SU-2S-LRSD |

64 Simulation in East

Following are the output quantities computed while simulating ‘Summary Statistics’.

| Column Name | Description | Applicability |
|---------------------|---|--|
| Scenario ID | Identification number of scenarios when multiple values are provided for a parameter(s) | All simulations |
| SimIndex | An identifier for the simulation | All simulations |
| Look Index | Identifier for the look number | Multi Look Simulations |
| Status | Variable denoting whether a simulation was successfully executed | All simulations |
| BdryStopCode | Lookwise stopping decision | Multi Look simulations (0=Continue, 1= lower efficacy stop, 2= Upper Efficacy stop, 3= Futility) |
| Accruals 0 | Total accrued subjects under control for a particular simulation | Enhanced simulations with accrual dPN-2S-RAout |
| DPN-2S-RAOuts0 | Total dPN-2S-RApedout subjects under control for a particular simulation | Enhanced simulations with accrual dPN-2S-RAout |
| Pendings0 | Number of pending subjects under control for a particular simulation | Enhanced simulations with accrual dPN-2S-RAout |
| Events0 | Total number of events happened for control | SU-2S-LRAR, SU-2S-LRSD |
| Accruals_< i > | Total accrued subjects under treatment i for a particular simulation | Enhanced simulations with accrual dPN-2S-RAout |
| DPN-2S-RAOuts_< i > | Total dPN-2S-RApedout subjects under treatment i for a particular simulation | Enhanced simulations with accrual dPN-2S-RAout |

| Column Name | Description | Applicability |
|-----------------------|---|--|
| Pendings_< <i>i</i> > | Number of pending subjects under treatment <i>i</i> for a particular simulation | Enhanced simulations with accrual dPN-2S-RAout |
| Events_< <i>i</i> > | Total number of events happened for treatment <i>i</i> | SU-2S-LRAR, SU-2S-LRSD |
| TotAccruals | Total accruals for all the treatments together in a particular simulation | Enhanced simulations with accrual dPN-2S-RAout |
| TotDPN-2S-RAOuts | Total dPN-2S-RA outs for all the treatments together in a particular simulation | Enhanced simulations with accrual dPN-2S-RAout |
| Tot Pendings | Total pending subjects for all treatments together | Enhanced simulations with accrual dPN-2S-RAout |
| Tot Events | Total events for all the treatments together in a particular simulation | SU-2S-LRAR, SU-2S-LRSD |
| Look Time | At what time a particular look was taken | Enhanced simulations with accrual dPN-2S-RAout |
| Avg FollowUp Time | Average followup time for a particular simulation | Enhanced simulations with accrual dPN-2S-RAout |
| LogRankScore | Numerator of log rank statistic | SU-2S-LRAR, SU-2S-LRSD |
| HRfMN-2S-RALRStat | HR estimated fMN-2S-RA Log Rank statistic | SU-2S-LRAR, SU-2S-LRSD |

64 Simulation in East

| Column Name | Description | Applicability |
|----------------------------------|---|--|
| StdError | Standard Error | All simulations |
| LRStat | Log rank statistic | SU-2S-LRAR, SU-2S-LRSD |
| LwrEffBdry | Lower efficacy boundary | All 2 sided simulations |
| UprEffBdry | Upper efficacy boundary | All 2 sided simulations |
| LwrFutBdry | Lower futility boundary | All 2 sided simulations |
| UprFutBdry | Upper futility boundary | All 2 sided simulations |
| AccrDurtn | Accrual duration | Enhanced simulations with accrual dropout |
| HazardRate0Strat_< i - 1 > | Hazard rate for control in stratum_< i > | SU-2S-LRAR, SU-2S-LRSD |
| HazardRate_< j > Strat_< i - 1 > | Hazard rate for treatment_< j > in stratum_< i > | SU-2S-LRAR, SU-2S-LRSD |
| Hazard Ratio Strat_< i - 1 > | Hazard ratio in stratum_< i > | SU-2S-LRAR, SU-2S-LRSD |
| Log Rank Score Strat_< i - 1 > | Numerator of log rank statistic corresponding to stratum_< i > | SU-2S-LRAR, SU-2S-LRSD SU-2S-LRAR, SU-2S-LSRD |
| Std Error Strat_< i - 1 > | Standard error of log rank score corresponding to stratum_< i > | SU-2S-LRAR, SU-2S-LRSD |
| Completers0 | Number of completers under control | All simulations |
| Completers_< i > | Number of completers under treatment_< i > | All simulations |

| Column Name | Description | Applicability |
|-----------------------|---|--|
| Tot Completers | Total number of completers across all treatments | All simulations |
| Sum 0 | Sum of responses for control | All continuous endpoint simulations |
| PPN-2S-RA0 | PPN-2S-RAortion of responses for control | All discrete simulations |
| Sum_< i > | Sum of responses for treatment_< i > | All continuous endpoint simulations |
| PPN-2S-RA_< i > | PPN-2S-RAortion of responses for treatment_< i > | |
| PPN-2S-RApId | Pooled pPN-2S-RAortion pooled across treatments | All discrete simulations |
| Rho | Ratio of pPN-2S-RAortions estimated fMN-2S-RA the data | All discrete simulations |
| HFactor | Standard error of ratio between pooled and unpooled standard errors | All discrete simulations |
| Info | Fisher's information corresponding to a particular look | All simulations |
| Adaptation | Whether adaptation happened for a particular simulation | All simulations with SSR option available |
| Zone | Sample size reestimation zones | All simulations with SSR options available |
| Delta Sign | Sign of delta | All simulations with SSR options available |
| Adapt ReEstCompleters | Reestimated number of completers after adaptation | All simulations with SSR options available |

64 Simulation in East

| Column Name | Description | Applicability |
|----------------------|--|--|
| WaldStatIncr | Incremental wald test statistic | All simulations with SSR option available |
| TestStat | Test statistic | All simulations with SSR option available |
| InterimCP | Conditional power at the adapt look before adaptation | All simulations with SSR option available |
| AttainedCP | Conditional power attained at the adapt look after adaptation using reestimated events | All simulations with SSR options available |
| AccrDurnLTAdptLkTime | Indicates whether accrual duration is less than adapt look time | CHW/ CDL simulations |
| Mean0 | Mean of control responses | All Continuous endpoint simulations |
| SumOfSquares0 | Sum of square of control responses | All Continuous endpoint simulations |
| StdDev0 | Standard deviation of control responses | All Continuous endpoint simulations |
| Mean_< i > | Mean of treatment_< i > responses | All Continuous endpoint simulations |
| SumOfSquares_< i > | Sum of square of treatment_< i > responses | All Continuous endpoint simulations |
| StdDev_< i > | Standard deviation of treatment_< i > responses | All Continuous endpoint simulations |
| StdDevPld | Pooled standard deviation pooled across treatments | All Continuous endpoint simulations |
| Delta | Difference of treatment and control mean response | All Continuous endpoint simulations |

| Column Name | Description | Applicability |
|-------------------------|---|--|
| Tstat | Calculated t statistic | MN-1S-SM, MN-2S-DI, MN-2S-RA, MN-MAMS-PC, PN-MAMS-PC |
| Zstat | Calculated Z statistic | MN-1S-SM, MN-2S-DI, MN-2S-RA, MN-MAMS-PC, PN-MAMS-PC |
| CPnull | Conditional type I error | MS simulations |
| AdaptReEstCompleters | Reestimated completers after adaptation adjusted for upper and lower limits | MS simulations |
| AdaptActReEstCompleters | Actual unadjusted reestimated completers after adaptation | MS simulations |
| MSActReEstCompleters | Actual look position in the post adapt looks | MS simulations |
| EstDeltaII | Estimated delta at second stage | MS simulations |
| SEII | Standard error of delta at second stage | MS simulations |
| TStatII | t statistic at second stage | MS simulations |
| LwrEffBdryII | Lower efficacy boundary for second stage | MS simulations |
| UprEffBdryII | Upper efficacy boundary for second stage | MS simulations |
| LwrFutBdryII | Lower futility boundary for second stage | MS simulations |
| UprFutBdryII | Upper futility boundary for second stage | MS simulations |
| RCILowerBound | Repeated confidence interval lower bound | MS simulations |
| RCIUpperBound | Repeated confidence interval upper bound | MS simulations |

64 Simulation in East

| Column Name | Description | Applicability |
|---------------------|---|------------------------|
| BWCILowerBound | Backward image confidence interval lower bound | MS simulations |
| BWCIUpperBound | Backward image confidence interval upper bound | MS simulations |
| BWCIMUE | BWCI median unbiased estimate | MS simulations |
| LwrStgIDsgnIndx | Stage I design index at which the stage II design power is less than the stage I conditional power for lower BWCI estimates | MS simulations |
| LwrStgITestStat | Test statistic value which gives the desired stage II design power for lower BWCI estimates | MS simulations |
| UprStgIDsgnIndx | Stage I design index at which the stage II design power is less than the stage I conditional power for upper BWCI estimates | MS simulations |
| UprStgITestStat | Test statistic value which gives the desired stage II design power for upper BWCI estimates | MS simulations |
| RawPValue_< i > | raw p value corresponding to treatment_< i > | MN-MAMS-PC, PN-MAMS-PC |
| RejectionFlag_< i > | Flag indicating whether null hypothesis corresponding to treatment_< i > is rejected | MN-MAMS-PC, PN-MAMS-PC |
| StopStatus | Status of a treatment after a look | MN-MAMS-PC, PN-MAMS-PC |
| Return_Code | Indicator variable denoting whether a simulation ran successfully | MN-2S-ME, PN-2S-ME |

| Column Name | Description | Applicability |
|--------------------------------|--|--------------------|
| PPN-2S-RA_0_Endpoint_< i > | Observed response rate corresponding to end point_< i > for control | PN-2S-ME |
| PPN-2S-RA_< j >.Endpoint_< i > | Observed response rate corresponding to end Point_< i > for treatment_< j > | PN-2S-ME |
| Delta_< j >.Endpoint1 | Difference of treatment_< j > and control response corresponding to endpoint_< i > | MN-2S-ME, PN-2S-ME |
| StdError_< j >.Endpoint_< i > | Standard error of delta_< j > | MN-2S-ME, PN-2S-ME |
| Test_Stat_< j >.Endpoint_< i > | Test statistic_< j > corresponding to endpoint_< i > | MN-2S-ME, PN-2S-ME |
| Pval_< j >.Endpoint1 | pvalue_< j > corresponding to endpoint_< i > | MN-2S-ME, PN-2S-ME |
| Maxpval_Fam_< i > | Maximum pvalue among family_< i > of endpoints | MN-2S-ME, PN-2S-ME |
| Adjpval_Endpoint_< i > | Adjusted pvalue corresponding to endpoint_< i > | MN-2S-ME, PN-2S-ME |
| Adjpval | Adjusted pvalue for last family | MN-2S-ME, PN-2S-ME |
| SampleSize_0 | Sample size corresponding to control | MN-2S-ME, PN-2S-ME |
| SampleSize_< i > | Sample size corresponding to treatment_< i > | MN-2S-ME, PN-2S-ME |
| Tot_SampleSize | Total sample Size | MN-2S-ME, PN-2S-ME |

64 Simulation in East

| Column Name | Description | Applicability |
|----------------------------|---|-----------------------|
| RejFlag_1_Endpoint_< i > | Rejection flag indicating whether the hypothesis corresponding to endpoint_< i > is rejected | MN-2S-ME, PN-2S-ME |
| IsNonNull_1_Endpoint_< i > | Indicator variable denoting whether endpoint_< i > is generated under null | MN-2S-ME, PN-2S-ME |
| FWERFlag_Fam_< i > | Indicator variable denoting whether a particular simulation contributes to FWER count for family_< i > | MN-2S-ME, PN-2S-ME |
| FWERFlag | Indicator variable denoting whether a particular simulation contributes to overall FWER count | MN-2S-ME, PN-2S-ME |
| ConPowFlag_Fam_< i > | Indicator variable denoting whether a particular simulation contributes to Conjunctive Power count for family_< i > | MN-2S-ME, PN-2S-ME |
| DisjnPowFlag_Fam_< i > | Indicator variable denoting whether a particular simulation contributes to Disjunctive Power count for family_< i > | MN-2S-ME, PN-2S-ME |
| DisjnPowFlag | Indicator variable denoting whether a particular simulation contributes to overall Disjunctive Power count | MN-2S-ME, PN-2S-ME |
| ConPowFlag | Indicator variable denoting whether a particular simulation contributes to overall Conjunctive Power count | MN-2S-ME, PN-2S-ME |
| Stage | Variable indicating whether we are in interim or final stage | Predict |

| Column Name | Description | Applicability |
|-------------------|---|--------------------|
| FABdryStopCode | Stopping decision at final analysis (when no response is pending) | PN-2S-DI, MN-2S-DI |
| FAAccruals0 | Accruals for control at final analysis | PN-2S-DI, MN-2S-DI |
| FACompleters0 | Completers for control at final analysis | PN-2S-DI, MN-2S-DI |
| FAAccruals1 | Accruals for control at final analysis | PN-2S-DI, MN-2S-DI |
| FAPendings1 | Pendings for control at final analysis | PN-2S-DI, MN-2S-DI |
| FACompleters1 | Completers for control at final analysis | PN-2S-DI, MN-2S-DI |
| FATotAccruals | Total accruals for control at final analysis | PN-2S-DI, MN-2S-DI |
| FATotPendings | Total pendings for control at final analysis | PN-2S-DI, MN-2S-DI |
| FATotCompleters | Total completers for control at final analysis | PN-2S-DI, MN-2S-DI |
| FALookTime | Look time at final analysis | PN-2S-DI, MN-2S-DI |
| FAAvgFollowUpTime | Average followup time at final analysis | PN-2S-DI, MN-2S-DI |
| FASum0 | Sum for control at final analysis | PN-2S-DI, MN-2S-DI |
| FAPPN-2S-RA0 | PPN-2S-RA for control at final analysis | PN-2S-DI, MN-2S-DI |
| FASum1 | Sum for treatment at final analysis | PN-2S-DI, MN-2S-DI |
| FAPPN-2S-RA1 | PPN-2S-RA for treatment at final analysis | PN-2S-DI, MN-2S-DI |

64 Simulation in East

| Column Name | Description | Applicability |
|----------------|---|-----------------------|
| FAPPN-2S-RAPId | Polled pPN-2S-RAortionat final analysis | PN-2S-DI |
| FADelta | Delta at final analysis | PN-2S-DI, MN-2S-DI |
| FAHFactor | HFactor at final analysis | PN-2S-DI, MN-2S-DI |
| FAStdError | Standard error at final analysis | PN-2S-DI, MN-2S-DI |
| FAInfo | Information at final analysis | PN-2S-DI, MN-2S-DI |
| FAWaldTestStat | Wald test statistic at final analysis | PN-2S-DI |
| FATStat | T test statistic at final analysis | MN-2S-DI |
| FAZStat | Z test statistic at final analysis | MN-2S-DI |

Following are the output quantities computed while simulating ‘Sitewise Summary Statistics’.

| Column Name | Description | Applicability |
|--------------------|--|--|
| SiteID | Identification number of sites | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| AvgInitiationTime | Identification number of sites | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| AvgLastSubjArrTime | Sitewise average last subject arrival time averaged over simulations | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| AvgNumOfSubj | Sitewise average number of subjects averaged over simulations | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| AvgAccrualDuration | Sitewise average accrual duration averaged over simulations | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| AvgAccrualRate | Sitewise average rate of accrual averaged over simulations | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| SiteOpenedSimCount | In how many simulations a particular site is opened | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |

64 Simulation in East

Following are the output quantities computed while simulating 'Site Parameters'.

| Column Name | Description | Applicability |
|--------------------|--|--|
| SimulationID | Identification number of simulations | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| SiteOpenFlag | Flag indicating whether a particular site is opened in a particular simulation | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| SiteAlreadyOpened | Flag indicating whether a particular site is already opened at the time of prediction in a particular simulation | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| SiteID | Identification number of sites | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| SiteInitiationTime | Time when a particular site is initiated in a particular simulation | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| SiteAccrRate | Accrual Rate corresponding to each site in a particular simulation | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| SubjectsAccrued | How many subjects are accrued at a particular site in a particular simulation | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| LastSubjectRand | Time when the last subject was randomized for a particular site | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| AccrualDuration | Duration of accrual corresponding to a particular site in a particular simulation | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |
| ObsrvdAccrualRate | Observed accrual rate corresponding to a site in a particular simulation | MN-2S-DI, PN-2S-DI, PN-2S-RA, SU-2S-LRAR, SU-2S-LRSD |

65

Predictive Interval Plots

65.1 Predicting the Future Course of a Trial with Predictive Interval Plots (PIPS)

At the design stage of clinical trial, when no data are available, one relies on initial assumptions about the efficacy of the treatment arms to perform power calculations. Once the trial is underway, however, data begin to accrue and can be utilized to make predictions about the future course of the trial. These predictions fall into two categories; predictions from data pooled by treatment arm and predictions from unpooled data. For the trial sponsor, who must remain blinded to the results while the trial is on-going one, predictions from pooled data are the only option. A data monitoring committee on the other hand does have access to data broken out by treatment arm and is thus in a position to make predictions about the future course of the trial in an unblinded manner. In this chapter, we focus only on predictions from unblinded data. A popular way to make such predictions is through the use of conditional power. We have provided numerous examples of conditional power throughout this manual and hence will not dwell on it here. In this chapter we present an alternative graphical approach to prediction, utilizing predictive interval plots (PIPS) proposed by Evans, Li and Wei (2007) and Li, Evans, Uno and Wei (2009). These plots provide us with a visual display of the possible future outcomes for the trial by generating a series of repeated confidence intervals for future time points that are conditional on the current data. Conditional power is an automatic by-product of these plots, which provide additional insights about the magnitude of the treatment effect and its associated uncertainty. Please see Appendix L for details on input, output, and formulas relating to Predictive Interval Plots.

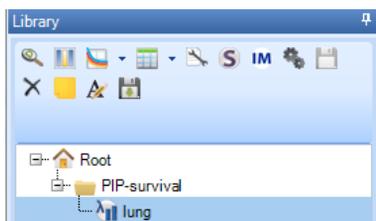
65.2 Example 1: PIP for Time to Event Data

A clinical trial of non small cell lung cancer was designed for 80% power to detect a hazard ratio of 0.8 at $\alpha = 0.05$ (two-sided) with three equally spaced looks using a Lan-DeMets O'Brien-Fleming type (**LD (OF)**) spending function. The primary endpoint was overall survival (OS). With these inputs, 641 OS events are needed to achieve 80% power. The median OS for the control arm was assumed to be 10 months. Based on 18 months of enrollment and an additional 12 months of follow-up this 30-month trial requires 639 events from a sample size of 897 patients.

The workbook **PIP-survival** containing this design named **lung** is already available to you in the sub-folder **Samples** in the East 6.4 installation folder in your computer. A typical path for this sub-folder is: **C:\Program Files (x86) \Cytel\Cytel Architect\East 6.4\Samples**. Open this workbook from File or Home menu.

65 Predictive Interval Plots

The Library nodes will appear as shown below.



Click on the design node **lung** and click on  icon to get the details of the design as shown below.

| Test Parameters | |
|--------------------------------|-------------|
| Design ID | lung |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.8 |
| Model Parameters | |
| HR = λ_1/λ_c | |
| Under H0 | 1 |
| Under H1 | 0.8 |
| Ratio of Medians: | 1.25 |
| Var (Log HR) | Null |
| Allocation Ratio (n_1/n_2) | 1 |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| Accrual/Dropout Parameters | |
| Accrual Duration | 18 |
| Max Study Duration | 30 |
| Dropout | No |

Variable Follow-Up Design: All subjects are followed until failure, drop out or end of study.

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|---------------------------------|-------------|---------------|---------|
| Sample Size (n) | | | |
| Maximum | 448 | 449 | 897 |
| Expected H1 | 446.332 | 446.332 | 892.664 |
| Expected H0 | 448.473 | 448.473 | 896.946 |
| Events (s) | | | |
| Maximum | 336 | 303 | 639 |
| Expected H1 | 298.833 | 263.963 | 546.044 |
| Expected H0 | 318.545 | 318.545 | 636.379 |
| Maximum Information (I): 159.75 | | | |

Accrual and Study Duration

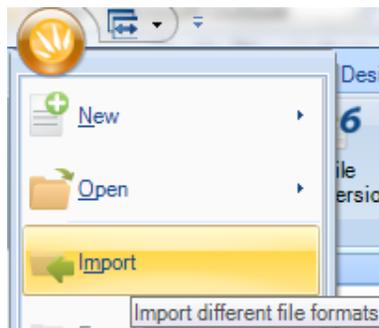
| | Accrual Duration | Study Duration |
|-------------|------------------|----------------|
| Maximum | 18 | 29.961 |
| Expected H1 | 17.913 | 25.757 |
| Expected H0 | 17.999 | 27.795 |

Stopping Boundaries: Look by Look

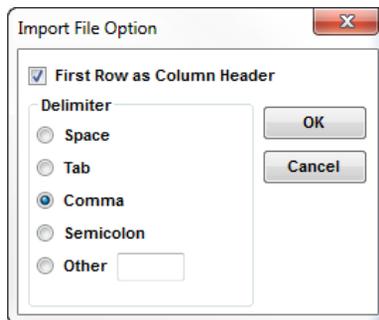
| Look # | Info. Fraction (s/s_max) | Events (s) | Cumulative α Spent | Boundaries Efficacy Z | |
|--------|--------------------------|------------|---------------------------|-----------------------|--------|
| | | | | Upper | Lower |
| 1 | 0.333 | 213 | 2.07E-4 | 3.71 | -3.71 |
| 2 | 0.667 | 426 | 0.012 | 2.511 | -2.511 |
| 3 | 1 | 639 | 0.05 | 1.993 | -1.993 |

The First Interim Analysis Although the first interim analysis was planned after 213 events, due to rapid enrollment it occurred earlier, after only 119 events. The dataset containing the data from the first interim analysis is saved in a .csv file named **PIP-Lung-Look01.csv** in the **Samples** folder. To illustrate the role of the PIPs at this interim analysis, you need this dataset. While you are on the design node **lung**, you can bring up this dataset into the workbook by clicking on the menu item File >

Import or Home > Import and locating the sub-folder Samples.



After clicking on the menu item **Import** and locating **Samples** sub-folder, you click on the dataset name **PIP-Lung-Look01.csv**. You will be presented with the following dialog box.



Keep the default choices selected, click OK, and keep the imported dataset in the workbook **PIP-survival**. Now a new node with the name **PIP-Lung-Look01.cydx** will appear under the design node **lung**. The data also

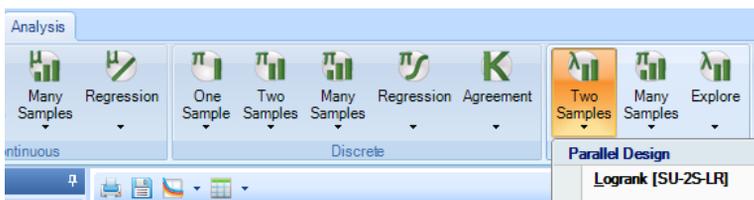
65 Predictive Interval Plots

will be displayed in the right side window.

| TrtmntID: 1 | | Value: 2 | | | | |
|-------------|----------|------------|-------------|---------|---------|-----|
| | TrtmntID | SRVMON | ArrivalTime | Censor1 | Censor2 | var |
| 1 | 2 | 1.4 | 0.033333333 | 1 | 1 | |
| 2 | 2 | 1.4 | 0.066666667 | 1 | 1 | |
| 3 | 2 | 22.6333333 | 0.066666667 | 0 | 0 | |
| 4 | 2 | 3.3 | 0.2 | 1 | 1 | |
| 5 | 2 | 5.3 | 0.533333333 | -1 | 0 | |
| 6 | 1 | 4 | 0.866666667 | 1 | 1 | |
| 7 | 1 | 2.8 | 1.166666667 | 1 | 1 | |
| 8 | 2 | 2.8 | 1.6 | 1 | 1 | |
| 9 | 2 | 3.4 | 1.933333333 | 1 | 1 | |
| 10 | 1 | 20.6666667 | 2.033333333 | 0 | 0 | |

The dataset is saved in the library as a Cytel file with extension **.cydx**. Examine this dataset. It contains five variables: TrtmntID (1=control, 2=experimental); SRVMON (time since entering the trial in months); ArrivalTime (time of entry into the trial), Censor1 (1=alive; 0=dead, -1=lost to follow up); Censor2 (1=alive, 0=dead or lost to follow up). Note the presence of two censor variables. Censor1, indicating drop-outs with -1, is utilized by the program that generates the PIPs. Censor2, indicating either drop-outs or administratively censored patients, is utilized by the Analysis program computing the Logrank test. This can be seen in the choice of the variables in the Analysis dialog box and PIP dialog box detailed below.

Before we can perform the first interim analysis, we must estimate the hazard ratio and its standard error from this interim analysis dataset. To that end, select the **Two Samples > Logrank** from the **Analysis** tab.



Enter the appropriate variables into the input dialog box.

Analysis: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank

Data Set: PIP-Lung-Look01.cyx

Main Advanced

Trial Type: Superiority Response Variable: SRVMON Frequency Variable:

Population ID: TrtmntID

Control: 1 Censor Indicator: Censor2

Treatment: 2 Censored: 0 Complete: 1

and click on the **OK** button at the bottom right side of the screen. You will get the analysis results with a summary as shown below.

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | | Average Follow-up Time |
|--------------|-----------------|------------|----------------|------------|----------------|------------------------|
| | | Count | % | Count | % | |
| 1 | 256 | 61 | 23.828% | 195 | 76.172% | 11.353 |
| 2 | 261 | 58 | 22.222% | 203 | 77.778% | 11.699 |
| Total | 517 | 119 | 23.017% | 398 | 76.983% | 11.528 |

Test of Hypothesis:

| Log Rank Score | Std. Error | Standardized Test Statistic | (1-Sided) | | (2-Sided) |
|----------------|------------|-----------------------------|-----------|---------|-----------|
| | | | Tail | p-value | p-value |
| -2.509 | 5.441 | -0.461 | L.E. | 0.322 | 0.645 |

Parameter Estimates from Cox Model:

| Hazard Ratio (HR) | ln(Hazard Ratio) (δ) | Std. Error of Estimate of δ | Wald Statistic | 95% Confidence Interval(2-Sided) | |
|-------------------|-------------------------------|------------------------------------|----------------|----------------------------------|-------------|
| | | | | Lower Limit | Upper Limit |
| 0.919 | -0.084 | 0.183 | -0.46 | 0.642 | 1.317 |

Estimated Hazard Rates:

| | |
|--------------------------------|---------|
| Control (λ_c) | 0.02099 |
| Treatment ($\lambda_c * HR$) | 0.01929 |

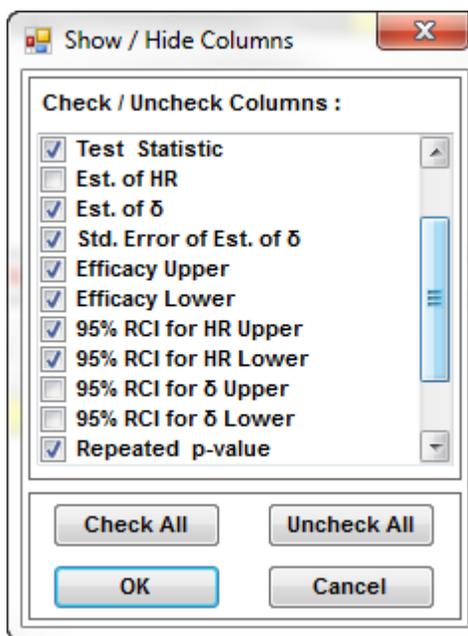
The hazard ratio is 0.919 and the total number of events is 119. These are the summary statistics we need to perform the first interim analysis. With the design node **lung** selected in the library, bring up the Interim Monitoring worksheet by clicking on the

65 Predictive Interval Plots

IM icon on the library toolbar.

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | | 95% RCI for HR | | Repeat... p-value | CP | Predictive Power |
|--------|----------------------|-------------------|----------------|------------------|--------------------------------|----------|-------|----------------|-------|-------------------|----|------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |

East gives you a facility to choose the columns to display in the IM sheet, by clicking on the show/hide icon  and choosing from the list displayed.



For convenient entry of the summary data into the interim monitoring worksheet, you can display the Interim Monitoring worksheet and the logrank analysis side by side in two windows, with the use of the menu item Home>Arrange>View Selected

Windows:

The screenshot displays the 'IM Interim Monitoring Dashboard' window. At the top, there is a table for 'Look #', 'Information Fraction', 'Cumulative Events', 'Test Statistic', 'Est. of δ ', 'Std. Error of Est. of δ ', and 'Efficacy' (Upper, Lower, U). Below this, a red instruction reads: 'Click the "Enter Interim Data" button to enter the Look # 1 data.' The dashboard includes several interactive panels: 'Stopping' (Cum. Events, Efficacy Upper/Lower), 'Conditions' (HR, CP), 'Error' (Info. Fraction, α), and 'Conf.' (Info. Fraction, RCI Upper/Lower). To the right, the 'Analysis: Time to Event Response: Logrank Test' window is open, showing the following text:

Let $\delta = \ln(A_1/A_2)$
 $H_0: \delta = 0$ Vs. $H_1: \delta \neq 0$ for 2-Sided test
 Either $H_1: \delta > 0$ Or $H_1: \delta < 0$ for 1-Sided test

Data File: PIP-Lung-Look01.cyx
 Trial Type: Superiority
 Population ID: TrtmntID(Treatment=2, Control=1)
 Response Variable: SRVMON
 Censor: Censor2(Censor Value=0, Complete=1)
 Confidence Level: 0.95

Output

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | |
|--------------|-----------------|--------|---------|----------|---------|
| | | Count | % | Count | % |
| 1 | 256 | 61 | 23.828% | 195 | 76.172% |
| 2 | 261 | 58 | 22.222% | 203 | 77.778% |
| Total | 517 | 119 | 23.017% | 398 | 76.983% |

Parameter Estimates:

| Hazard Ratio (HR) | 95% Confidence Interval(2-Sided) | | |
|-------------------|----------------------------------|-------------|--|
| | Lower Limit | Upper Limit | |
| 0.919 | 0.641 | 1.317 | |

Click on the **Enter Interim Data** button and the **Test Statistic Calculator** will appear. Here you have two options, either you can read and transfer directly from the results of analysis node or enter the estimate and SE of delta manually. Let us follow the first option. Click on the Recalc button and it will transfer the results from

65 Predictive Interval Plots

the analysis node to the test statistic calculator.

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Read from Analysis Node

Select Workbook: PIP-survival

Select Analysis Node: PIP-Lung-Look01.cydx:Analysis:

Cumulative Events: 119

Input for Survival end point

Estimate of δ : -0.084

$\delta = \ln(\lambda_t / \lambda_c)$

Standard Error of Estimate of δ : 0.183

Output

Test Statistic: -0.46

Recalc OK Cancel

If you had chosen the second option in the test statistic calculator, you would enter 119 for the cumulative number of events, $\ln(0.919)$ for $\hat{\delta}$ and $2/\sqrt{119}$ for the standard error of $\hat{\delta}$.

Click **OK** to enter the first-look data into the interim monitoring worksheet.

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | | 95% RCI for HR | | Repeat... p-value | CP | Predicti... Power |
|--------|----------------------|-------------------|----------------|------------------|--------------------------------|----------|--------|----------------|-------|-------------------|-------|-------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.186 | 119 | -0.46 | -0.084 | 0.183 | 5.063 | -5.063 | 2.326 | 0.363 | 1 | 0.156 | 0.403 |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |

The results at this first interim analysis are not very promising. The conditional power under the current trend (HR=0.919) is only 0.156 and the predictive power is only 0.403. The predictive interval plots (PIPs) can provide some additional insights by simulating the future course of the trial conditional on the data already obtained and assumptions about the hazard rates of the two survival curves. To generate these plots select the **Look # 1** row. Then click on the **PIP** icon to open the PIP dialog box. Enter the inputs into the left panel of the dialog box as shown below.

PIP for Look: Required No. of Events:

Specify Subject Info

Select Workbook:

Select Subject Data:

Choose Variables

Population ID: Status Indicator:

Control: 1=Complete

Treatment: 0=Censored

Arrival Time: Response Variable:

-1 = Dropout

The information in the PIP-Lung-Look01.cyx is now available to East.

Entries into the right hand panel of the PIP inputs dialog box may either be user specified or estimated directly from the PIP-Lung-Look01.cyx dataset. To begin with, let us estimate the entries from the data. Accordingly click on the **Optional: Estimate Parameters from Data** button. The right hand panel fills up as

65 Predictive Interval Plots

shown below.

Parameters of Current Dataset (Editable)

| | | |
|---|---------|-----------|
| Sample Size: | 517 | |
| No. of Events: | 119 | |
| Accrual Rate: | 44.957 | |
| Hazard Ratio (λ_t/λ_c): | 0.919 | |
| | Control | Treatment |
| Hazard Rate (λ): | 0.021 | 0.019 |
| Dropout Hazard Rate (γ): | 0.002 | 0.001 |

We are now in a position to generate the predictive interval plots. As stated earlier these are repeated confidence intervals based on the data already observed and estimates of the hazard ratio for future looks. Since the first interim look was taken earlier than planned, there are still three additional interim looks (looks, 2, 3, or 4) to be encountered.

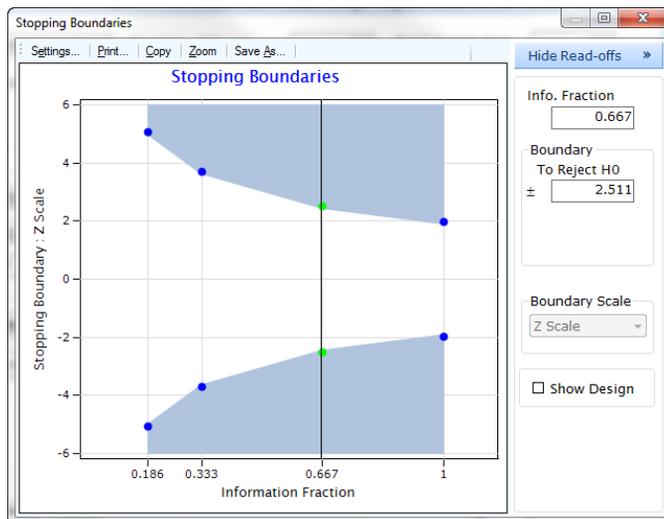
Input for Predicted Intervals Plot

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

PIP for Look: Any Future Look Total No. of Events: 639 Total Sample Size: 897

The boundaries for these future looks have been re-computed based on the specified error spending function. To view the re-computed 4-look design, click on the 

icon at the top right of the input dialog box.



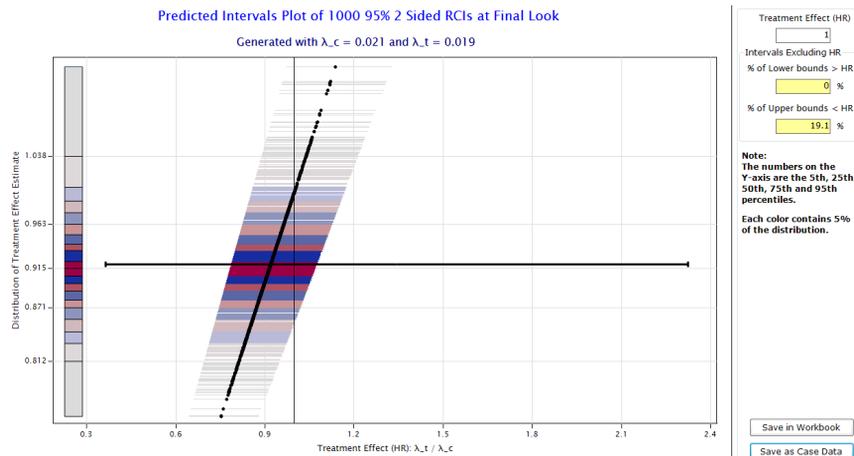
Suppose we wish to generate 1000 PIPs for look 4, ignoring the intermediate looks. Select Final Look from the drop-down box .

The dialog box is titled "Input for Predicted Intervals Plot" and contains the following information:

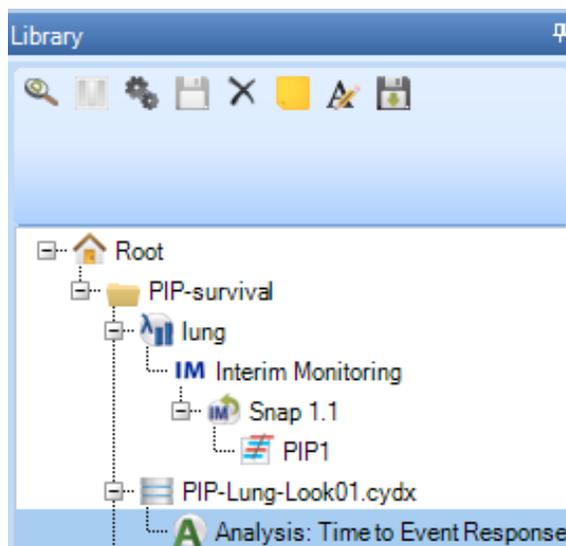
- Simulation: Survival Endpoint - Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration
- PIP for Look: Final Look (selected in a dropdown)
- Required No. of Events: 639
- Total Sample Size: 897

65 Predictive Interval Plots

and press the **Simulate** button. The following plot is generated.



One thousand repeated confidence intervals (RCIs) are generated for look 4, sorted in increasing order of their corresponding estimated hazard ratios, and stacked on top of each other. Save this PIP in the library by clicking on the **Save in Workbook** button on the bottom right of the plot. The library should now look as shown below.



Let us examine the generated PIP. The black horizontal line is the RCI for the current look (look 1). Notice how much narrower the RCIs for look 4 are compared to the current RCI. By default, the vertical cursor is positioned at HR=1 on the X-axis. In this position it is seen that 19.1% of the RCI's have upper bounds that are less than 1, suggesting that under the current trend with HR=0.919, the probability of a successful outcome for this trial at look 4 (ignoring all intermediate looks) is 0.191.

One can drag the vertical cursor to the right or left to see what percentage of trials will successfully cut-off hazard ratios other than 1. For the present let us leave the vertical cursor at HR=1. Notice the thick vertical bar with colored bands near the Y-axis. This band displays quantiles of the distribution of the hazard ratios generated by the simulations. Each color on either side of the median contains 5% of the generated hazard ratios. Thus, for example, the lowest five bands on the bar, ending at HR=0.871 represent 25% of the generated hazard ratios. In other words, the lower 25-th quantile of the hazard ratios is 0.871.

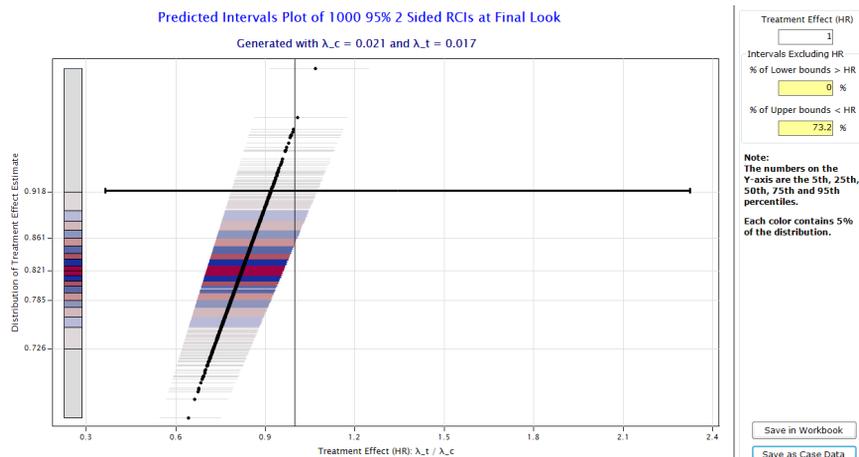
Since only 19.1% of the RCIs in this PIP resulted in a statistically significant outcome (upper bound of RCI less than 1), one might weigh the option of terminating the trial for futility. The above PIP was, however, generated under the assumption that the hazard ratio is 0.919, estimated from the look 1 data, is the actual hazard ratio. There is uncertainty associated with this estimate. Thus it would be desirable to take a conservative approach to futility termination and re-run the PIPs under the assumption, made at the design stage, that the underlying HR=0.8. To that end, we retrieve the input dialog box that was used for the current PIP by clicking on **PIP1** in the library and clicking on the **Edit** tool in the library toolbar. While on the node **PIP1**, click on . In the ensuing dialog box, change the value of the hazard rate for the Treatment arm from the current $\lambda(Treatment)$ value to $0.8 \times \lambda(Control)$.

| | Control | Treatment |
|-----------------------------------|---------|---------------|
| Hazard Rate (λ): | 0.021 | $0.8 * 0.021$ |
| Dropout Hazard Rate (γ): | 0.002 | 0.001 |

Now generate a new PIP by clicking on the **Simulate** button and save it in the

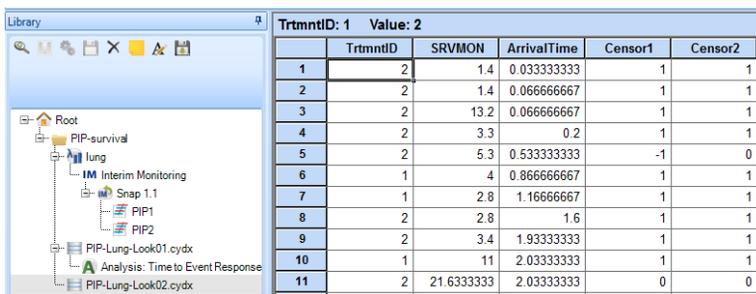
65 Predictive Interval Plots

workbook.



In this PIP, 73.2% of the RCIs have upper bounds that exclude HR=1. Therefore, given the uncertainty about the true value of the HR, it is premature to terminate this trial for futility and the trial continues to the next interim analysis.

The Second Interim Analysis The dataset for the second interim analysis is contained in a .csv file named **PIP-Lung-02.csv** on your computer. Import this .csv into East as shown below.



Next perform the logrank test on the look 2 data by invoking it from the Analysis tab in the same manner as you did for look 1.

The results will appear as shown below.

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | |
|--------------|-----------------|--------|---------|----------|---------|
| | | Count | % | Count | % |
| 1 | 395 | 126 | 31.899% | 269 | 68.101% |
| 2 | 403 | 132 | 32.754% | 271 | 67.246% |
| Total | 798 | 258 | 32.331% | 540 | 67.669% |

Parameter Estimates:

| Hazard Ratio (HR) | 95% Confidence Interval(2-Sided) | |
|-------------------|----------------------------------|-------------|
| | Lower Limit | Upper Limit |
| 1.019 | 0.799 | 1.301 |

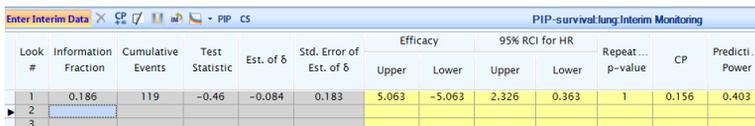
Test of Hypothesis:

| Log Rank Score | Std. Error | Standardized Test Statistic | (1-Sided) | | (2-Sided) |
|----------------|------------|-----------------------------|-----------|---------|-----------|
| | | | Tail | p-value | p-value |
| 1.231 | 8.008 | 0.154 | G.E. | 0.439 | 0.878 |

Estimated Hazard Rates:

| | |
|--------------------------------|-------|
| Control (λ_c) | 0.033 |
| Treatment ($\lambda_c * HR$) | 0.034 |

At this look, taken after 258 events, the hazard ratio estimate is 1.019. We must enter this information into the interim monitoring worksheet. Select the node **Interim Monitoring** from the library and click on the  icon in the library toolbar. You will see the IM worksheet as shown below.

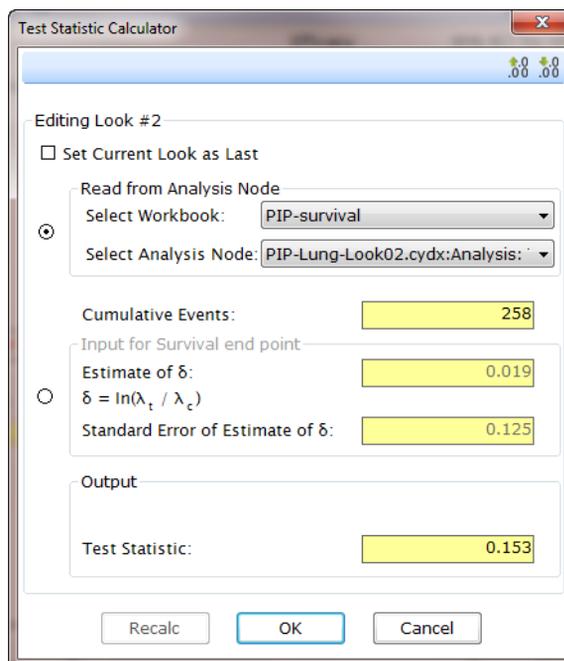


| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of 6 | Std. Error of Est. of 6 | Efficacy | | 95% RCI for HR | | Repeat p-value | CP | Predicted Power |
|--------|----------------------|-------------------|----------------|-----------|-------------------------|----------|-------|----------------|-------|----------------|----|-----------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| | | | | | | 1 | 0.186 | 119 | -0.46 | | | |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |

With **Look #2** selected, click on the **Enter Interim Data** button and choose the option to read values from look 2 analysis node. Click on Recalc button to see the test

65 Predictive Interval Plots

calculator computations as shown below.



enter in the resulting **Test Statistic Calculator**, the values for **Cumulative Events = 258**, **Estimate of delta = $\ln(1.019)$** , and **Standard Error of Estimate of delta = $2/\sqrt{258}$** , and click the **OK** button. The interim monitoring worksheet gets updated.

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | | 95% RCI for HR | | Repeat ... p-value | CP | Predicti... Power |
|--------|----------------------|-------------------|----------------|------------------|--------------------------------|----------|-------|----------------|--------|--------------------|-------|-------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| | | | | | | 1 | 0.186 | 119 | -0.461 | | | |
| 2 | 0.404 | 258 | 0.151 | 0.019 | 0.125 | 3.34 | -3.34 | 1.544 | 0.672 | 1 | 0.013 | 0.108 |
| 3 | | | | | | | | | | | | |

Now the conditional power under the current trend is only 0.014 and the predictive power is only 0.108. It is very unlikely that the trial will succeed and termination for futility appears to be a reasonable option. Before taking a final decision, however, it may be advisable to obtain a PIP for the future course of the trial under the assumption that HR=0.8 is still correct and the observed value HR=1.019 is due to variability in the data. Accordingly we invoke the PIP dialog box, enter a value of 0.8 for the hazard

ratio, and simulate the remainder of the trial 1000 times.

Input for Predicted Intervals Plot

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

PIP for Look: Final Look Required No. of Events: 639 Total Sample Size: 897

Specify Subject Info

Select Workbook: PIP-survival

Select Subject Data: PIP-Lung-Look02.cydx

Choose Variables

Population ID: TrtmtID Status Indicator: Censor1

Control: 1 1=Complete
 0=Censored

Treatment: 2 -1 = Dropout

Arrival Time: ArrivalTime Response Variable: SRVMON

Optional: Estimate Parameters from Data

Parameters of Current Dataset (Editable)

Sample Size: 797

No. of Events: 258

Accrual Rate: 50.764

Hazard Ratio (λ_t / λ_c): 1.026

Hazard Rate (λ): Control 0.033 Treatment 0.8*0.033

Dropout Hazard Rate (γ): Control 0.003 Treatment 0.003

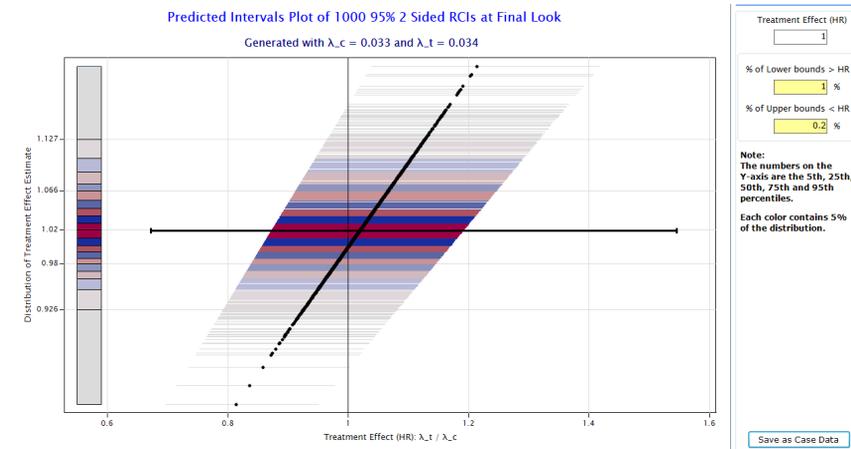
Number of Simulations: 1000

Random Number Seed

Clock

Fixed 100

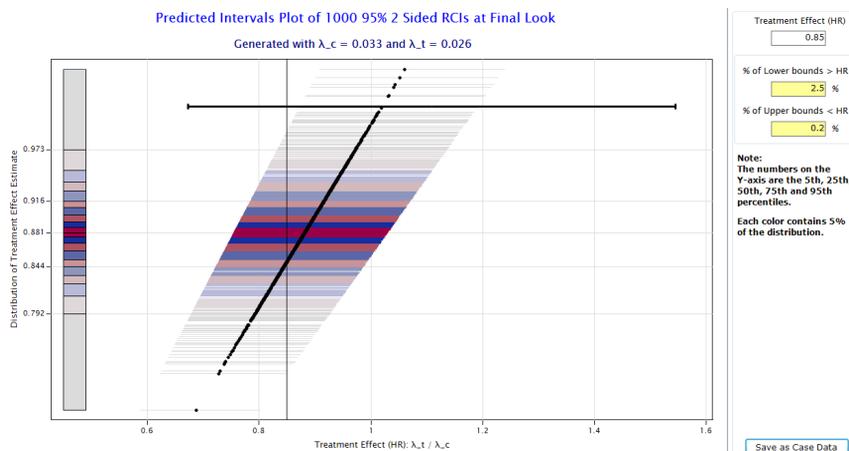
Simulate Cancel



We observe that 32.2% of the RCIs have upper bounds that are below 1. This suggests that if the trial continued and the true hazard ratio was indeed 0.8 the chance of a successful trial is 0.322. But how many of these successful outcomes would be considered clinically meaningful? Suppose that trials with observed values of HR that exceed 0.85 are not of any interest to the sponsor since there are other compounds on

65 Predictive Interval Plots

the market for this therapeutic area that have had smaller hazard ratios. Then the question becomes, how many of the 1000 RCIs would have upper bounds that are below 0.85. To answer this question, move the vertical cursor to 0.85 on the X-axis. This can be done either by dragging the cursor or (more conveniently) by entering the value 0.85 in the edit box at the top of the **Read-offs** panel of the PIP.



It is seen that 0.2% of the RCIs have upper bounds that are below 0.85 even though we generated the PIP under the optimistic assumption that the true HR=0.8. It is clearly desirable to terminate the trial for futility.

This example has shown that the RCIs provide more information than can be obtained from a conditional power calculation. The PIP may be used to determine whether a clinically meaningful treatment effect can be ruled out.

65.3 Example 2: PIP for Binomial Data

CAPTURE (Lancet, 1997; 349: 1429-35) was a randomized placebo-controlled trial of abciximab before and during coronary intervention in refractory unstable angina. After angiography, patients received a randomly assigned infusion of abciximab or placebo followed by percutaneous transluminal coronary intervention (PTCA). The primary endpoint was death from any cause within 30 days after the PTCA. The planned enrollment was 1400 patients with four equally spaced looks and stopping boundaries generated by the LD(OF) spending function. This study has 80% power to detect a 5% difference in mortality rates, from 15% on the placebo arm to 10% on the abciximab arm, at two sided $\alpha = 0.05$.

The workbook **PIP-binomial** containing this design named **capture** is already available to you in the sub-folder **Samples** in the East 6.4 installation folder in your computer. Open the **PIP-binomial** workbook in the East library. The design details are shown below.

| Test Parameters | |
|-----------------------------------|-------------------|
| Design ID | Capture |
| Design Type | Superiority |
| Number of Looks | 4 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.802 |
| Model Parameters | |
| Prop. under Control (π_c) | 0.15 |
| Prop. under Treatment (π_t) | 0.1 |
| $\delta = \pi_t - \pi_c$ | |
| Under H0 | 0 |
| Under H1 | -0.05 |
| Allocation Ratio (n_t/n_c) | 1 |
| Variance | Unpooled Estimate |
| Casagrande-Pike-Smith Correction | Not Applied |
| Boundary Parameters | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|-----------------------------------|-------------|---------------|----------|
| Sample Size (n) | | | |
| Maximum | 700 | 700 | 1400 |
| Expected H1 | 575.117 | 575.117 | 1150.235 |
| Expected H0 | 698.043 | 698.043 | 1396.086 |
| Maximum Information (I): 3218.391 | | | |

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|------------|---|----------|
| | | | | | Under H0 | Under H1 |
| | | | | | Efficacy Z | Efficacy |
| 1 | 0.25 | 350 | 7.367E-6 | -4.333 | 7.367E-6 | 0.002 |
| 2 | 0.5 | 700 | 0.002 | -2.963 | 0.002 | 0.167 |
| 3 | 0.75 | 1050 | 0.01 | -2.359 | 0.008 | 0.373 |
| 4 | 1 | 1400 | 0.025 | -2.014 | 0.015 | 0.26 |

The table below displays the results observed at each interim look.

Table 65.1: Results Observed At Each Interim Look

| Look Number | Sample Size | Placebo | Abciximab | p-value |
|-------------|-------------|----------------|----------------|---------|
| 1 | 350 | 30/175 (17.2%) | 14/175 (8%) | 0.010 |
| 2 | 700 | 55/353 (15.6%) | 37/347 (10.6%) | 0.047 |
| 3 | 1050 | 84/532 (15.8%) | 55/518 (10.6%) | 0.010 |

The stopping boundary was crossed and the Data Monitoring Committee stopped the trial

Let us enter the data from the first two looks into the interim monitoring worksheet. Select the CAPTURE design in the workbook library and click on the **IM** tool from the library toolbar.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 97.5% RCI for δ | | Repeat ... p-value | CP | Predicted ... Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|------------------------|-------|--------------------|----|---------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | | | | | | | | | | | |
| 2 | | | | | | | | | | | |
| 3 | | | | | | | | | | | |
| 4 | | | | | | | | | | | |

Now enter the data for the first two looks into the IM dashboard. For each look you

65 Predictive Interval Plots

will have to click on the **Enter Interim Data** button to invoke the test statistic calculator and enter the data look by look as described below.

For the first look, enter the data in the test statistic calculator, click on **Recalc** and **OK** buttons.

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Sample Size and Responses

| | Control | Treatment |
|-------------------------|---------|-----------|
| Cumulative Sample Size: | 175 | 175 |
| Cumulative Response: | 30 | 14 |

Cumulative Sample Size: 350

Input for Binomial end point

Estimate of δ : -0.091

$\delta = (\pi_1 - \pi_2)$

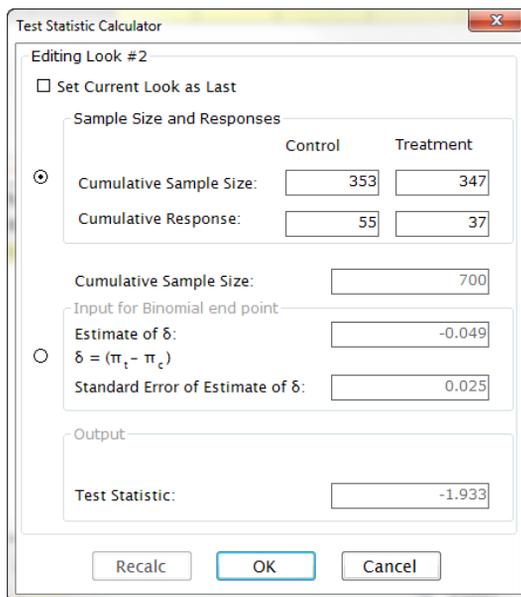
Standard Error of Estimate of δ : 0.035

Output

Test Statistic: -2.605

Recalc OK Cancel

Similarly post the data for the second look into the IM worksheet.



Now you will see the computed results posted into the IM worksheet as shown below.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 97.5% RCI for δ | | Repeat ... p-value | CP | Predicti... Power |
|--------|----------------------|------------------------|----------------|--------|----------------|----------|-----------------|-------|--------------------|-------|-------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | 0.25 | 350 | -2.605 | -0.091 | 0.035 | -4.333 | 0.061 | -1 | 0.157 | 1 | 0.971 |
| 2 | 0.5 | 700 | -1.933 | -0.049 | 0.025 | -2.963 | 0.026 | -1 | 0.118 | 0.858 | 0.778 |
| 3 | | | | | | | | | | | |
| 4 | | | | | | | | | | | |

It is evident from the above results that new drug looks promising. The conditional power is 0.858 and the predictive power is 0.778. It might be instructive at this stage to run a PIP for the future course of the trial. The data for the first two looks are stored on your computer in a .csv file named "PIP-Capture-Look02.csv". Import this file into

65 Predictive Interval Plots

East. It will be added to the library with the name "PIP-Capture-Look02.cydx".

| | Trtmnt_ID | Response |
|----|-----------|----------|
| 1 | 1 | 0 |
| 2 | 1 | 1 |
| 3 | 0 | 0 |
| 4 | 0 | 0 |
| 5 | 1 | 0 |
| 6 | 0 | 1 |
| 7 | 0 | 0 |
| 8 | 0 | 0 |
| 9 | 0 | 0 |
| 10 | 1 | 1 |

Now return to the IM dashboard by selecting Interim Monitoring node in the library. To produce the PIP for the next look (look 3), select the Look 2 row on the IM dashboard and click on the **PIP** button. Complete the Input dialog box as shown below. (Remember to click on the **Optional: Estimate Parameters from Data** button if you want East to compute the sample size and estimate the event rates from the look 2 dataset and post these parameters directly into the dialog box.)

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

PIP for Look: User Specified Lo Required Sample Size: 1050

Specify Subject Info
 Select Workbook: PIP-binomial
 Select Subject Data: PIP-Capture-Look02.cydx

Choose Variables
 Population ID: Trtmnt_ID Data contains delayed responses
 Control: 0
 Treatment: 1

Response Variable: Response
 Response Value: 1

Optional: Estimate Parameters from Data

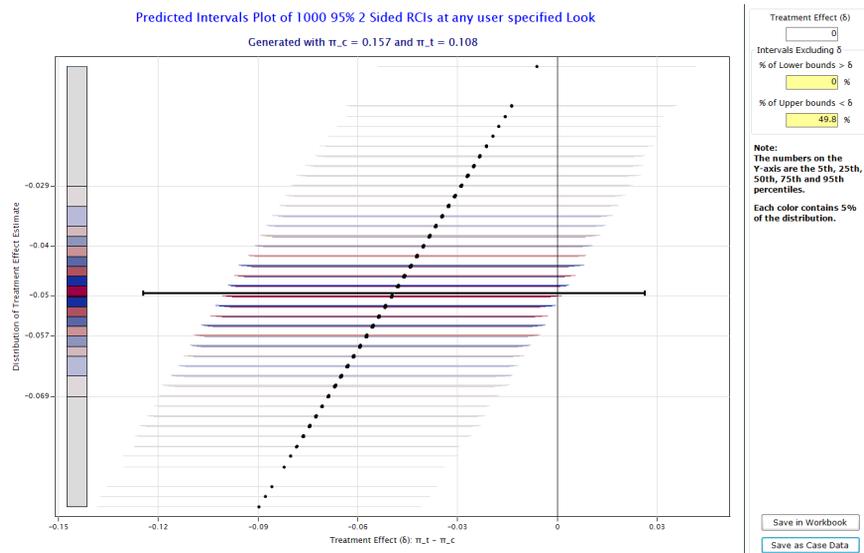
Parameters of Current Dataset (Editable)
 Sample Size: 700
 Diff. of Proportions (δ): -0.049
 Proportion under Control (π_c): 0.157
 Proportion under Treatment (π_t): 0.108

Number of Simulations: 1000
 Refresh Frequency: 100

Random Number Seed
 Clock Suppress Intermed. Output
 Fixed 100 Pause after Refresh

Simulate Cancel

Click on the **Simulate** button to generate the PIP for look 3 with 1000 repeated confidence intervals.



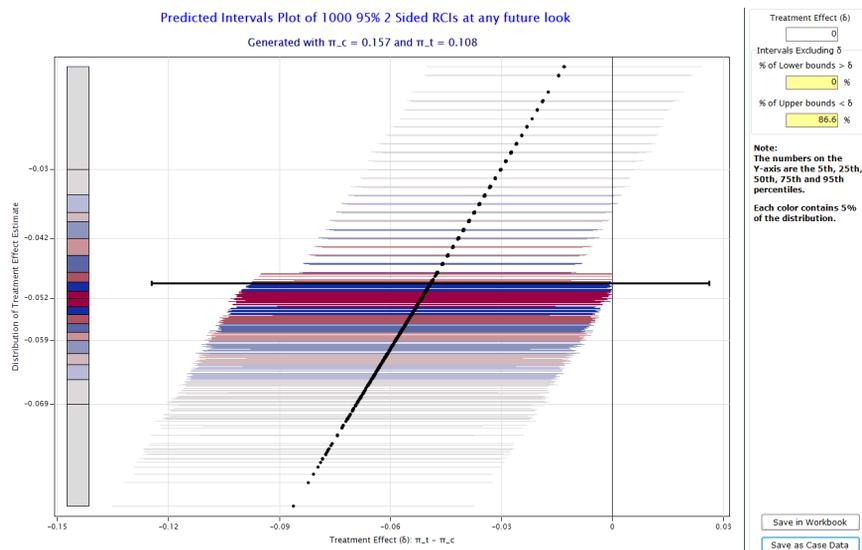
Observe that for 49.8% of the RCIs have upper bounds that exclude 0. Thus, conditional on current data and the current estimates of the event rates, there is a 49.8% chance of crossing the early-stopping boundary at the very next look. Save this PIP in the library. This can be done by clicking on the **Save in Workbook** button at the bottom right of the screen.

Suppose we wish to generate a PIP for any future look, not simply the next one. With the cursor on the PIP1 node in the library, click on the edit tool, and specify in the

65 Predictive Interval Plots

resulting input dialog box that you wish to create a PIP for "any future look".

Upon clicking the **Simulate** button the requested PIP is generated.



Overall, 86.6% of the RCIs have upper confidence bounds that are less than 0. The wider intervals are generated at Look 3 and the narrower ones are generated at Look 4. This PIP shows that the overall probability that this trial will be a success, conditional on current trends, is 0.866. The vertical rectangle with the colored bands displays the distribution of the estimated risk reductions, From this PIP we see that only the 5% of the estimated risk reductions will be less than 0.029. We now return to the IM dashboard and enter the data for look 3.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | 97.5% RCI for δ | | Repeat ... p-value | CP | Predicti... Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|------------------------|-------|--------------------|-------|-------------------|
| | | | | | | | Upper | Lower | | | |
| 1 | 0.25 | 350 | -2.605 | -0.091 | 0.035 | -4.333 | 0.061 | -1 | 0.157 | 1 | 0.971 |
| 2 | 0.5 | 700 | -1.933 | -0.049 | 0.025 | -2.963 | 0.026 | -1 | 0.118 | 0.858 | 0.778 |
| 3 | | | | | | | | | | | |
| 4 | | | | | | | | | | | |

Test Statistic Calculator

Editing Look #3

Set Current Look as Last

Sample Size and Responses

| | Control | Treatment |
|-------------------------|---------|-----------|
| Cumulative Sample Size: | 532 | 518 |
| Cumulative Response: | 84 | 55 |

Cumulative Sample Size: 1050

Input for Binomial end point

Estimate of δ : -0.052

$\delta = (\pi_1 - \pi_2)$

Standard Error of Estimate of δ : 0.021

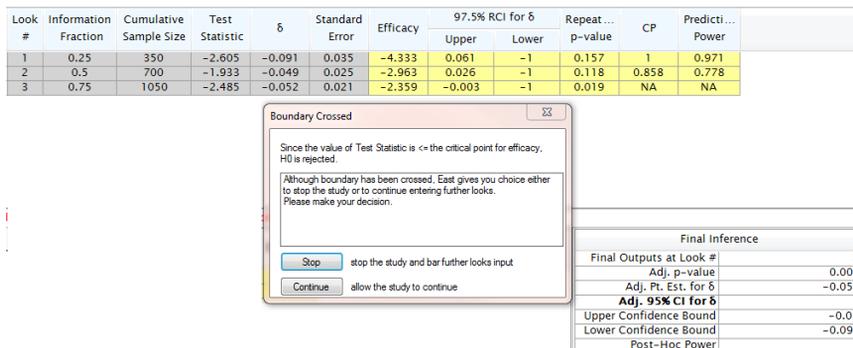
Output

Test Statistic: -2.485

Click **OK** on the test statistic calculator to post the computed values for third look. Now

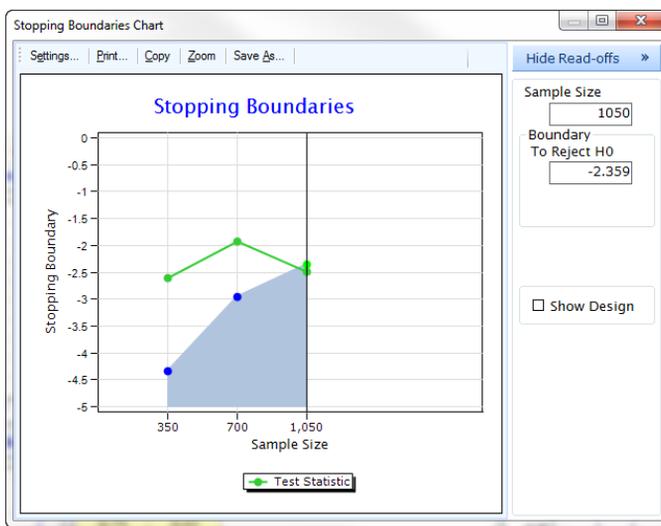
65 Predictive Interval Plots

the boundary is crossed and you are presented with the message on boundary crossing.



You have to decide now, on the choice of either stopping or continuing the trial. In the actual trial the data monitoring committee (DMC) recommended that the trial be terminated and the sponsor agreed with the recommendation. Thus abciximab was declared to be superior to placebo in this class of patients with respect to all causes mortality, at the two-sided 5% level of significance.

Notice, however, that the stopping boundary was barely crossed.



The efficacy boundary is -2.359 while the corresponding test statistic is -2.485. Had there been one less event on the Control arm and one more on the Treatment arm, the efficacy boundary would not have been crossed and the study would have continued to the final look after enrolling 1400 patients. Now the DMC is charged with examining the totality of evidence, including safety issues and consistency across secondary endpoints before recommending that a trial be terminated. Therefore sometimes, in close situations like this one, the DMC might well recommend that the trial not be terminated prematurely but rather that it continue to the end so as to achieve a robust result that can alter medical practice. In such a situation the DMC might find a PIP for the final look to be a valuable additional piece of information to help it with the decision making. In order to illustrate this, now stop the trial first. Next, click on PIP button while on the third look in the IM worksheet. You will be presented with PIP dialog box where you fill in the details. The dialog box will look as shown below.

Input for Predicted Intervals Plot

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

PIP for Look: Final Look Required Sample Size: 1400

Specify Subject Info

Select Workbook: PIP-binomial

Select Subject Data: PIP-Capture-Look03.cydx

Choose Variables

Population ID: Trtmnt_ID Data contains delayed responses

Control: 0

Treatment: 1

Response Variable: Response

Response Value: 1

Optional: Estimate Parameters from Data

Parameters of Current Dataset (Editable)

Sample Size: 1050

Diff. of Proportions (δ): -0.052

Proportion under Control (π_c): 0.159

Proportion under Treatment (π_t): 0.107

Number of Simulations: 1000

Refresh Frequency: 100

Random Number Seed

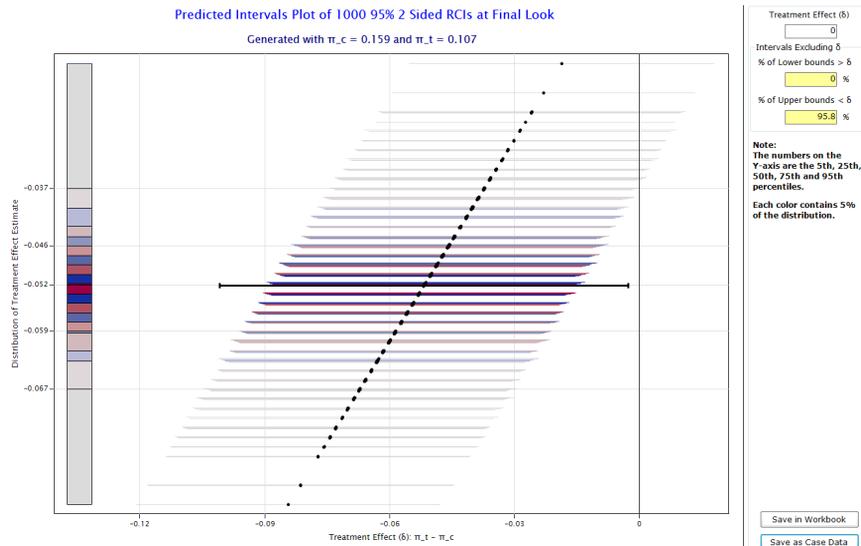
Clock Suppress Intermed. Output

Fixed 100 Pause after Refresh

Simulate Cancel

65 Predictive Interval Plots

Click on **Simulate** button. The PI plot will be generated as shown below.



We see that 95.8% of the RCIs have upper confidence bounds that are below zero. Thus the trial were to continue to look 4, there is only a 4% chance that it would fail to achieve statistical significance. Moreover, the vertical bar near the Y-axis displaying the distribution of the estimates of treatment effect shows in 95% of the simulations the absolute risk reduction is at least -0.037. This is the type of robust result that the trial needs to obtain in order to alter medical practice. Thus the DMC might weigh the trade-off between terminating the trial immediately with a relatively marginal result or proceeding to take one more look with a high probability of achieving a stronger result.

65.4 Example 3: PIP for Continuous Outcome Data

We thank the AIDS Clinical Trials Group (ACTG) for permitting us to use this dataset. NARC 009 was a prospective, randomized, double-blind, placebo-controlled, multicenter, clinical trial of Prosaptide (PRO) conducted by the Neurologic AIDS Research Consortium for the treatment of HIV-associated neuropathic pain. Subjects were randomized to a daily dose of 2, 4, 8 or 16 mg PRO or placebo via subcutaneous injection. The primary endpoint was the 6-week reduction from baseline in the weekly average of random daily Gracely pain scale prompts, collected using an electronic diary. The trial randomized a total of 390 subjects in equal proportion to the five treatment arms. With 78 patients/arm the trial is capable of detecting a difference (treatment minus control) in the change from baseline of $\delta = -0.2$ Gracely units with

93% power, for each dose versus placebo comparison, assuming a common $\sigma = 0.35$, two-sided $\alpha = 0.05$, and no correction for multiplicity. One interim analysis was planned when about half the patients had enrolled.

In this example, we will only consider the comparison of the 2 mg dose and placebo. The design is saved in the East workbook named PIP-normal. Please bring up this workbook into your East library.

Design: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means

Test Parameters

| | |
|--------------------|-------------|
| Design ID | NARC009 |
| Design Type | Superiority |
| Number of Looks | 2 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.931 |

Model Parameters

| | |
|--------------------------------|---------------------|
| Test Statistic | Z |
| Input Method | Difference of Means |
| $\delta = \mu_1 - \mu_2$ | 0 |
| Under H0 | 0 |
| Under H1 | -0.2 |
| Std. Deviation (σ) | 0.35 |
| Allocation Ratio (n_1/n_2) | 1 |

Boundary Parameters

| | |
|-------------------|---------|
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OP) |

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|-----------------|-------------|---------------|---------|
| Sample Size (n) | 73 | 73 | 146 |
| Maximum | 62,163 | 61,862 | 124,025 |
| Expected H1 | 72,89 | 72,887 | 145,777 |

Maximum Information (I): 297,959

Stopping Boundaries: Look by Look

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | | Boundary Crossing Probability (Incremental) | | | |
|--------|--------------------------|-----------------|---------------------------|------------|--------|---|-------|----------|-------|
| | | | | Efficacy Z | | Under H0 | | Under H1 | |
| | | | | Upper | Lower | Upper | Lower | Upper | Lower |
| 1 | 0.5 | 73 | 0.003 | 2.963 | -2.963 | 0.002 | 0.002 | 3.262E-8 | 0.301 |
| 2 | 1 | 146 | 0.05 | 1.969 | -1.969 | 0.023 | 0.023 | 2.9E-8 | 0.63 |

At the time of the interim analysis a total of 65 patients had enrolled to the two arms of the trial. The interim analysis data are stored in a .csv file named **NARC_02mg.csv**. Import this dataset into East.

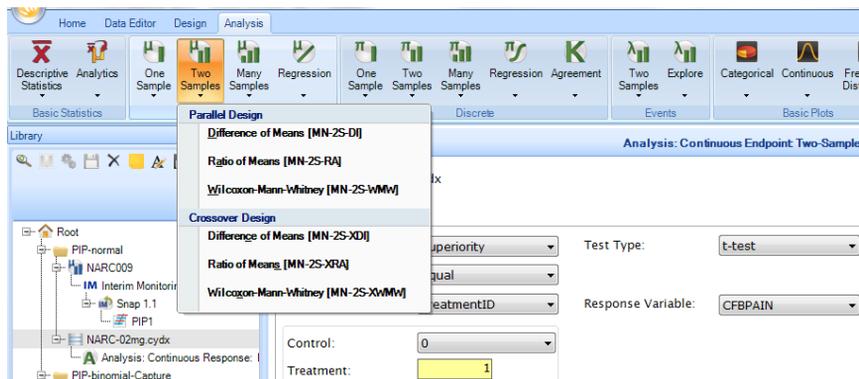
PatientID: 1 Value: 1

| | PatientID | TreatmentID | CFBPAIN |
|----|-----------|-------------|--------------|
| 1 | 1 | 0 | -1.17777778 |
| 2 | 2 | 0 | -0.80825 |
| 3 | 3 | 0 | -0.669901786 |
| 4 | 4 | 0 | -0.665297619 |
| 5 | 5 | 0 | -0.658944444 |
| 6 | 6 | 0 | -0.567191667 |
| 7 | 7 | 0 | -0.541198413 |
| 8 | 8 | 0 | -0.533261905 |
| 9 | 9 | 0 | -0.416964286 |
| 10 | 10 | 0 | -0.356419048 |

Then perform the **Two Samples > Difference of Means** test on the

65 Predictive Interval Plots

imported dataset.



Click on **OK** at the bottom of Analysis input dialog box. You will get the following analysis results.

Output

Response Variable: CFBPAIN
 Total Number of Records: 65
 Number of Records Rejected: 0
 Summary of the Observed Data:

| Treatment Group | Min | Max | Median | Mean | Std.Dev | n |
|-----------------|--------|-------|--------|--------|---------|----|
| 0 | -1.178 | 0.238 | -0.086 | -0.231 | 0.333 | 31 |
| 1 | -1.36 | 0.366 | -0.174 | -0.251 | 0.366 | 34 |

Test of Hypothesis:

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | 1-Sided | | 2-Sided | 95% Confidence Interval(2-Sided) | |
|----|---------------------|------------------|------------|-------|----|---------|------|---------|----------------------------------|-------------|
| | | | | | | p-value | Tail | | p-value | Lower Limit |
| 65 | -0.019 | -0.055 | 0.087 | -0.22 | 63 | 0.413 | L.E. | 0.826 | -0.193 | 0.155 |

The observed value of δ is only -0.019 with a standard error of 0.087. We will enter these results into the interim monitoring worksheet. Select the NARC009 node in the library and click on the **IM** tool in the library toolbar. Then click on the **Enter Interim Data** button and enter the sample size, estimate of δ and standard error

into the test statistic calculator and press **OK**.

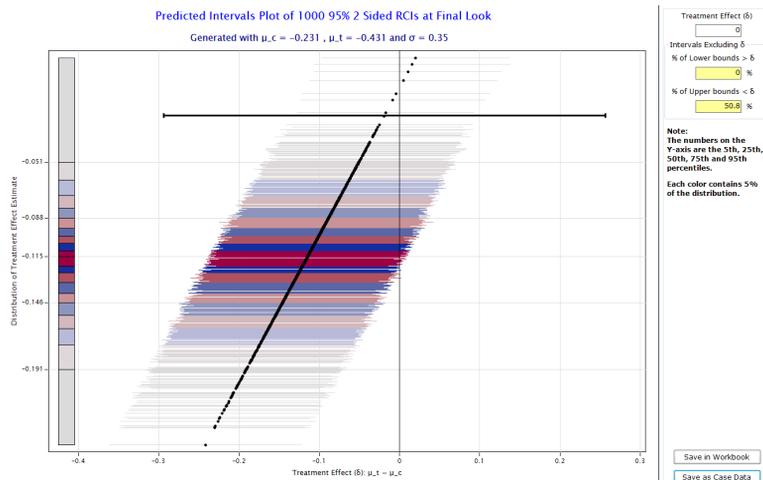
The computed values will now be posted in the IM worksheet.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | | 95% RCI for δ | | Repeat ... p-value | CP | Predicti... Power |
|--------|----------------------|------------------------|----------------|----------|----------------|----------|--------|----------------------|--------|--------------------|-------|-------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.445 | 65 | -0.218 | -0.019 | 0.087 | 3.163 | -3.163 | 0.256 | -0.294 | 1 | 0.015 | 0.091 |
| 2 | | | | | | | | | | | | |

These interim results are rather poor. With a conditional power of only 0.015 and a predictive power of only 0.091 under the current trend, this trial is likely to fail. Before terminating the trial for futility, however, it would be useful to generate a PIP with 1000 RCIs generated under the design assumption that the true value of $\delta = -0.2$. With the Look 1 row selected on the Interim Monitoring worksheet, click on the PIP button, click on 'Optional: Estimate Parameters from Data' button, and complete the

65 Predictive Interval Plots

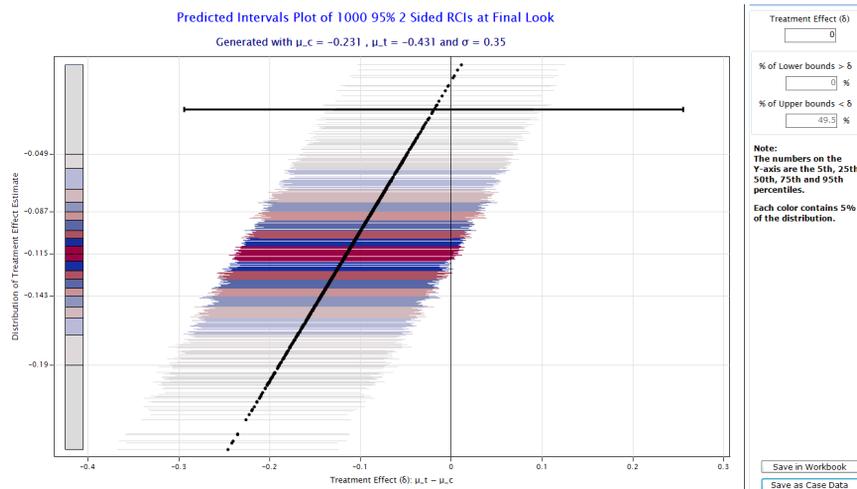
dialog box as shown.



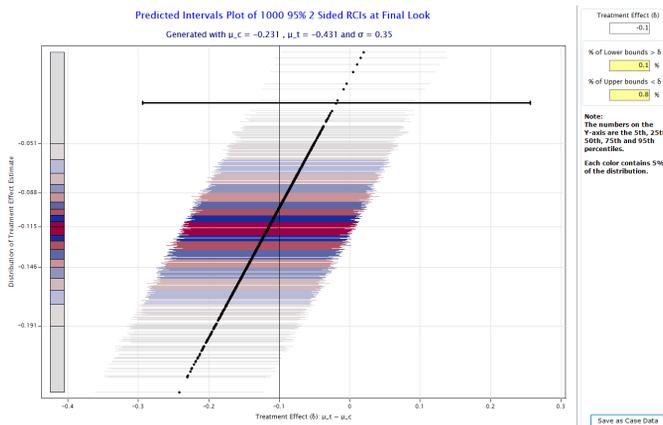
Notice that the PIP is generated for final look is for look 3, not for look 2.

Since look 1 was taken earlier than scheduled, after 65 subjects, the look that was actually designated as look 1 with 73 subjects, is becomes look 2. Thus the final look, with 146 subjects becomes look 3. Also the value of $\mu_t = \mu_c - 0.2 = -0.431$.

Click on the **Simulate** button to generate the PIP.



Under the optimistic assumption that the true value of $\delta = -0.2$ we see that 50.8% of the RCIs have upper bounds that are less than 0. This would suggest that the trial continue to the next look. It is important to point out, however, that the smallest value of δ that would be considered clinically meaningful is $\delta = -0.1$. Accordingly, drag the cursor to -0.1 on the X-axis (or type -0.1, in the edit box at the top of the Read-offs panel).



65 *Predictive Interval Plots*

It is now seen that only 1% of the RCI's have upper bounds that are less than -0.1. Moreover, these RCI's were generated under the optimistic assumption that the true $\delta = -0.2$. We may thus feel confident that terminating the trial for futility is the correct decision.

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

EastPredict is an enrollment/events prediction procedure that models the subject enrollment process. In general, the enrollment rate for a specific trial can be estimated based on past experience and any relevant information on that trial. However, this rate is only an estimate and the actual enrollment in a period needs to be treated as a random variable with a certain probability distribution. EastPredict module models this uncertainty in enrollment through the assumption that the subject arrival pattern follows a known probability distribution. In this chapter, we demonstrate the features of EastPredict (henceforth ‘East’) using examples of studies with normal, binomial, and survival endpoints.

Important Note: In this chapter, we will use four examples for three endpoints - normal (Orlistat trial), binomial (Capture trial), and survival (Rales trial and Oncox trial) to illustrate enrollment/events prediction procedures. The main purpose of these procedures is to predict at any time point of the study, the likely cumulative enrollment/completers/dropouts for normal and binomial studies and enrollment/events/dropouts for survival studies. A study may be terminated at a particular time point, because of a decision as per group sequential procedure. In that case, any prediction made for a subsequent time point will have no meaning. So the procedures described in this and the next chapter, predict what would materialize if the study reaches any particular time point, ignoring the possibility of earlier termination by crossing a group sequential boundary. In this way, the predictive procedures cover all possible scenarios, whether the study is likely to terminate earlier or later.

66.1 Normal Design

66.1.1 The Orlistat Trial: Initial Design

66.1.2 Simulating the Orlistat Trial

66.1.3 Output

This section uses inputs from the Orlistat trial described in Chapter 10 and extends the example by adding site information and accrual information to the simulation design.

66.1.1 The Orlistat Trial: Initial Design

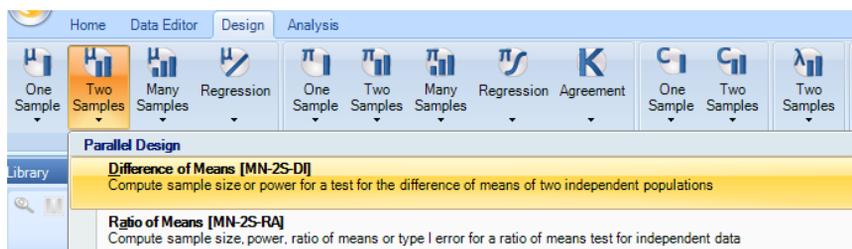
The drug Orlistat was developed to treat obesity by promoting weight loss. Its efficacy was tested by randomizing patients into the treatment group or the control group according to the ratio 3:1, and comparing the resulting weight loss of the two groups after one year. The following assumptions were made:

- Expected mean weight loss in the treatment group: 9 kg
- Expected mean weight loss in the control group: 6 kg
- Standard deviation of weight change: 8 kg

Eighteen sites participated in the trial. The accrual rate was expected to be 100 subjects per year with a dropout rate of 10% and a response lag of 1 year.

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

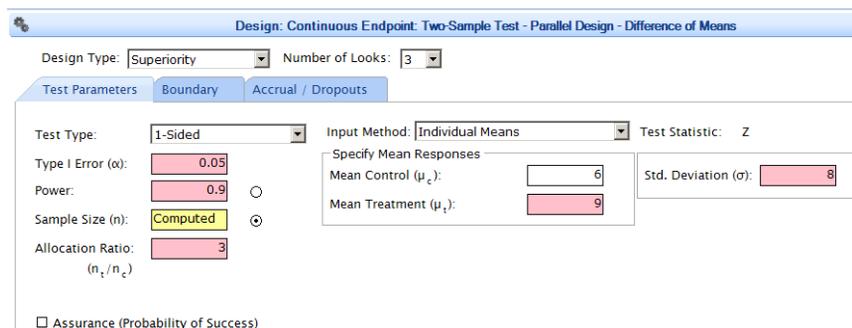
To design this trial navigate to the Design ribbon and select **Two Samples** under the Continuous tab and then **Difference of Means**, the first option under Parallel Designs.



This will open an input dialog box, where you enter the following design parameters of the Orlistat trial in the corresponding fields:

- Design Type: Superiority
- Number of Looks: 3
- Test Type: 1-Sided
- Type-1 Error: 0.05
- Power: 0.9
- Allocation Ratio (n_t/n_c): 3
- Mean Control (μ_c): 6
- Mean Treatment (μ_t): 9
- Std. Deviation (σ): 8

Click on the **Include Options** button in the top right-hand corner and select **Accrual/Dropouts** which opens a third tab of the same name. The design window then appears as follows:



In the **Boundary** tab, we specify the details for the **Efficacy** boundary, and the spacing of the looks. We keep the spending function as the default Lan DeMets (OF) function. The spacing of the looks is defined in the column **Info. Fraction** which has a range of (0, 1.000]. When set to **Equal** the looks are distributed equidistantly across this range. Setting the spacing of looks to **Unequal** allows us to define at which points the looks occur by changing the corresponding information fractions.

Let us assume that in the Orlistat trial all three looks are equally spaced and the dialog box will appear as shown below.

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary Accrual / Dropouts

Efficacy

Boundary Family: Spending Functions
 Spending Function: Lan-DeMets
 Parameter: OF
 Type I Error (α): 0.05

Futility

Boundary Family: None

Spacing of Looks: Equal Unequal

Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary |
|--------|----------------|---------------------|-------------------|
| 1 | 0.333 | 0.001 | 3.200 |
| 2 | 0.667 | 0.016 | 2.141 |
| 3 | 1.000 | 0.050 | 1.695 |

The final step is to add the accrual/dropout information. Click on the **Accrual/Dropouts** tab and set **Accrual Rate** to 100, **Response Lag** to 1 and **Probability of Dropout** to 0.1. Note that East does not require the unit of time to be specified explicitly as long as consistency is maintained in the parameters given. In other words, we may choose the unit of analysis to be years, months or weeks as long as all time-related data (overall accrual rate, response lag, dropout rate, individual site accrual rates, etc.) is also expressed in terms of the same unit. Later examples will

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

demonstrate the use of months and weeks as units of analysis.

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary Accrual / Dropouts

Accrual Info
Accrual Rate: 100

Response Lag Info
Response Lag: 1

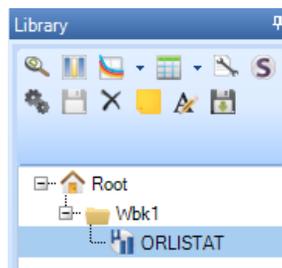
Dropout Info
Probability of Dropout: 0.1

We have entered all the parameters required for East to determine the sample size. Click **Compute** in the bottom right-hand corner of the design window. The following output preview is displayed in the lower panel when the computation is complete.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Accrual Rate | Response Lag | Dropout Prob. | Study Duration | Sample Size |
|------|-------------|--------------|-----------|--------------------|-------|-------|------------------|-------------------|--------------|--------------|---------------|----------------|-------------|
| Des1 | Superiority | 3 | 1-Sided | 0.05 | 0.9 | 3 | Equal | LD (OP) | 100 | 1 | 0.1 | 4.68 | 368 |

East has determined that a sample size of 368 subjects is required to attain a power of 0.9. The trial is expected to be around 4.68 years long. In the next section we introduce the simulation feature to explore the enrollment process of this trial given information about the sites over which it will be conducted.

Rename the design 'ORLISTAT' using the button in the **Output Preview** pane and then save it using . It will then appear in the **Library** pane on the left-hand side of the East interface in a workbook named 'Wbk1', which you can rename as 'Orlistat'.



66.1.2 Simulating the Orlistat Trial

The primary input in the simulation is the enrollment plan which contains the following information for each site:

- Site initiation period: the time period over which the site is expected to be initialized so that it is ready to begin enrolling subjects
- Site accrual rate: the number of subjects expected to arrive at the site over the unit of time chosen (in this case, 'year')
- Enrollment cap: the maximum number of subjects that may be enrolled at the site. This enrollment cap also applies to the entire study. This means that no single site or all the sites put together can enroll more than this enrollment cap.

The table below shows a sample enrollment plan for Orlistat.

| Site Name | Site Initiation (Start) | Site Initiation (End) | Accrual Rate | Enrollment Cap |
|-----------|-------------------------|-----------------------|--------------|----------------|
| Site 1 | 0 | 0 | 10 | 368 |
| Site 2 | 0 | 0.25 | 10 | 368 |
| Site 3 | 0 | 0.25 | 10 | 368 |
| Site 4 | 0 | 0.25 | 10 | 368 |
| Site 5 | 0 | 0.25 | 10 | 368 |
| Site 6 | 0 | 0.25 | 5 | 368 |
| Site 7 | 0 | 0.25 | 5 | 368 |
| Site 8 | 0 | 0.25 | 5 | 368 |
| Site 9 | 0 | 0.25 | 5 | 368 |
| Site 10 | 0 | 0.25 | 5 | 368 |
| Site 11 | 0 | 0.25 | 5 | 368 |
| Site 12 | 0 | 0.25 | 5 | 368 |
| Site 13 | 0 | 0.25 | 5 | 368 |
| Site 14 | 0 | 0.25 | 2 | 368 |
| Site 15 | 0 | 0.25 | 2 | 368 |
| Site 16 | 0 | 0.25 | 2 | 368 |
| Site 17 | 0 | 0.25 | 2 | 368 |
| Site 18 | 0 | 0.25 | 2 | 368 |

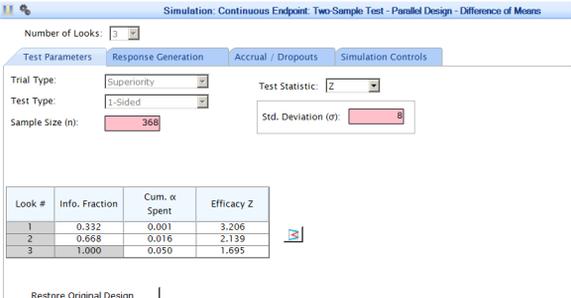
Recall that all parameters in this example are in annual terms, thus a site initiation end time of 0.25 for Sites 2 to 18 indicates that these sites must be initiated within 3 months. In the case of Site 1, the start and end times of '0' indicate that the site is ready to begin enrolling subjects immediately. In addition, note that the individual site accrual rates must sum up to the overall accrual rate specified during the design time,

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

such as in the plan above where all the site accrual rates sum to 100. Lastly, the enrollment cap for each site is generally set to the sample size.

Let us simulate the enrollment process of the Orlistat trial under this enrollment plan.

Select ‘ORLISTAT’ in the **Library** pane and click . This opens the simulation input dialog box containing four tabs: **Test Parameters**, **Response Generation**, **Accrual/Dropouts**, and **Simulation Controls**. Select the **Test Statistic Z**.



| Look # | Info. Fraction | Cum. α Spent | Efficacy Z |
|--------|----------------|---------------------|------------|
| 1 | 0.332 | 0.001 | 3.206 |
| 2 | 0.668 | 0.016 | 2.139 |
| 3 | 1.000 | 0.050 | 1.695 |

The first three tabs contain the trial details we had entered in the initial design phase. In the **Simulation Controls** tab we can specify the number of simulation runs we wish to make as well as general output options.

The enrollment plan is to be specified in the **Accrual/Dropouts** tab. Click on **Include Options** in the upper right-hand corner and select **Site**. The **Accrual/Dropouts** tab then provides an option to select the accrual model and a grid in which the enrollment plan can be filled in:

Accrual Model East models the variation in accrual rate by assuming that subjects arrive according to one of two probability distributions: Uniform or Poisson. Under the uniform model, the arrival times of subjects are sampled from a uniform distribution over the given time interval. The Poisson model assumes that subjects arrive according to a Poisson process and thus their inter-arrival times are sampled from an exponential distribution. Experience suggests that arrivals follow a Poisson process and so for all examples in this chapter we select the Poisson accrual model.

Enrollment Plan When entering the enrollment plan we must select whether we will specify it by region or by site. When we select **Sites by Region** it is assumed that all sites within a region have the same parameters (site initiation periods, accrual rates and enrollment caps), while selecting **Sites** allows us to specify enrollment parameters

individually by site. The enrollment plan of Orlistat shown above was specified by site, thus we select **Sites**.

The site parameters can be entered manually in the grid after specifying Number of Sites. Alternatively, you may create a spreadsheet such as the one shown below and save it as a comma-separated values (CSV) file and then import it.

| | A | B | C | D | E |
|----|--------|----------|--------|-------|------|
| 1 | SiteID | SIPstart | SIPend | Arate | Ecap |
| 2 | SITE1 | 0 | 0 | 10 | 368 |
| 3 | SITE2 | 0 | 0.25 | 10 | 368 |
| 4 | SITE3 | 0 | 0.25 | 10 | 368 |
| 5 | SITE4 | 0 | 0.25 | 10 | 368 |
| 6 | SITE5 | 0 | 0.25 | 10 | 368 |
| 7 | SITE6 | 0 | 0.25 | 5 | 368 |
| 8 | SITE7 | 0 | 0.25 | 5 | 368 |
| 9 | SITE8 | 0 | 0.25 | 5 | 368 |
| 10 | SITE9 | 0 | 0.25 | 5 | 368 |
| 11 | SITE10 | 0 | 0.25 | 5 | 368 |
| 12 | SITE11 | 0 | 0.25 | 5 | 368 |
| 13 | SITE12 | 0 | 0.25 | 5 | 368 |
| 14 | SITE13 | 0 | 0.25 | 5 | 368 |
| 15 | SITE14 | 0 | 0.25 | 2 | 368 |
| 16 | SITE15 | 0 | 0.25 | 2 | 368 |
| 17 | SITE16 | 0 | 0.25 | 2 | 368 |
| 18 | SITE17 | 0 | 0.25 | 2 | 368 |
| 19 | SITE18 | 0 | 0.25 | 2 | 368 |

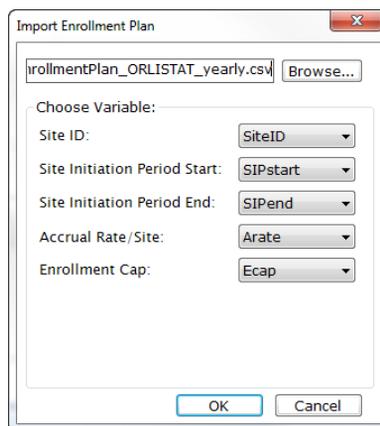
In the above data, ‘SiteID’ corresponds to the site name, ‘SIPstart’ refers to the Site Initiation Start and ‘SIPend’ to Site Initiation End. ‘Arate’ and ‘Ecap’ refer to the site accrual rate and enrollment cap respectively.

For your convenience, this CSV file is already created and stored in **Samples** subfolder in your East installation folder, under the name **EnrollmentPlan_ORLISTAT_yearly.csv**. You may import this CSV file by clicking on **Home** -- > **Import** menu item and choosing the CSV file from **Samples** subfolder. This imported CSV file will appear as a node under ORLISTAT workbook with the extension .cydx which is the format for East data files.

Click on **Specify Enrollment Plan...** button and select the workbook and the imported CSV file, now with the extension .cydx. Next, use the dropdown boxes in the **Choose Variable** panel to match the header names in your .cydx file to the column names shown in the East interface. In our example the final Import **Enrollment Plan**

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

window would appear as follows:



When all these inputs are entered correctly the **Accrual/Dropouts** tab appears as shown below:

Test Parameters Response Generation **Accrual/Dropout** Simulation Controls

Accrual Model:

Sites By Regions Sites

Number of Sites:

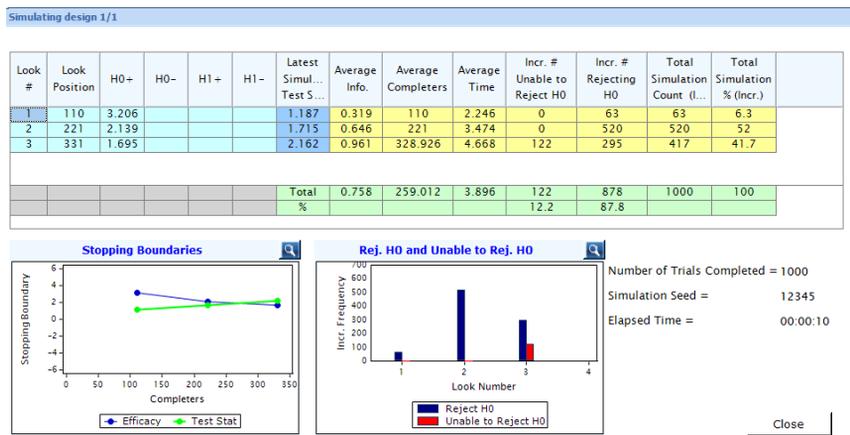
Response Lag Info
Response Lag:

Dropout Info
Probability of Dropout:

| Site ID | Site Initiation Period | | Accrual Rate /Site | Enrollment Cap |
|---------|------------------------|------|--------------------|----------------|
| | Start | End | | |
| SITE1 | 0 | 0 | 10 | 368 |
| SITE2 | 0 | 0.25 | 10 | 368 |
| SITE3 | 0 | 0.25 | 10 | 368 |
| SITE4 | 0 | 0.25 | 10 | 368 |
| SITE5 | 0 | 0.25 | 10 | 368 |
| SITE6 | 0 | 0.25 | 5 | 368 |
| SITE7 | 0 | 0.25 | 5 | 368 |

As a final step, let us navigate to the **Simulation Controls** and set the number of simulations to 1000. Choose the **Fixed** seed as 12345. We can now simulate this design by clicking the **Simulate** button in the lower right-hand corner. East displays

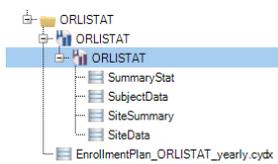
the following window after it carries out the required simulation runs:



Once the specified number of simulations has been run we can close the simulating design window and see a one-line summary of the output in the **Output Preview** pane with the ID Sim1:

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Accrual Rate | Response Lag | Dropout Prob. | Study Duration | Sample Size |
|----------|-------------|--------------|-----------|--------------------|-------|-------|------------------|-------------------|--------------|--------------|---------------|----------------|-------------|
| Orlistat | Superiority | 3 | 1-Sided | 0.05 | 0.9 | 3 | Equal | LD (OF) | 100 | 1 | 0.1 | 4.68 | 368 |
| Sim1 | Superiority | 3 | 1-Sided | | 0.878 | | User Specified | User Specified | | 1 | 0.1 | | |

Then save Sim1 by clicking on the button. It will appear as a sub-node of the design in 'Orlistat' in the **Library** pane on the left-hand side of the East interface along with four spreadsheets containing detailed information from the simulation runs. Click on the Sim1 node and rename it 'ORLISTAT' using the . Note that East uses the blue icon to denote designs and the brown icon to denote simulations.



66 Enrollment/Events Prediction - At Design Stage (By Simulation)

66.1.3 Output

All outputs from the simulation can be accessed from the **Library** pane. Double-clicking on the 'ORLISTAT' simulation opens a general summary page containing four output tables.

The first output table, **Average Sample Size, Dropouts and Look Times**, displays the average over all 1000 simulations of the sample size (the number of subjects enrolled in the study), completers (subjects who completed the one year period till follow-up), dropouts (subjects who dropped out of the study) and pipeline (subjects who enrolled but did not complete or drop out of the system formally). The table also contains the average look time for all three looks, for instance we observe that on average the first look took place at 2.336 years.

The table **Simulation Boundaries and Boundary Crossing Probabilities** displays the efficacy boundary at each look and the number of simulations in which the boundary was crossed. In total, the trial was stopped for efficacy in 878 simulations resulting in the average power at termination of 0.878.

Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size (n) | Average Completers (s) | Average Dropouts (d) | Average Pipeline (n-s-d) | Average Look Time | Final Analysis For (Incr.) |
|---------|-------------------------|------------------------|----------------------|--------------------------|-------------------|----------------------------|
| | | | | | | Efficacy (%) |
| 1 | 222.205 | 110 | 12.262 | 99.943 | 2.336 | 4.7 |
| 2 | 345.271 | 221 | 24.662 | 99.609 | 3.572 | 50 |
| 3 | 368 | 328.926 | 36.144 | 2.93 | 4.779 | 30 |
| Average | 347.356 | 259.012 | 28.753 | 59.591 | 3.996 | |
| % | | | | | | 84.700% |

Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Completers | Boundaries | Stopping For | Total Simulations | |
|--------|------------|------------|--------------|-------------------|---------|
| | | Efficacy | | Count | % |
| | | | Efficacy | | |
| 1 | 110 | 3.206 | 63 | 63 | 6.300% |
| 2 | 221 | 2.139 | 520 | 520 | 52.000% |
| 3 | 331 | 1.695 | 295 | 417 | 41.700% |
| Total | | | 878 | 1000 | |
| % | | | 87.800% | | |

The third table summarizes the enrollment plan, and the final table **Overall Look-Wise Output** shows the number of completers, accruals and dropouts over a

range of percentiles at each look.

Enrollment Plan

Accrual Model: Poisson

Number of Sites: 18

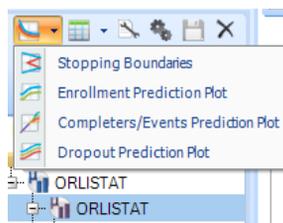
| Site ID | Site Initiation Period Start | Site Initiation Period End | Accrual Rate /Site | Enrollment Cap |
|---------|------------------------------|----------------------------|--------------------|----------------|
| SITE1 | 0 | 0 | 10 | 368 |
| SITE2 | 0 | 0.25 | 10 | 368 |
| SITE3 | 0 | 0.25 | 10 | 368 |
| SITE4 | 0 | 0.25 | 10 | 368 |
| SITE5 | 0 | 0.25 | 10 | 368 |
| SITE6 | 0 | 0.25 | 5 | 368 |
| SITE7 | 0 | 0.25 | 5 | 368 |
| SITE8 | 0 | 0.25 | 5 | 368 |
| SITE9 | 0 | 0.25 | 5 | 368 |
| SITE10 | 0 | 0.25 | 5 | 368 |
| SITE11 | 0 | 0.25 | 5 | 368 |
| SITE12 | 0 | 0.25 | 5 | 368 |
| SITE13 | 0 | 0.25 | 5 | 368 |
| SITE14 | 0 | 0.25 | 2 | 368 |
| SITE15 | 0 | 0.25 | 2 | 368 |
| SITE16 | 0 | 0.25 | 2 | 368 |
| SITE17 | 0 | 0.25 | 2 | 368 |

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

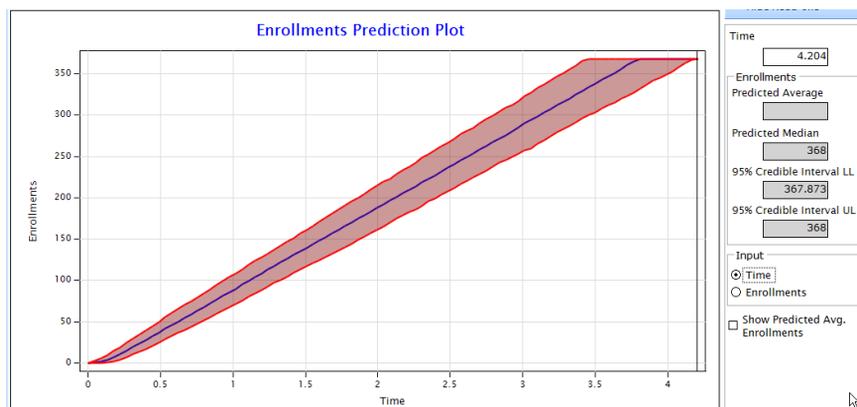
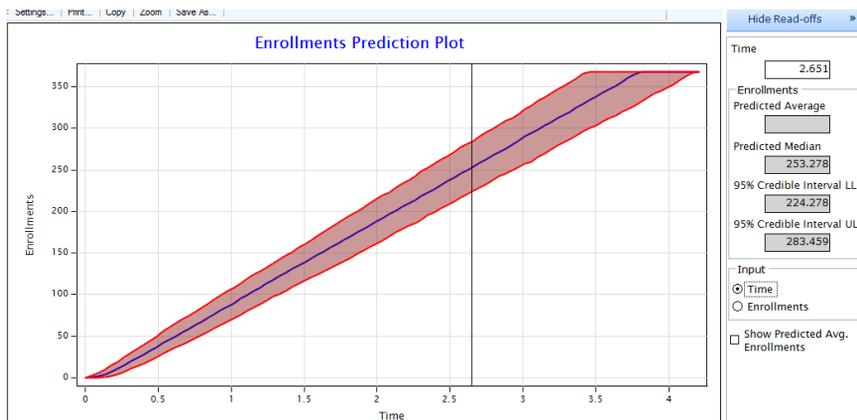
Overall Look-Wise Output

| | Look No. | Percentile | No. of Completers | No. of Sites Opened | No. of Accruals | No. of Dropouts |
|---|----------|------------|-------------------|---------------------|-----------------|-----------------|
| - | 1 | 5% | 110 | 18 | 205 | 6 |
| | | 25% | | 18 | 214 | 10 |
| | | 50% | | 18 | 222 | 12 |
| | | 75% | | 18 | 230 | 15 |
| | | 95% | | 18 | 240 | 19 |
| | | Average | | 18 | 222.205 | 12.262 |
| - | 2 | 5% | 221 | 18 | 326 | 16 |
| | | 25% | | 18 | 338 | 21 |
| | | 50% | | 18 | 346 | 24 |
| | | 75% | | 18 | 353 | 28 |
| | | 95% | | 18 | 363 | 33 |
| | | Average | | 18 | 345.271 | 24.662 |
| - | 3 | 5% | 331 | 18 | 368 | 25 |
| | | 25% | | 18 | 368 | 32 |
| | | 50% | | 18 | 368 | 37 |
| | | 75% | | 18 | 368 | 41 |
| | | 95% | | 18 | 368 | 46 |
| | | Average | | 18 | 368 | 36.144 |

In addition, East generates a series of plots depicting the timelines of enrollment, completers and dropouts. These plots can be accessed by selecting the ‘ORLISTAT’ simulation in the **Library** pane and clicking the  button.



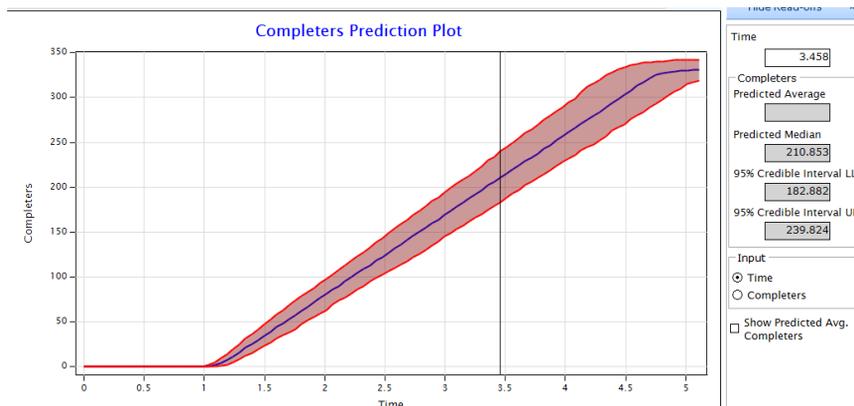
The **Enrollment Prediction Plot** displays the number of enrollments against time. It shows the predicted median and average enrollments across all simulations as well as the 95% confidence interval. For instance, at the time 2.651, indicated by the vertical marker the number of enrollments reached 253 in 97.5% of the simulations, while in 2.5% of the simulations the number of enrollments was below 224. Overall, East predicts a maximum accrual duration of around 4.2 years to enroll 368 subjects.



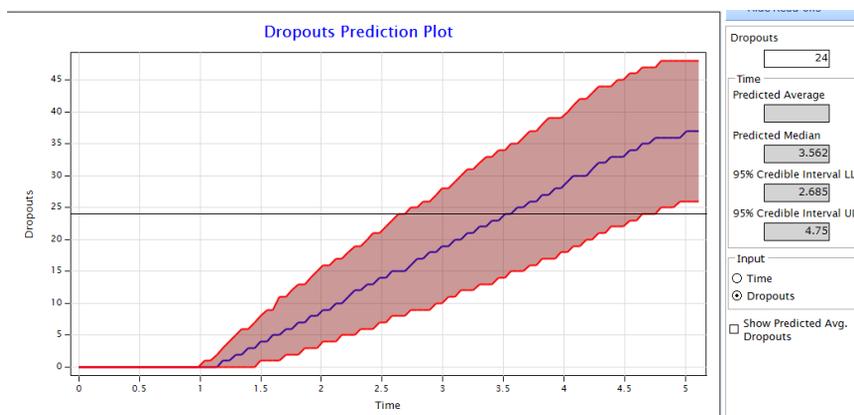
The **Completers Prediction Plot** displays the number of completers over time in terms of the 95% confidence interval, mean and median. In the case of normal and binomial designs the shape of the **Completers Prediction Plot** resembles that of the **Enrollment Prediction Plot**, with the main difference being that it is off-set to the right corresponding to the length of the response lag (one year, in the case of Orlistat). In addition, the prediction lines of the completers are slightly lower than the

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

enrollments due to the number of dropouts.



The **Dropout Prediction Plot** shows the fairly steady increase in dropouts as the trial progresses. The median number of dropouts by the end of the study is 36, as we would expect given the 10% dropout rate.



Lastly, there are four output files nested below the 'ORLISTAT' simulation node containing the full details of all the simulation runs. These files, named **SummaryStat**, **SubjectData**, **SiteSummary**, and **SiteData**, are the source of the data displayed in the tables and plots described above.

66.2 Binomial Design

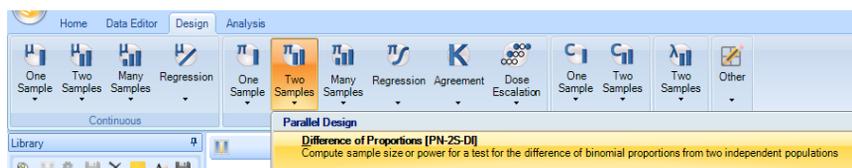
- 66.2.1 The CAPTURE Trial: Initial Design
- 66.2.2 Simulating the CAPTURE Trial
- 66.2.3 Output

In the following section we simulate the CAPTURE trial introduced in Chapter 3. It is an example of a binomial design, where the aim was to compare two independent samples in terms of the difference of proportions in event rate.

66.2.1 The CAPTURE Trial: Initial Design

The CAPTURE trial compared the performance of the drug Abciximab and a placebo on event rate. The null hypothesis H_0 stated that both the drug and the placebo had an event rate of 15%, versus the alternative hypothesis H_1 that Abciximab reduces the event rate from 15% to 10%. The study was 2-sided with a power of 0.8 and an α of 0.05. The accrual rate was 12 subjects/week, the probability of dropout was 5% and the response lag was 4 weeks.

To design this trial, click on the Design ribbon and select ‘Two Samples’ under the Discrete tab and then click on ‘Difference of Proportions’:



This opens an input dialog box:

In the relevant fields of the dialog box, fill in the design parameters of the CAPTURE trial that are summarized below:

- Design Type: Superiority
- Number of Looks: 3
- Test Type: 2-Sided
- Type-1 Error: 0.05
- Power: 0.8
- Prop. Under Control (π_c): 0.15
- Prop. Under Treatment (π_t): 0.1
- Allocation Ratio : 1

Next, click on the **Include Options** button, in the top right-hand corner and select **Accrual/Dropouts**. This opens an additional tab in which we can specify the accrual rate, response lag and the probability of subjects dropping out of the trial.

When the design parameters are filled in correctly the Test Parameters window appears

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

as follows:

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary Accrual / Dropouts

Test Type: 2-Sided

Type I Error (α): 0.05

Power: 0.8

Sample Size (n): Computed

Allocation Ratio: 1 (n_1/n_2)

Specify Proportion Response
Prop. under Control (π_c): 0.15

Specify Alternative Hypothesis
Prop. under Treatment (π_t): 0.1
Diff. in Prop. ($\delta_1 = \pi_t - \pi_c$): -0.05

Specify Variance
 Pooled Estimate
 Unpooled Estimate

Use Casagrande-Pike-Smith Correction (Ignored if alloc. ratio is not 1)

Assurance (Probability of Success):

In the **Boundary** tab we specify the details for the **Efficacy** boundary, the spacing of the looks, the boundary families and spending functions. We keep the default spending function of Lan DeMets (OF) for this design. The spacing of the looks is defined in the column **Info. Fraction** which has a range of (0, 1.000]. When the spacing of the looks is set to **Equal** the values of the information fraction are distributed equally across the range. If we wish specify when each interim look will be taken we can set the spacing of looks to **Unequal** and then enter the desired information fractions corresponding to the time points at which the interim looks shall occur. For this example let us assume all three looks are equally spaced.

Design Type: Superiority Number of Looks: 3

Test Parameters Boundary Accrual / Dropouts

Efficacy

Boundary Family: Spending Functions

Spending Function: Lan-DeMets

Parameter: OF

Type I Error (α): 0.05

Futility

Boundary Family: None

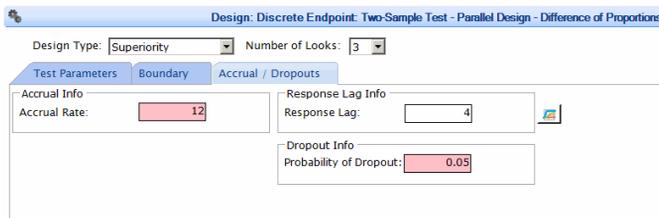
Spacing of Looks Equal Unequal

Efficacy Boundary: Z Scale

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|---------------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.333 | 0.000 | 3.710 | -3.710 |
| 2 | 0.667 | 0.012 | 2.511 | -2.511 |
| 3 | 1.000 | 0.050 | 1.993 | -1.993 |

In the **Accrual/Dropouts** tab, set **Accrual Rate** to 12, **Response Lag** to 4 and

Probability of Dropout to 0.05.

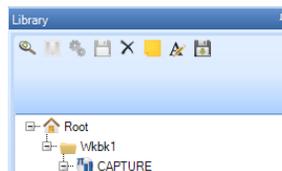


We have entered all the parameters required for East to determine the sample size. Click **Compute** to obtain a preview of the output.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Accrual Rate | Response Lag | Dropout Prob. | Study Duration | Sample Size |
|------|-------------|--------------|-----------|--------------------|-------|-------|------------------|-------------------|--------------|--------------|---------------|----------------|-------------|
| Des1 | Superiority | 3 | 2-Sided | 0.05 | 0.8 | 1 | Equal | LD (OF) | 12 | 4 | 0.05 | 125.333 | 1456 |

East has determined that a sample size of 1456 subjects is required to attain a power of 0.8. The trial is expected to be approximately 125 weeks long. Let us simulate this trial to explore its enrollment timeline.

Rename the design ‘CAPTURE’ using the  tool in the **Output Preview** pane. It will then appear in the **Library** pane on the left-hand side of the East interface in a workbook named ‘Wkbk1’, which also you can rename as ‘CAPTURE’.



66.2.2 Simulating the CAPTURE Trial

The primary input in the simulation is the enrollment plan which contains the following information for each site:

- Site initiation period: the time period over which the site is expected to be initialized so that it is ready to begin enrolling subjects
- Site accrual rate: the number of subjects expected to arrive at the site over the unit of time chosen (in this case, ‘week’)

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

- Enrollment cap: the maximum number of subjects that may be enrolled at the site. This enrollment cap also applies to the entire study. This means that no single site or all the sites put together can enroll more than this enrollment cap.

The table below shows a sample enrollment plan for the CAPTURE trial.

| Site Name | Site Initiation (Start) | Site Initiation (End) | Accrual Rate | Enrollment Cap |
|-----------|-------------------------|-----------------------|--------------|----------------|
| Site 1 | 0 | 0 | 1 | 1456 |
| Site 2 | 0 | 10 | 1 | 1456 |
| Site 3 | 0 | 10 | 1 | 1456 |
| Site 4 | 0 | 10 | 1 | 1456 |
| Site 5 | 0 | 10 | 1 | 1456 |
| Site 6 | 0 | 10 | 0.75 | 1456 |
| Site 7 | 0 | 10 | 0.75 | 1456 |
| Site 8 | 0 | 10 | 0.75 | 1456 |
| Site 9 | 0 | 10 | 0.75 | 1456 |
| Site 10 | 0 | 10 | 0.75 | 1456 |
| Site 11 | 0 | 10 | 0.5 | 1456 |
| Site 12 | 0 | 10 | 0.5 | 1456 |
| Site 13 | 0 | 10 | 0.5 | 1456 |
| Site 14 | 0 | 10 | 0.25 | 1456 |
| Site 15 | 0 | 10 | 0.25 | 1456 |
| Site 16 | 0 | 10 | 0.25 | 1456 |
| Site 17 | 0 | 10 | 0.25 | 1456 |
| Site 18 | 0 | 10 | 0.25 | 1456 |
| Site 19 | 0 | 10 | 0.25 | 1456 |
| Site 20 | 0 | 10 | 0.25 | 1456 |

Under this enrollment plan, Site1 initiates immediately and the remaining 19 sites must initiate within 10 weeks of the start time. The accrual rates are given per site per week and sum up to the overall accrual rate of 12. The enrollment cap of each site is set to the estimated total sample size of the study. We shall simulate the CAPTURE trial using this enrollment plan.

To access the simulation tool select 'CAPTURE' in the **Library** pane and click . This opens the simulation input dialog box containing four tabs: **Test Parameters**, **Response Generation**, **Accrual/Dropouts** and **Simulation Controls**.

The **Simulation Controls** tab is where we specify the number of simulation runs. The remaining three tabs contain the trial details we had entered in the initial design phase.

Click on **Include Options** in the upper right-hand corner and select **Site** to add information about the number of sites and their enrollment parameters.

Accrual Model We have the choice to specify whether the arrival times of subjects are to be sampled under a uniform model or a Poisson model. Under the uniform model,

the arrival times of subjects are sampled from a uniform distribution over the given time interval. The Poisson model assumes that subjects arrive according to a Poisson process and thus their inter-arrival times are sampled from an exponential distribution. Let us use the Poisson accrual model as it is known to be a more realistic representation of the subject arrival process.

Enrollment Plan We must choose whether to specify the enrollment plan by region or by site. Under **Sites by Region** East assumes that all sites within a region have the same parameters (site initiation periods, accrual rates and enrollment caps), while selecting **Sites** allows us to specify enrollment parameters individually. Let us specify the CAPTURE enrollment plan by **Sites**.

Enter the parameters in the enrollment plan grid either manually or by creating a spreadsheet such as the one shown below, saving it as a CSV file, import it using **Home** -- > **Imports** menu item to appear as a node with extension .cydx, and then select it using the **Specify Enrollment Plan...** button.

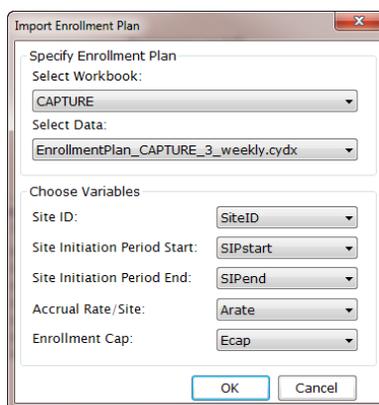
| Site Name | Site Initiation (Start) | Site Initiation (End) | Accrual Rate | Enrollment Cap |
|-----------|-------------------------|-----------------------|--------------|----------------|
| Site 1 | 0 | 0 | 1 | 1456 |
| Site 2 | 0 | 10 | 1 | 1456 |
| Site 3 | 0 | 10 | 1 | 1456 |
| Site 4 | 0 | 10 | 1 | 1456 |
| Site 5 | 0 | 10 | 1 | 1456 |
| Site 6 | 0 | 10 | 0.75 | 1456 |
| Site 7 | 0 | 10 | 0.75 | 1456 |
| Site 8 | 0 | 10 | 0.75 | 1456 |
| Site 9 | 0 | 10 | 0.75 | 1456 |
| Site 10 | 0 | 10 | 0.75 | 1456 |
| Site 11 | 0 | 10 | 0.5 | 1456 |
| Site 12 | 0 | 10 | 0.5 | 1456 |
| Site 13 | 0 | 10 | 0.5 | 1456 |
| Site 14 | 0 | 10 | 0.25 | 1456 |
| Site 15 | 0 | 10 | 0.25 | 1456 |
| Site 16 | 0 | 10 | 0.25 | 1456 |
| Site 17 | 0 | 10 | 0.25 | 1456 |
| Site 18 | 0 | 10 | 0.25 | 1456 |
| Site 19 | 0 | 10 | 0.25 | 1456 |
| Site 20 | 0 | 10 | 0.25 | 1456 |

For your convenience this CSV file is already created and stored in the **Samples** subfolder in your East installation folder, under the name **EnrollmentPlan_CAPTURE_3_weekly.csv**. In this CSV file, ‘SiteID’ corresponds to the site name, ‘SIPstart’ refers to the Site Initiation Start and ‘SIPend’ to Site Initiation End. ‘Arate’ and ‘Ecap’ refer to the site accrual rate and enrollment cap respectively. You may import this CSV file by clicking on **Home** -- > **Import** menu item and choosing the CSV file from **Samples** subfolder. This imported CSV

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

file will appear as a node under CAPTURE workbook with the extension .cydx which is the format for East data files.

Click on **Specify Enrollment Plan...** button and specify the workbook and the imported CSV file, now with the extension .cydx. After selecting the .cydx file, use the dropdown boxes in the **Choose Variable** panel to match the header names in your .cydx file to the column names shown in the East interface. Using the names in our .cydx file the final **Specify Enrollment Plan** window would appear as follows:



After clicking **OK** the grid should contain the CAPTURE enrollment plan and the complete **Accrual/Dropout** tab should appear as shown below:

Simulation: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

Number of Looks: 3

Test Parameters | Response Generation | **Accrual / Dropouts** | Simulation Controls

Accrual Model: Poisson

Sites By Regions Sites

Specify Enrollment Plan...

Number of Sites: 20

Response Lag Info
Response Lag: 4

Dropout Info
Probability of Dropout: 0.05

| Site ID | Site Initiation Period Start | Site Initiation Period End | Accrual Rate /Site | Enrollment Cap |
|---------|------------------------------|----------------------------|--------------------|----------------|
| SITE1 | 0 | 0 | 1 | 1456 |
| SITE2 | 0 | 10 | 1 | 1456 |
| SITE3 | 0 | 10 | 1 | 1456 |
| SITE4 | 0 | 10 | 1 | 1456 |
| SITE5 | 0 | 10 | 1 | 1456 |

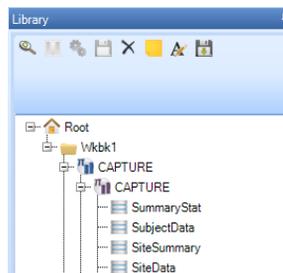
Set the number of simulations to 1000 in the **Simulation Control Info** tab; select **Random Number Seed** as **Fixed** equal to 12345 and then simulate the design by clicking the **Simulate** button in the lower right-hand corner.

Once the simulation is complete and we close the simulating window a one-line summary of the output is shown in the **Output Preview** pane:

| ID | Design Type | Test Type | Sample Size | Completers | Power | πc (Data) | πt (Data) | No. of Looks |
|------|-------------|-----------|-------------|------------|-------|----------------|----------------|--------------|
| Sim1 | Superiority | 2-Sided | 1456 | 1383 | 0.811 | 0.15 | 0.1 | 3 |

Click on the summary, rename it ‘CAPTURE’ using the  button and then save it by clicking on the  button.

It will appear as a sub-node of the design in ‘Wkbk1’ in the **Library** pane on the left-hand side of the East interface along with four spreadsheets containing detailed information from the simulation runs.



Note that East uses the blue icon  to denote designs and the brown icon  to denote simulations.

66.2.3 Output

The **Library** pane contains all the output from the simulation of the CAPTURE trial. The general summary is accessed by double-clicking on the ‘CAPTURE’ simulation. The first table, **Average Sample Size, Dropouts and Look Times**, shows us the average over all 1000 simulations of the sample size (the number of subjects enrolled in the study), completers (subjects who completed the one year period till follow-up), dropouts (subjects who dropped out of the study) and pipeline (subjects who enrolled but did not complete or drop out of the system formally). In addition it provides the

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

average look time for all three looks. For instance, the first look took place on average at around 49.0 weeks, the second look at 89.4 weeks and the final look at 129.428 weeks. These interim looks are approximately 40 weeks apart, reflecting the equally spaced look times we specified in the **Boundary Info** tab.

In the table **Simulation Boundaries and Boundary Crossing Probabilities** we can see the efficacy boundary and number of completers at each look. By end of the study the null hypothesis was rejected in 818 out of 1000 simulations, resulting in the power of 0.818.

Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size (n) | Average Completers (s) | Average Dropouts (d) | Average Pipeline (n-s-d) | Average Look Time | Final Analysis For (Incr.) | |
|---------|-------------------------|------------------------|----------------------|--------------------------|-------------------|----------------------------|--------------------|
| | | | | | | Upper Efficacy (%) | Lower Efficacy (%) |
| 1 | 533.019 | 461 | 24.192 | 47.827 | 49.023 | 0 | 1.3 |
| 2 | 1018.363 | 922 | 48.415 | 47.948 | 89.459 | 0 | 37 |
| 3 | 1456 | 1379.72 | 72.482 | 3.798 | 129.428 | 0 | 40.8 |
| Average | 1268.842 | 1184.682 | 62.257 | 21.903 | 112.43 | | |
| % | | | | | | 0.000% | 79.100% |

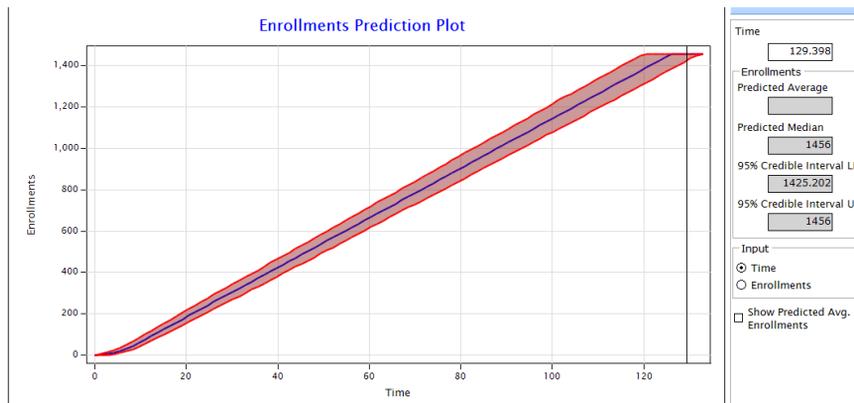
Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Completers | Boundaries | | Stopping For | | Total Simulations | |
|--------|------------|------------|--------|----------------|----------------|-------------------|---------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 461 | 3.712 | -3.712 | 0 | 15 | 15 | 1.500% |
| 2 | 922 | 2.511 | -2.511 | 0 | 396 | 396 | 39.600% |
| 3 | 1383 | 1.993 | -1.993 | 0 | 407 | 589 | 58.900% |
| Total | | | | 0 | 818 | 1000 | |
| % | | | | 0.000% | 81.800% | | |

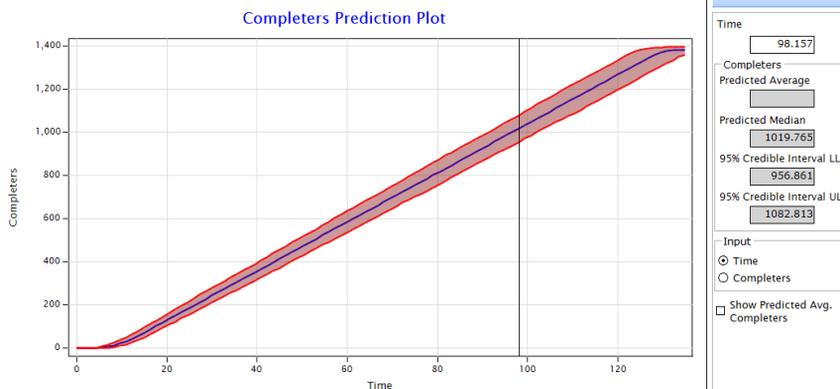
East aggregates the data contained in these data files to generate plots showing the enrollment process over time. These plots can be accessed by clicking the  button in the **Library** pane.

The **Enrollment Prediction Plot** displays the number of enrollments against time and shows us how long it is expected to take for the target number of enrollments to be reached. In this case the predicted median enrollment of 1456 was completed at

around 129 weeks, closer to the initial computation of 125 weeks.



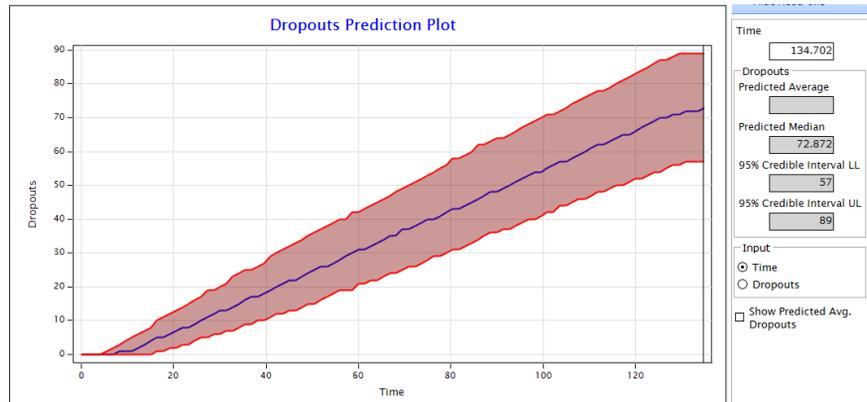
In the **Completers Prediction Plot** we can see the number of completers over time. While the number of subjects is lower due to dropouts, the plot itself is very similar to the **Enrollment Prediction Plot** owing to the relatively short response lag of 4 weeks.

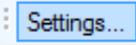


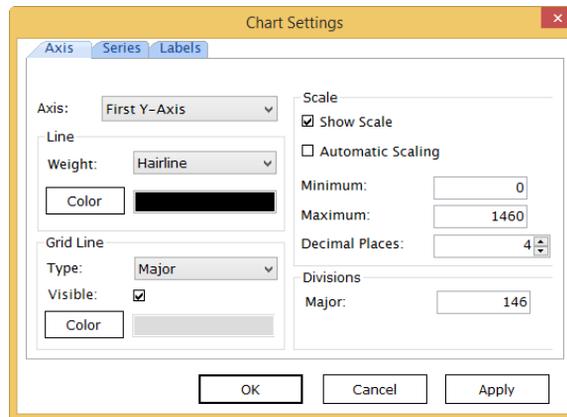
Lastly, the **Dropout Prediction Plot** shows the cumulative number of dropouts over the accrual duration and indicates that the median number of dropouts at the end of the

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

trial was about 73.



For all the plots, the axes and labels can be adjusted using the  button which invokes the **Chart Settings** menu:



Finally, East produces four files containing the full data generated in the simulations. These files, named **SummaryStat**, **SubjectData**, **SiteSummary**, and **SiteData** are the source of the data displayed in the tables and plots described above and can be accessed from the **Library**.

66.3 Survival Design- Example 1

- 66.3.1 The RALES Trial:
Initial Design
- 66.3.2 Simulating the
RALES Trial
- 66.3.3 Output

The next example is based on the RALES trial described in Chapter 43. The aim of this trial was to compare survival in two groups: a treatment group receiving Aldactone for heart failure, and a control group. As in the previous examples, we extend the RALES simulation to incorporate accrual information and study the enrollment and events prediction.

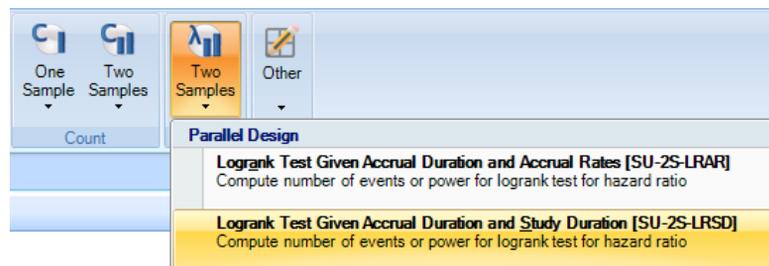
66.3.1 The RALES Trial: Initial Design

Aldactone was developed to treat patients with severe heart failure. The randomized aldactone evaluation study (RALES) was a six-year double blind trial comparing survival rates of a treatment group that was administered Aldactone and a control group that received a placebo. The placebo group was known to have a mortality rate of 38%, and the aim of RALES was to ascertain with a power of 0.9 whether Aldactone was successful in reducing that mortality rate by 17% (from 38% to 31.54%) in the treatment group. The study was a two-sided test with $\alpha = 0.05$ and an expected dropout rate of 5% in both groups. Subjects were enrolled over a period of 1.7 years and there were 6 interim looks scheduled over the duration of the study.

Suppose we wish to design this trial using ‘months’ as our unit of analysis instead of ‘years’. In that case, the relevant parameters would be adjusted as follows:

- Accrual rate: $960/12 = 80$ subjects/month
- Accrual duration: $1.7 \times 12 = 20.4$ months
- Study duration: $6 \times 12 = 72$ months
- Hazard rate (treatment): $0.3154/12 = 0.0263$
- Hazard rate (control): $0.38/12 = 0.0317$

Let us implement this design in East. Click on the **Two Sample** button in the Survival category on the Design ribbon and select **Logrank Test Given Accrual Duration and Study Duration**.



This opens the survival design dialog box with default values.

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

Enter the following design parameters of the RALES trial in the corresponding fields:

- Design Type: Superiority
- Number of Looks: 6
- Test Type: 2-Sided
- Type-1 Error: 0.05
- Power: 0.9
- Allocation Ratio: 1

The next step is to enter the survival information in the right-hand portion of the **Test Parameters** tab. Set **# of Hazard Pieces** to '1' and let **Input Method** be 'Hazard Rates'. Fill in **Hazard Rate (Control)** as '0.0317' and **Hazard Rate (Treatment)** as '0.0263'. The **Hazard Ratio** is then automatically computed as 0.83:

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates

Design Type: Superiority Number of Looks: 6

Test Parameters Boundary Accrual / Dropouts

Test Type: 2-Sided # of Hazard Pieces: 1 Input Method: Hazard Rates

Hazard Ratio (Optional) Alternative

Hazard Ratio (λ_t / λ_c) 0.83

Log Hazard Ratio $\ln(\lambda_t / \lambda_c)$ -0.187

| Hazard Rate | |
|-------------|-----------------|
| Control | Treatment: Alt. |
| 0.0317 | 0.0263 |

Variance of Log Hazard Ratio

Null Alternative

Assurance (Probability of Success)

In the **Boundary** tab we specify the details for the **Efficacy** boundary and the spacing of the interim looks. We keep the default spending function of Lan DeMets (OF). When set to **Equal** the looks are distributed equidistantly across the (0, 1.000] range of the **Info. Fraction**. Setting the spacing of looks to **Unequal** allows us to choose when the interim looks take place by setting the information fractions accordingly. For this example let us assume all looks are equally spaced.

Design Parameters Boundary Info **Accrual / Dropout Info**

Efficacy
 Boundary Family: Spending Functions
 Spending Function: Lan-DeMets
 Parameter: OF
 Type I Error (α): 0.05

Futility
 Boundary Family: None

Spacing of Looks Efficacy Boundary: Z Scale

Equal Unequal

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|--------|----------------|--------------|-------------------|--------|
| | | | Upper | Lower |
| 1 | 0.167 | 0.000 | 5.367 | -5.367 |
| 2 | 0.333 | 0.000 | 3.710 | -3.710 |
| 3 | 0.500 | 0.003 | 2.970 | -2.970 |
| 4 | 0.667 | 0.012 | 2.539 | -2.539 |
| 5 | 0.833 | 0.028 | 2.252 | -2.252 |

In the final tab we can enter the accrual and dropout information. Recall that RALES had an accrual duration of 20.4 months (1.7 years) and a total study duration of 72 months (6 years). Enter these values in their respective fields while leaving **# of Accrual Periods** as '1'. Also, in the RALES trial 5% of the subjects are expected to drop out. This can be specified in the **Piecewise Dropout Information** panel either in terms of hazard rates or probability. Achieving the 5% dropout is a trial and error process as described in Chapter 50. Set **# of Pieces** to '1' and **Input Method** to 'Prob. of Dropout'. Set both **Prob. of Dropout (Control)** and **Prob. of Dropout (Treatment)** to 0.05 and initially set **By Time** to 12 (months). The final **Accrual/Dropout** tab should appear as follows:

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

Design Type: Superiority Number of Looks: 6

Test Parameters Boundary **Accrual / Dropouts**

Subjects are followed: Until End of Study

Accrual Info
 Accrual Duration: 20.4 Study Duration: 72
 # of Accrual Periods: 1

| Period # | By Time | Cum. % Accrued |
|----------|---------|----------------|
| 1 | 20.400 | 100.000 |

Piecewise Dropout Information
 # of Pieces: 1 Input Method: Prob. of Dropout
 By Time: 12
 Prob. of Dropout (Control): 0.05
 Prob. of Dropout (Treatment): 0.05

Note: Period 1 hazard rates apply after time 12.

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

Click **Compute** to determine the required accruals and events for this trial.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events |
|------|-------------|--------------|-----------|--------------------|-------|-------|------------------|-------------------|-------------|------------------|------------------|----------------|
| Des1 | Superiority | 6 | 2-Sided | 0.05 | 0.9 | 1 | Equal | LD (OF) | 1638 | 1637.996 | 1637.987 | 1238 |

Rename this design 'RALES' using  and then save it in the library using . Let us simulate this trial to study its enrollment process.

66.3.2 Simulating the RALES Trial

The primary input in the simulation is the enrollment plan which contains the following information for each site:

- Site initiation period: the time period over which the site is expected to be initialized so that it is ready to begin enrolling subjects
- Site accrual rate: the number of subjects expected to arrive at the site over the unit of time chosen (in this case, 'month')
- Enrollment cap: the maximum number of subjects that may be enrolled at the site. This enrollment cap also applies to the entire study. This means that no single site or all the sites put together can enroll more than this enrollment cap.

The table below shows a sample enrollment plan for the RALES trial.

| Site Name | Site Initiation (Start) | Site Initiation (End) | Accrual Rate | Enrollment Cap |
|-----------|-------------------------|-----------------------|--------------|----------------|
| SITE1 | 0 | 0 | 8 | 1000 |
| SITE2 | 0 | 1 | 8 | 1000 |
| SITE3 | 0 | 1 | 4 | 1000 |
| SITE4 | 0 | 1 | 4 | 1000 |
| SITE5 | 0 | 1 | 4 | 1000 |
| SITE6 | 0 | 1 | 4 | 1000 |
| SITE7 | 0 | 1 | 4 | 1000 |
| SITE8 | 0 | 1 | 4 | 1000 |
| SITE9 | 0 | 1 | 4 | 1000 |
| SITE10 | 0 | 1 | 4 | 1000 |
| SITE11 | 0 | 1 | 4 | 1000 |
| SITE12 | 0 | 1 | 4 | 1000 |
| SITE13 | 0 | 1 | 4 | 1000 |
| SITE14 | 0 | 1 | 4 | 1000 |
| SITE15 | 0 | 1 | 4 | 1000 |
| SITE16 | 0 | 1 | 4 | 1000 |
| SITE17 | 0 | 1 | 2 | 1000 |
| SITE18 | 0 | 1 | 2 | 1000 |
| SITE19 | 0 | 1 | 2 | 1000 |
| SITE20 | 0 | 1 | 2 | 1000 |

We see from this enrollment plan that there are 20 sites participating in the study and each site may enroll a maximum of 1000 subjects. Site 1 initiates immediately and the remaining 19 sites must initiate within 1 month of the start of the study. The accrual rates are given in terms of subjects arriving per site per month and sum up to the overall monthly accrual rate of 80. We shall use this enrollment plan in our simulation of the RALES trial.

Select 'RALES' in the **Library** pane and click  to open the simulating design window containing the tabs **Simulation Parameters**, **Response Generation**, **Accrual/Dropouts**, and **Simulation Controls**. The first three tabs contain the trial details we had entered in the initial design phase. The **Simulation Controls** tab is

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

where we specify the number of runs.

Simulation Parameters
Response Generation Info
Accrual/Dropout Info
Simulation Control Info

Trial Type: Superiority

Test Type: 2-Sided

Max. # of Events: 1238

Fix at Each Look: Total No. of Events

| Look # | Info. Fraction | Cum. α Spent | | Efficacy Z | |
|--------|----------------|---------------------|-------|------------|--------|
| | | Upper | Lower | Upper | Lower |
| 1 | 0.166 | 0.000 | 0.000 | 5.371 | -5.371 |
| 2 | 0.334 | 0.000 | 0.000 | 3.709 | -3.709 |
| 3 | 0.500 | 0.002 | 0.002 | 2.970 | -2.970 |
| 4 | 0.666 | 0.006 | 0.006 | 2.539 | -2.539 |
| 5 | 0.834 | 0.014 | 0.014 | 2.252 | -2.252 |

Click on the **Include Options** box to add **Site** :

Include Options

Site
 Randomization
 User Defined R. Function
 Stratification

The main inputs we must provide in the **Accrual/Dropouts** tab are the accrual model and the enrollment plan.

Accrual Model We have the choice to specify whether the arrival times of subjects are to be sampled from a uniform distribution or from an exponential distribution under the Poisson process. Let us use the Poisson accrual model as it is known to be a more realistic representation of the subject arrival process. Furthermore, let us specify the enrollment plan in terms of **Sites**; when we specify in terms of **Sites by Region** it is assumed that all sites within a region have the same parameters, which is not the case

in our enrollment plan.

Accrual Model: ▾

Sites By Regions Sites

Enrollment Plan Enter the parameters of the RALES enrollment plan in the grid manually. Alternatively, create a spreadsheet such as the one shown below and save it as a CSV file, import it using the menu item **Home-->Import** to add it as a node with the extension .cydx, and then select it using the **Specify Enrollment Plan...** button.

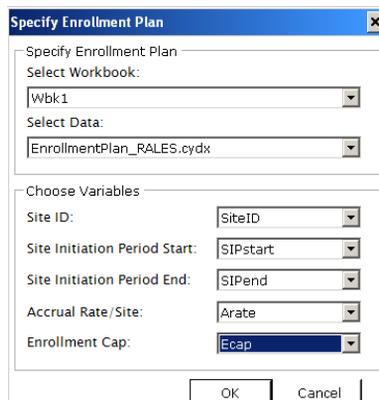
| | A | B | C | D | E |
|----|--------|----------|--------|-------|------|
| 1 | SiteID | SIPstart | SIPend | Arate | Ecap |
| 2 | SITE1 | 0 | 0 | 8 | 1000 |
| 3 | SITE2 | 0 | 1 | 8 | 1000 |
| 4 | SITE3 | 0 | 1 | 4 | 1000 |
| 5 | SITE4 | 0 | 1 | 4 | 1000 |
| 6 | SITE5 | 0 | 1 | 4 | 1000 |
| 7 | SITE6 | 0 | 1 | 4 | 1000 |
| 8 | SITE7 | 0 | 1 | 4 | 1000 |
| 9 | SITE8 | 0 | 1 | 4 | 1000 |
| 10 | SITE9 | 0 | 1 | 4 | 1000 |
| 11 | SITE10 | 0 | 1 | 4 | 1000 |
| 12 | SITE11 | 0 | 1 | 4 | 1000 |
| 13 | SITE12 | 0 | 1 | 4 | 1000 |
| 14 | SITE13 | 0 | 1 | 4 | 1000 |
| 15 | SITE14 | 0 | 1 | 4 | 1000 |
| 16 | SITE15 | 0 | 1 | 4 | 1000 |
| 17 | SITE16 | 0 | 1 | 4 | 1000 |
| 18 | SITE17 | 0 | 1 | 2 | 1000 |
| 19 | SITE18 | 0 | 1 | 2 | 1000 |
| 20 | SITE19 | 0 | 1 | 2 | 1000 |
| 21 | SITE20 | 0 | 1 | 2 | 1000 |

For your convenience this CSV file is already created and stored in the **Samples** subfolder in your East installation folder, under the name **EnrollmentPlan RALES.csv**. In this CSV file, ‘SiteID’ corresponds to the site name, ‘SIPstart’ refers to the Site Initiation Start and ‘SIPend’ to Site Initiation End. ‘Arate’ and ‘Ecap’ refer to the site accrual rate and enrollment cap respectively.

Click the **Specify Enrollment Plan...** button to load the file into the enrollment plan grid using **Browse**:

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

Ensure that the header names in your CSV file match the column names indicated in the **Specify Enrollment Plan** window by selecting the corresponding variable names in the dropdown boxes:

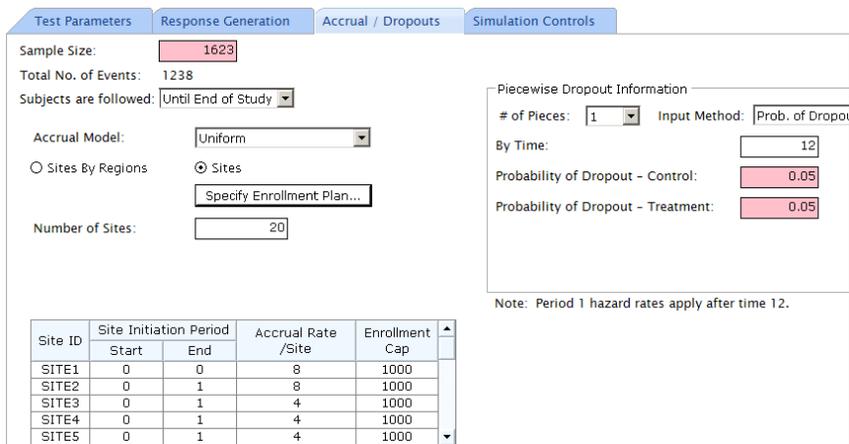


The 'Specify Enrollment Plan' dialog box contains the following fields:

- Select Workbook:** Wbk 1
- Select Data:** EnrollmentPlan_RALES.cydx
- Choose Variables:**
 - Site ID: SiteID
 - Site Initiation Period Start: SIPstart
 - Site Initiation Period End: SIPend
 - Accrual Rate / Site: Arate
 - Enrollment Cap: Ecap

Buttons: OK, Cancel

Click **OK**. When the final **Accrual/Dropouts** tab appears as displayed below we can set the number of simulations to 1000 in the **Simulation Control** tab and then simulate the design by clicking **Simulate**.



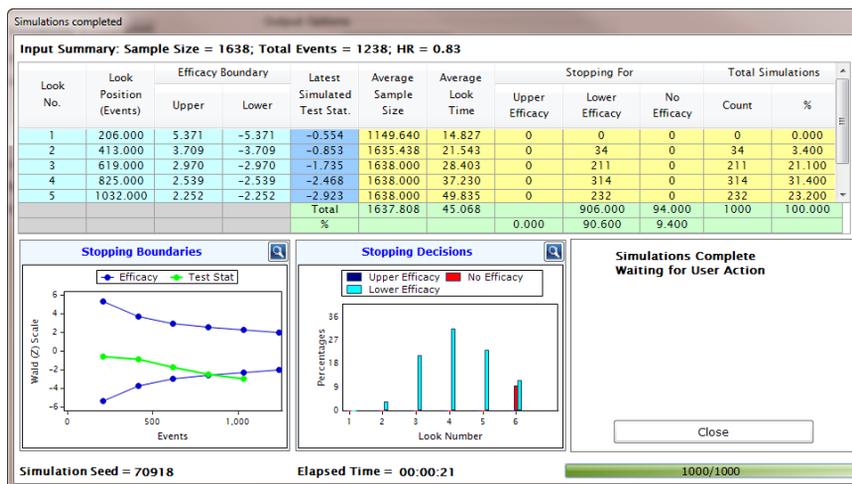
The 'Simulation Control' dialog box shows the following settings:

- Sample Size:** 1623
- Total No. of Events:** 1238
- Subjects are followed:** Until End of Study
- Accrual Model:** Uniform
- Sites By Regions:** Sites
- Number of Sites:** 20
- Piecewise Dropout Information:**
 - # of Pieces: 1
 - Input Method: Prob. of Dropout
 - By Time: 12
 - Probability of Dropout - Control: 0.05
 - Probability of Dropout - Treatment: 0.05

Note: Period 1 hazard rates apply after time 12.

| Site ID | Site Initiation Period Start | Site Initiation Period End | Accrual Rate /Site | Enrollment Cap |
|---------|------------------------------|----------------------------|--------------------|----------------|
| SITE1 | 0 | 0 | 8 | 1000 |
| SITE2 | 0 | 1 | 8 | 1000 |
| SITE3 | 0 | 1 | 4 | 1000 |
| SITE4 | 0 | 1 | 4 | 1000 |
| SITES | 0 | 1 | 4 | 1000 |

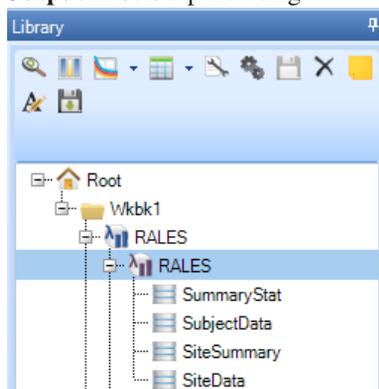
East displays the following window as it carries out the simulation runs:



Once the specified number of simulations has been run we can close the simulating design window and see a one-line summary of the output in the **Output Preview** pane:

| ID | Design Type | Test Type | Sample Size | Maximum Events | Power | No. of Looks | Spacing of Looks | Efficacy Boundary | No. of Hazard Pieces | Test Statistic | Followup Duration | Average Study Duration | Average Sample Size |
|------|-------------|-----------|-------------|----------------|-------|--------------|------------------|-------------------|----------------------|----------------|--------------------|------------------------|---------------------|
| Sim1 | Superiority | 2-Sided | 1638 | 1238 | 0.906 | 6 | User Specified | User Specified | 1 | Logrank | Until End of Study | 45.068 | 1637.808 |

Save the output from the **Output Preview** pane using



Note that East uses the blue icon  to denote designs and the brown icon  to denote simulations.

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

66.3.3 Output

Double-click on the 'RALES' simulation node in the **Library** pane to open the output summary. Here we can see data such as the estimations of the average sample size, number of events and dropouts at each look. In the table **Simulation Boundaries and Boundary Crossing Probabilities** we observe that by the end of the trial in 906 out of 1000 simulations we are able to reject the null hypothesis that the hazard rates of the treatment and control group are equal. In other words, Aldactone was effective in reducing the mortality rate by 17% as hypothesized.

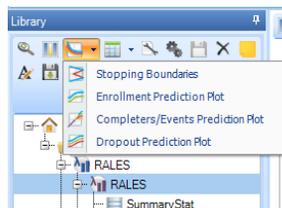
Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 1149.64 | 111.552 | 94.448 | 14.889 | 15.492 | 14.827 | 6.167 |
| 2 | 1635.438 | 222.574 | 190.426 | 29.925 | 30.994 | 21.543 | 8.71 |
| 3 | 1638 | 331.055 | 287.945 | 44.622 | 46.509 | 28.403 | 13.061 |
| 4 | 1638 | 435.103 | 389.897 | 59.155 | 62.301 | 37.23 | 17.394 |
| 5 | 1638 | 536.558 | 495.442 | 73.689 | 78.277 | 49.835 | 21.729 |
| 6 | 1638 | 634.22 | 603.78 | 88.234 | 93.081 | 71.911 | 26.032 |
| Average | 1637.808 | 477.906 | 423.961 | 64.182 | 68.64 | 45.068 | 19.024 |

Simulation Boundaries and Boundary Crossing Probabilities

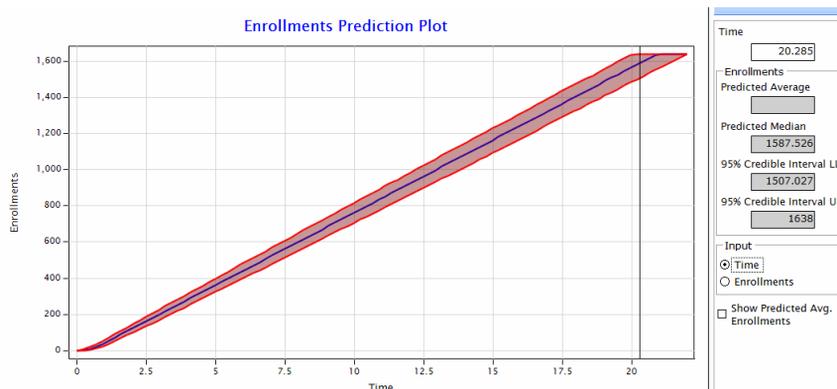
| Look # | Events | Boundaries | | Stopping For | | Total Simulations | |
|--------|--------|------------|--------|----------------|----------------|-------------------|---------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 1 | 206 | 5.371 | -5.371 | 0 | 0 | 0 | 0.000% |
| 2 | 413 | 3.709 | -3.709 | 0 | 34 | 34 | 3.400% |
| 3 | 619 | 2.97 | -2.97 | 0 | 211 | 211 | 21.100% |
| 4 | 825 | 2.539 | -2.539 | 0 | 314 | 314 | 31.400% |
| 5 | 1032 | 2.252 | -2.252 | 0 | 232 | 232 | 23.200% |
| 6 | 1238 | 2.045 | -2.045 | 0 | 115 | 209 | 20.900% |
| Total | | | | 0 | 906 | 1000 | |
| % | | | | 0.000% | 90.600% | | |

Click on the  button in the **Library** pane and select **Enrollment Prediction Plot**.



The **Enrollment Prediction Plot** displays the cumulative enrollments over time. It shows the predicted median and average enrollments along with the 95% confidence

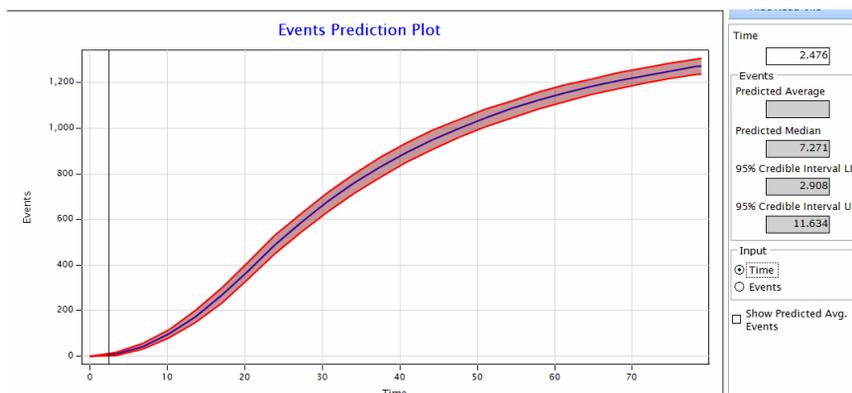
interval over all simulations.



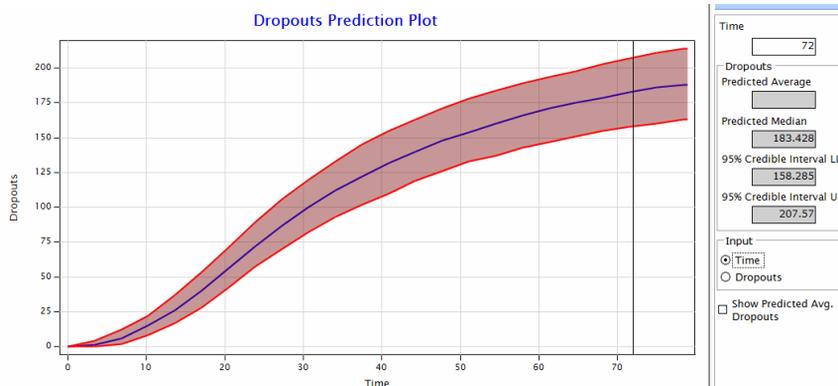
From our simulation of the RALES trial, it is expected that the full sample size will be enrolled earliest by about 20 months and latest by about 22 months. Furthermore, the confidence interval band is fairly narrow; indicating that there is not expected to be a great degree of variation in the predicted enrollment.

In the **Events Prediction Plot** we can observe the timeline of the events throughout the study period of around 72 months and beyond, while the **Enrollment Prediction Plot** only covered the accrual duration of about 20 months. From the graph, we may conclude that it is likely that the study will take the estimated length of a median of about 73 months and a maximum of 79 months.

66 Enrollment/Events Prediction - At Design Stage (By Simulation)



Lastly, the **Dropouts Prediction Plot** shows the progression of dropouts over the study duration. The predicted median dropouts by the end of the study period of 72 months is about 182, with the 95% confidence interval spanning a range of 158 to 208.



Lastly, four output files nested below the ‘RALES’ simulation node in the **Library** pane contain the full details of all the simulation runs. These files, named **SummaryStat**, **SubjectData**, **SiteSummary**, and **SiteData**, are the source of the data displayed in the tables and plots described above.

SummaryStat contains the look-wise details of each of the 1000 simulation runs including the number of accruals, completers, dropouts, look times, average follow-up

times and so on.

| | ScenarioID | SimIndex | LookIndex | Status | BdryStopCod | Accruals0 | DropOuts0 | Pendings0 | Events0 | Accruals1 |
|----|------------|----------|-----------|---------|-------------|-----------|-----------|-----------|---------|-----------|
| 1 | 1 | 1 | 1 | SUCCESS | 0 | 569 | 14 | 435 | 120 | 538 |
| 2 | 1 | 1 | 2 | SUCCESS | 0 | 819 | 34 | 560 | 225 | 819 |
| 3 | 1 | 1 | 3 | SUCCESS | 0 | 819 | 50 | 438 | 331 | 819 |
| 4 | 1 | 1 | 4 | SUCCESS | 1 | 819 | 70 | 313 | 436 | 819 |
| 5 | 1 | 2 | 1 | SUCCESS | 0 | 592 | 16 | 465 | 111 | 589 |
| 6 | 1 | 2 | 2 | SUCCESS | 0 | 819 | 33 | 561 | 225 | 819 |
| 7 | 1 | 2 | 3 | SUCCESS | 0 | 819 | 47 | 445 | 327 | 819 |
| 8 | 1 | 2 | 4 | SUCCESS | 1 | 819 | 56 | 325 | 438 | 819 |
| 9 | 1 | 3 | 1 | SUCCESS | 0 | 617 | 13 | 491 | 113 | 624 |
| 10 | 1 | 3 | 2 | SUCCESS | 0 | 819 | 29 | 573 | 217 | 819 |

The **SubjectData** sheet displays the following data corresponding to each subject:

- ScenarioID: it is possible to simulate a design under different scenarios by entering multiple parameter values in certain fields (refer to Section 3.7). East then assigns an identification number to each scenario. In this example we simulated a single design without varying the parameters and so the ScenarioID is always '1'.
- SimulationID: the identification number of the simulation.
- PatientID: a unique identification number assigned to each subject.
- SiteID: the identification number of the site at which the subject arrived.
- ArrivalTime: the time at which the subject arrived.
- TreatmentID: the type of treatment the patient received.
- SurvivalTime: the observed survival time of the subject over the course of the study duration
- DropOutTime: the time at which the subject dropped out of the study.
- CensorInd: this variable corresponds to censoring information. '1' represents completers and '0' represents dropouts and subjects in the pipeline.

| | ScenarioID | SimulationID | SubjectID | SiteID | ArrivalTime | TreatmentID | SurvivalTime | DropOutTime | CensorInd |
|----|------------|--------------|-----------|--------|-------------|-------------|--------------|-------------|-----------|
| 1 | 1 | 2 | 1 | SITE1 | 0.13760122 | 0 | 9.41891265 | 714.631726 | 1 |
| 2 | 1 | 2 | 2 | SITE4 | 0.186764959 | 0 | 0.0820291139 | 246.059898 | 1 |
| 3 | 1 | 2 | 3 | SITE8 | 0.320618213 | 1 | 190.372803 | 216.397981 | 0 |
| 4 | 1 | 2 | 4 | SITE4 | 0.380756242 | 0 | 14.6001508 | 4.26732488 | 0 |
| 5 | 1 | 2 | 5 | SITE2 | 0.407131713 | 1 | 146.593565 | 54.4256413 | 0 |
| 6 | 1 | 2 | 6 | SITE10 | 0.40928505 | 0 | 36.3955346 | 15.1181987 | 0 |
| 7 | 1 | 2 | 7 | SITE1 | 0.427330167 | 0 | 20.1346786 | 8.87775966 | 0 |
| 8 | 1 | 2 | 8 | SITE12 | 0.467855551 | 0 | 18.0266611 | 443.918518 | 1 |
| 9 | 1 | 2 | 9 | SITE4 | 0.46983433 | 0 | 1.84320828 | 411.790456 | 1 |
| 10 | 1 | 2 | 10 | SITE9 | 0.484827099 | 1 | 105.510231 | 172.044815 | 0 |

SiteSummary contains the site-level data:

- SiteID: the identification number of the site.
- RegionID (if applicable): the ID of the region to which the site belongs. In this

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

example there is no RegionID column because we chose to define the enrollment plan by individual sites.

- AvgInitiationTime: the average initiation time of this site over all simulations in which this site was opened.
- AvgLastSubjectArrTime: the average time at which the last subject was enrolled at this site over all simulations in which this site was opened.
- AvgNumOfSubj: the average number of subjects enrolled at this site over all simulations in which this site was opened.
- AvgAccrualDuration: the average of the accrual duration for the site computed in every simulation in which the site was opened. The accrual duration is calculated as the last subject randomization time - site initiation time of that site.
- AvgAccrualRate: the average of the observed accrual rate computed in each simulation in which the site was opened.
- SiteOpenedSimCount: the number of simulations in which the site was opened.

| | SiteID | AvgInitiation | AvgLastSubj | AvgNumOfSu | AvgAccrualD | AvgAccrualR | SiteOpenedS |
|----|--------|---------------|-------------|------------|-------------|-------------|-------------|
| 1 | SITE1 | 0 | 20.8234621 | 167.702 | 20.9368551 | 8.01436138 | 1000 |
| 2 | SITE2 | 0.502006848 | 20.8239329 | 163.458 | 20.4348482 | 8.00508055 | 1000 |
| 3 | SITE3 | 0.500614453 | 20.7031027 | 81.885 | 20.4362406 | 4.00954025 | 1000 |
| 4 | SITE4 | 0.512975231 | 20.7066329 | 80.926 | 20.4238798 | 3.96492613 | 1000 |
| 5 | SITE5 | 0.493601696 | 20.6904496 | 81.768 | 20.4432534 | 4.0025906 | 1000 |
| 6 | SITE6 | 0.504012862 | 20.6989217 | 81.626 | 20.4328422 | 3.99760218 | 1000 |
| 7 | SITE7 | 0.492161186 | 20.6963453 | 81.855 | 20.4446939 | 4.00579845 | 1000 |
| 8 | SITE8 | 0.50711503 | 20.6996943 | 81.99 | 20.42974 | 4.01551697 | 1000 |
| 9 | SITE9 | 0.504618226 | 20.7073291 | 81.749 | 20.4322368 | 4.00287669 | 1000 |
| 10 | SITE10 | 0.505181138 | 20.6899161 | 81.533 | 20.4316739 | 3.99293502 | 1000 |

The final output file, **SiteData**, contains the following information for each site:

- SimulationID: the identification number of the simulation.
- SiteOpenFlag: indicates whether the site has been initiated. The flag is '1' if the site has been initiated and '0' if it has not.
- SiteID: the identification number of the site.
- RegionID (if applicable): the ID of the region to which the site belongs. In this example there is no RegionID column because we chose to define the enrollment plan by individual sites.
- SiteReadyTime: the site initiation time generated as part of the simulations.
- SiteAccrRate: the site accrual rate specified in the enrollment plan.
- SubjectsAccrued: the number of subjects accrued at the site.
- LastSubjectRand: the randomization time of the last subject arriving at the site.
- AccrualDuration: if SiteOpenFlag = 1 for the ith site the accrual duration is computed as follows: AccrualDuration = maximum of the LastSubjRand times

across all sites - SiteReadyTime of the ith site. If SiteOpenFlag = 0 for the site then the AccrualDuration field will be blank.

- ObsrvdAccrualRate: the observed accrual rate for the site. It is computed as follows: $\text{ObsrvdAccrualRate} = \text{SubjectsAccrued} / \text{AccrualDuration}$.

| | SimulationID | SiteOpenFla | SiteID | SiteInitiation | SiteAccrRate | SubjectsAccr | LastSubjectR | AccrualDurat | ObsrvdAccru |
|----|--------------|-------------|--------|----------------|--------------|--------------|--------------|--------------|-------------|
| 1 | 2 | 1 | SITE1 | 0 | 8 | 147 | 20.3582609 | 20.7573887 | 7.08181563 |
| 2 | 2 | 1 | SITE2 | 0.29579935 | 8 | 168 | 20.5090119 | 20.4615894 | 8.21050588 |
| 3 | 2 | 1 | SITE3 | 0.95491147 | 4 | 73 | 20.7521585 | 19.8024773 | 3.68640746 |
| 4 | 2 | 1 | SITE4 | 0.116551483 | 4 | 92 | 20.7176 | 20.6408372 | 4.45718354 |
| 5 | 2 | 1 | SITE5 | 0.560744428 | 4 | 78 | 20.7572397 | 20.1966443 | 3.86202771 |
| 6 | 2 | 1 | SITE6 | 0.61768908 | 4 | 74 | 20.6467691 | 20.1396997 | 3.67433484 |
| 7 | 2 | 1 | SITE7 | 0.701100702 | 4 | 74 | 20.1872397 | 20.056288 | 3.68961594 |
| 8 | 2 | 1 | SITE8 | 0.0874838217 | 4 | 93 | 20.754555 | 20.6699049 | 4.49929501 |
| 9 | 2 | 1 | SITE9 | 0.297379938 | 4 | 85 | 20.6148794 | 20.4600088 | 4.15444592 |
| 10 | 2 | 1 | SITE10 | 0.181911783 | 4 | 83 | 20.383544 | 20.5754769 | 4.03392836 |

66.4 Survival Design-Example 2

The final example is based on the ONCOX time to event trial. The aim of this trial was to compare survival in two groups: a treatment group receiving a new drug for cancer, and a control group. As in the previous examples, we extend the ONCOX simulation to incorporate accrual information and study the enrollment and events prediction.

66.4.1 The ONCOX Trial: Initial Design

The randomized ONCOX study was a 30 months double blind efficacy and futility trial comparing survival rates of a treatment group and a control group with one interim look. The control group was known to have a median survival period of 5 months and the aim of ONCOX was to ascertain with a power of 0.9 that the median survival in the treatment group would be a longer period of 7 months. The study was a one-sided test with $\alpha = 0.025$ and an expected annualized dropout rate of 4% in both the groups. The efficacy and futility boundaries were to be based on spending function of $\gamma(-5)$. Subjects were enrolled over a period of 24 months. The sample size was fixed to be 460.

Let us implement this design in East. Click on the **Two Sample** button in the Survival category on the Design ribbon and select **Logrank Test Given Accrual Duration and Study Duration**.

This opens the survival design dialog box with default values.

Enter the following design parameters of the ONCOX trial in the corresponding fields Design Parameters tab:

- Design Type: Superiority

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

- Number of Looks: 2
- Test Type: 1-Sided
- Type-1 Error: 0.025
- Sample Size: 460
- Power: (to be computed)
- No.of Events (to be computed)
- # of Hazard Pieces: 1
- Median Survival Time
- Input Method: Median Survival Time (Control): 5
- Input Method: Median Survival Time (Treatment): 7
- Allocation Ratio: 1

The dialog box will appear as shown below.

Design Type: Superiority Number of Looks: 2

Test Parameters Boundary Accrual / Dropouts

Test Type: 1-Sided # of Hazard Pieces: 1 Input Method: Median Survival Times

Type I Error (α): 0.025

Power: Computed

Sample Size (n): 460

No. of Events: Computed

Allocation Ratio: 1 (n_1/n_2)

Hazard Ratio (Optional) Alternative

Hazard Ratio (λ_1/λ_2) 0.714

Ratio of Medians (m_1/m_2) 1.4

| Med.Surv.Time | |
|---------------|-----------------|
| Control | Treatment: Alt. |
| 5 | 7 |

Variance of Log Hazard Ratio

Null Alternative

Assurance (Probability of Success)

Notice that the hazard ratio is computed to 0.714.

Next, in the **Boundary** tab we specify the details for the **Efficacy** boundary, **Futility** boundary, and the spacing of the interim looks. As indicated in the beginning of this chapter, we modify the spending function from the default Lan DeMets (OF) to Gamma family(-5). Set the spacing of looks as **Equal**. Futility boundary Non-binding Gamma family with parameter -5 is chosen.

Now the **Test Parameters** tab will appear as shown below.

Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

Design Type: Superiority Number of Looks: 2

Test Parameters Boundary Accrual / Dropouts

Efficacy
Boundary Family: Spending Functions
Spending Function: Gamma Family
Parameter (γ): -5
Type I Error (α): 0.025

Futility
Boundary Family: Spending Functions
Spending Function: Gamma Family
Parameter (γ): -5
Type II Error (β): 0.101

Info. Fraction at Interim Look: 0.5 Boundary Scale: Z Scale

| Look # | Info. Fraction | Stop for Efficacy | Stop for Futility | Cum. α Spent | Efficacy Boundary | Cum. β Spent | Futility Boundary |
|--------|----------------|-------------------------------------|-------------------------------------|--------------|-------------------|--------------|-------------------|
| 1 | 0.500 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.002 | -2.895 | 0.008 | 0.125 |
| 2 | 1.000 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 0.025 | -1.972 | 0.100 | -1.972 |

In the final tab, we can enter the accrual duration (24 months), study duration (30 months) and dropout information (prob.of dropout as 0.04 in a 12 month survival period). Now the **Accrual/Dropouts** tab should appear as follows:

Test Parameters Boundary **Accrual / Dropouts**

Subjects are followed: Until End of Study

Accrual Info
Accrual Duration: 24 Study Duration: 30
of Accrual Periods: 1

| Period # | By Time | Cum. % Accrued |
|----------|---------|----------------|
| 1 | 24.000 | 100.000 |

Piecewise Dropout Information
of Pieces: 1 Input Method: Prob. of Dropo
By Time: 12
Prob. of Dropout (Control): 0.04
Prob. of Dropout (Treatment): 0.04

Note: Period 1 hazard rates apply after time 12.

Click **Compute** to determine the required events and the power attained for this trial.

| ID | Design Type | No. of Looks | Test Type | Specified α | Attained α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Futility Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events |
|-------|-------------|--------------|-----------|-------------|------------|-------|-------|------------------|-------------------|-------------------|-------------|------------------|------------------|----------------|
| ONCOX | Superiority | 2 | 1-Sided | 0.025 | 0.025 | 0.899 | 1 | Equal | Gm (-5) | Gm (-5) (NB) | 460 | 393.693 | 423.62 | 374 |

Rename this design 'ONCOX' using and then save it in the library using . Let us simulate this trial to study its enrollment process.

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

66.4.2 Simulating the ONCOX Trial

The primary input in the simulation is the enrollment plan which contains the following information for each site:

- Site initiation period: the time period over which the site is expected to be initialized so that it is ready to begin enrolling subjects
- Site accrual rate: the number of subjects expected to arrive at the site over the unit of time chosen (in this case, ‘month’)
- Enrollment cap: the maximum number of subjects that may be enrolled for the entire country as well for each site within the country. Thus, one site in a country can enroll number of subjects equal to the cap, provided all other sites in the country enroll none.

The table below shows a sample enrollment plan for the ONCOX trial.

| Country | No.ofSites | Site Initiation | | Accrual Rate | Enrollment Cap |
|------------|------------|-----------------|-------|--------------|----------------|
| | | (Start) | (End) | | |
| Austria | 4 | 3 | 6 | 0.3835 | 64 |
| Belgium | 6 | 5 | 6 | 0.3185 | 72 |
| CZ | 3 | 4 | 6 | 0.247 | 36 |
| France | 9 | 4 | 5 | 0.4485 | 168 |
| Germany | 10 | 4 | 5 | 0.3965 | 164 |
| Hungary | 4 | 6 | 8 | 0.1755 | 28 |
| ITALY | 7 | 5 | 7 | 0.2665 | 73 |
| Poland | 4 | 4 | 5 | 0.481 | 65 |
| Spain | 7 | 4 | 6 | 0.1755 | 47 |
| UK | 8 | 4 | 6 | 0.2795 | 92 |
| Australia | 11 | 4 | 6 | 0.299 | 181 |
| NewZealand | 4 | 5 | 6 | 0.2535 | 55 |
| Canada | 5 | 6 | 7 | 0.234 | 64 |
| US | 39 | 0 | 11 | 0.2535 | 544 |

We see from this enrollment plan that there are 14 countries each with different number of sites, participating in the study and each site may enroll a maximum of the number of subjects specified as ‘Enrollment Cap’. Sites in US initiates immediately and the remaining sites in remaining 13 countries must initiate within a maximum of 8 months of the start of the study. The accrual rates are given in terms of subjects arriving per site per month. We shall use this enrollment plan in our simulation of the ONCOX trial.

Select ‘ONCOX’ design in the **Library** pane and click  to open the simulating design window containing the tabs **Test Parameters**, **Response Generation**, **Accrual/Dropouts**, and **Simulation Controls**. The first three tabs contain the trial details we had entered in the initial design phase. The **Simulation Controls** tab is

where we specify the number of runs.

| Test Parameters | Response Generation | Accrual/Dropout | Simulation Controls | |
|----------------------|---------------------|---------------------|---------------------|------------|
| Trial Type: | Superiority | | | |
| Test Type: | 1-Sided | | | |
| Fix at Each Look: | Total No. of Events | | | |
| Total No. of Events: | 374 | | | |
| Look # | Info. Fraction | Cum. α Spent | Efficacy Z | Futility Z |
| 1 | 0.500 | 0.002 | -2.895 | 0.121 |
| 2 | 1.000 | 0.025 | -1.972 | -1.972 |

Click on the **Include Options** box to add **Site Info**:

Include Options

- Site
- Randomization
- User Defined R Function
- Stratification
- Sample Size Re-estimation

The main inputs we must provide in the **Accrual/Dropouts** tab are the accrual model and the enrollment plan.

Accrual Model We have the choice to specify whether the arrival times of subjects are to be sampled from a uniform distribution or from an exponential distribution under the Poisson process. Let us use the Poisson accrual model as it is known to be a more realistic representation of the subject arrival process. Furthermore, let us specify the enrollment plan in terms of **Sites by Regions**; when we specify in terms of **Sites by**

66 Enrollment/Events Prediction - At Design Stage (By Simulation)

Region it is assumed that all sites within a region have the same parameters.

Accrual Model:

Sites By Regions Sites

Enrollment Plan Enter the parameters of the ONCOX enrollment plan in the grid manually. Alternatively, create a spreadsheet such as the one shown below and save it as a CSV file so that it can be imported using the menu item **Home-->Import** to appear as a node with extension .cydx.

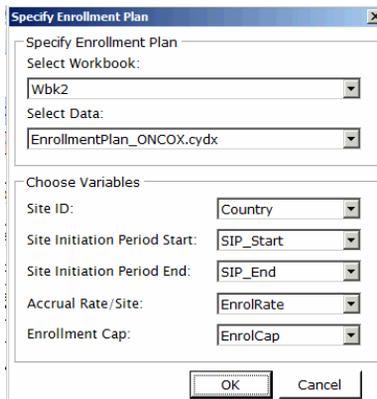
| | A | B | C | D | E | F |
|----|-----------|--------|-----------|---------|-----------|----------|
| 1 | Country | Nsites | SIP_Start | SIP_End | EnrolRate | EnrolCap |
| 2 | Austria | 4 | 3 | 6 | 0.3835 | 64 |
| 3 | Belgium | 6 | 5 | 6 | 0.3185 | 72 |
| 4 | CZ | 3 | 4 | 6 | 0.247 | 36 |
| 5 | France | 9 | 4 | 5 | 0.4485 | 168 |
| 6 | Germany | 10 | 4 | 5 | 0.3965 | 164 |
| 7 | Hungary | 4 | 6 | 8 | 0.1755 | 28 |
| 8 | ITALY | 7 | 5 | 7 | 0.2665 | 73 |
| 9 | Poland | 4 | 4 | 5 | 0.481 | 65 |
| 10 | Spain | 7 | 4 | 6 | 0.1755 | 47 |
| 11 | UK | 8 | 4 | 6 | 0.2795 | 92 |
| 12 | Australia | 11 | 4 | 6 | 0.299 | 181 |
| 13 | NewZeala | 4 | 5 | 6 | 0.2535 | 55 |
| 14 | Canada | 5 | 6 | 7 | 0.234 | 64 |
| 15 | US | 39 | 0 | 11 | 0.2535 | 544 |

For your convenience this CSV file is already created and stored in the **Samples** subfolder in your East installation folder, under the name **EnrollmentPlan.ONCOX.csv**. In this CSV file, the column titles are self explanatory.

Click the **Specify Enrollment Plan...** button to specify the .cydx file and get it into the enrollment plan grid.

Ensure that the header names in your CSV file which is now a .cydx file, match the column names indicated in the **Specify Enrollment Plan** dialog box by selecting the

corresponding variable names in the dropdown boxes:



Click **OK**. When the final **Accrual/Dropouts** tab appears as displayed below we can set the number of simulations to 1000 in the **Simulation Controls** tab. Fix the seed at 12345 and then simulate the design by clicking **Simulate**.

Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

Number of Looks: 2

Test Parameters | Response Generation | **Accrual / Dropouts** | Simulation Controls

Sample Size: 460
 Total No. of Events: 374
 Subjects are followed: Until End of Study

Accrual Model: Poisson
 Sites By Regions Sites
 Specify Enrollment Plan...
 Number of Sites: 14
 Include Site Risk as Stratum Variable

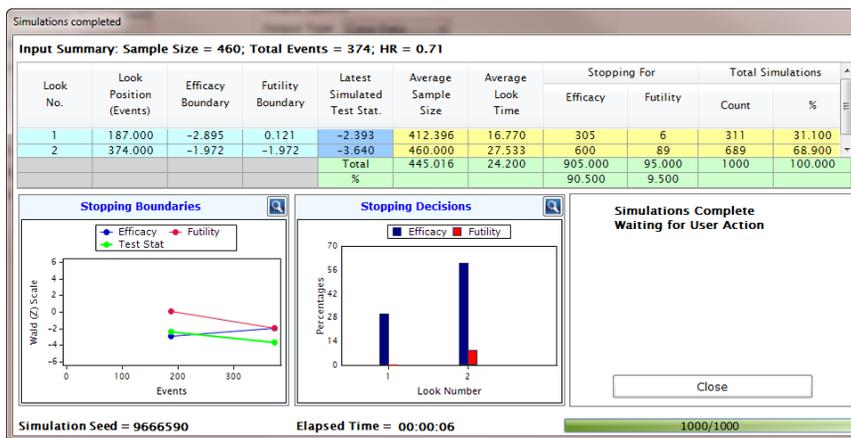
Piecewise Dropout Information
 # of Pieces: 1 Input Method: Prob. of Dropout
 By Time: 12
 Probability of Dropout - Control: 0.04
 Probability of Dropout - Treatment: 0.04

Note: Period 1 hazard rates apply after time 12.

| Site ID | Site Initiation Period Start | End | Accrual Rate /Site | Enrollment Cap |
|---------|------------------------------|-----|--------------------|----------------|
| Austria | 3 | 6 | 0.384 | 64 |
| Belgium | 5 | 6 | 0.319 | 72 |
| CZ | 4 | 6 | 0.247 | 36 |
| France | 4 | 5 | 0.449 | 168 |

East displays the following window as it carries out the simulation runs:

66 Enrollment/Events Prediction - At Design Stage (By Simulation)



Once the specified number of simulations has been run we can close the simulating design window and see a one-line summary of the output in the **Output Preview** pane:

| ID | Design Type | No. of Looks | Test Type | Power | Efficacy Boundary | Sample Size | Maximum Events | Futility Boundary | Spacing of Looks | No. of Hazard Pieces | Hazard Ratio | Test Statistic |
|------|-------------|--------------|-----------|-------|-------------------|-------------|----------------|-------------------|------------------|----------------------|--------------|----------------|
| Sim1 | Superiority | 2 | 1-Sided | 0.9 | User Specified | 460 | 374 | User Specified | User Specified | 1 | 0.714 | Logrank |

Save the simulation output from the **Output Preview** pane to the library. Note that East uses the blue icon  to denote designs and the brown icon  to denote simulations.

66.4.3 Output

Double-click on the 'ONCOX' simulation node in the **Library** pane to open the output summary. Here we can see data such as the estimations of the average sample size, number of events and dropouts at each look. In the table **Simulation Boundaries and Boundary Crossing Probabilities** we observe that by the end of the trial in 900 out of 1000 simulations we are able to reject the null hypothesis that the hazard rates of the

treatment and control group are equal, matching with the 90% power of the study.

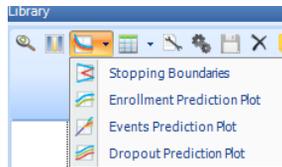
☰ Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 227.726 | 96.448 | 90.552 | 2.476 | 3.045 | 59.227 | 7.052 |
| 2 | 420.22 | 190.439 | 183.561 | 4.865 | 6.132 | 104.726 | 7.622 |
| Average | 366.101 | 164.277 | 156.989 | 4.163 | 5.279 | 91.919 | 7.49 |

☰ Simulation Boundaries and Incremental Boundary Crossing Probabilities

| Look # | Events | Boundaries | | Stopping For | | Total Simulations | |
|--------|--------|------------|----------|--------------|----------|-------------------|---------|
| | | Efficacy | Futility | Efficacy | Futility | Count | % |
| 1 | 187 | -2.895 | 0.125 | 271 | 11 | 282 | 28.200% |
| 2 | 374 | -1.972 | -1.972 | 629 | 89 | 718 | 71.800% |
| Total | | | | 900 | 100 | 1000 | |
| % | | | | 90.000% | 10.000% | | |

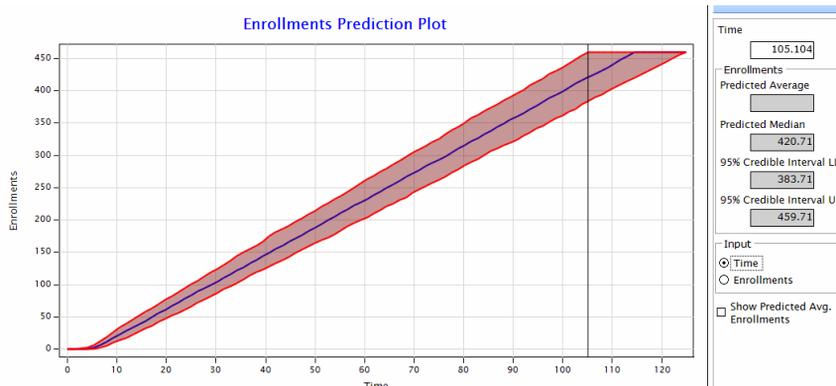
Click on the  button in the **Library** pane and select **Enrollment Prediction Plot**.



The **Enrollment Prediction Plot** displays the cumulative enrollments over time. It shows the predicted median and average enrollments along with the 95% confidence

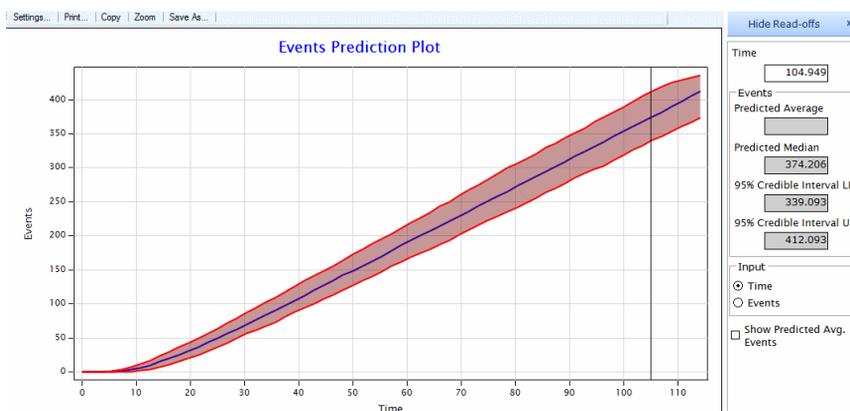
66 Enrollment/Events Prediction - At Design Stage (By Simulation)

interval over all simulations.



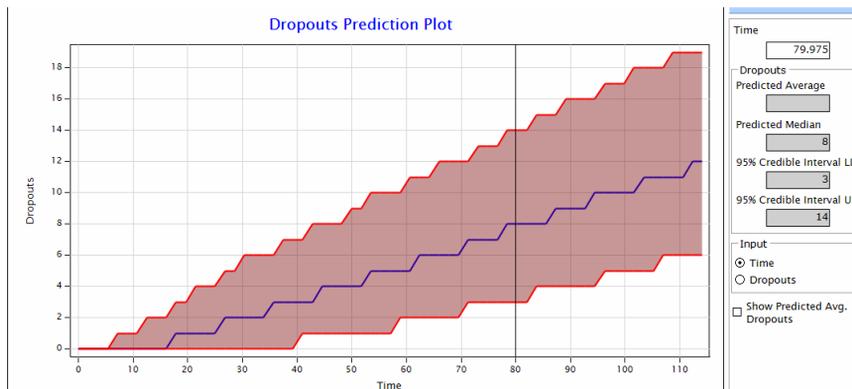
From our simulation of the ONCOX trial, it is expected that the full sample size will be enrolled earliest by about 105 months and latest by about 125 months.

In the **Events Prediction Plot**, we may observe that it is likely that the study will take the targeted median of 374 events in about 105 months and latest by 112 months with 95% confidence.



Lastly, the **Dropouts Prediction Plot** shows the progression of dropouts over the study

duration.



Lastly, four output files nested below the 'ONCOX' simulation node in the **Library** pane contain the full details of all the simulation runs. These files, named **SummaryStat**, **SubjectData**, **SiteSummary**, and **SiteData**, are the source of the data displayed in the tables and plots described above.

67 *Conditional Simulation*

During the design stage in the previous chapter we used simulation to explore the enrollment timeline and event prediction. The inputs of the design stage simulations were based on estimates of accrual rates and other parameters. In the case of Survival designs, once the trial begins and we obtain data on the realized enrollments, we can use the interim monitoring (IM) feature of EastPredict to update the parameters and generate new predictions about the enrollment and event timelines.

67.1 Survival Design- Example 1

67.1.1 Interim Data Preparation

67.1.2 Interim Analysis

67.1.3 Simulation

67.1.4 Output

The simulation of the RALES trial in the previous chapter indicated a required sample size of 1638 subjects with an expected accrual period of around 20 months. The total duration of the study was around 72 months. This example continues from the unconditional simulation performed in the previous chapter and assumes that the **Rales.cyx** workbook which is available in the **Samples** folder is open in East. In this section we perform a conditional simulation at the first interim look.

67.1.1 Interim Data Preparation

Data preparation for conditional simulation involves compiling the required data from various sources at a certain cut-off point as described below:

Subject Data Subject data refers to information collected about each subject accrued so far, namely:

- Arrival time: the time at which the subject arrived at the site.
- Censor information: whether the subject is a completer, a dropout or still in the pipeline.
- Treatment information: whether the subject was randomized to the treatment arm or the placebo arm.
- Survival information: the survival time of the subject.

For our example, we prepare the data on the basis of a simulated trial which was the output of our design time (unconditional) simulation.

The file **RALES_iLook1_SubjectData** contains a list of subjects accrued so far and the following data for each subject:

- **ArrivalTime**: the time at which the subject arrived.
- **TreatmentID**: a variable indicating which group the subject was randomized to ('1' for treatment, '0' for placebo).
- **TimeOnStudy**: the length of time the subject has been in the study, corresponding to survival time.

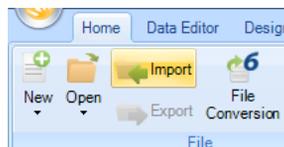
- CensorIndicator: a variable indicating whether the subject is a completer ('1'), a dropout ('-1') or in the pipeline ('0').
- CensorInd: a variable indicating whether the subject is a completer ('1') or a non-completer ('0'). A non-completer can be either a dropout or in the pipeline.

A portion of the RALES_iLook1_SubjectData file is shown below:

| | A | B | C | D | E | F | G |
|----|-----------|--------|------------|-----------|-----------------|----------|-----------|
| 1 | PatientID | SiteID | ArrivalTim | Treatment | CensorIndicator | TimeOnSt | CensorInd |
| 2 | 1 | SITE9 | 0.033264 | 1 | 0 | 15.52795 | 0 |
| 3 | 2 | SITE9 | 0.083005 | 1 | 0 | 15.47821 | 0 |
| 4 | 3 | SITE3 | 0.123138 | 0 | 1 | 7.790942 | 1 |
| 5 | 4 | SITE1 | 0.129578 | 0 | 0 | 15.43164 | 0 |
| 6 | 5 | SITE9 | 0.187238 | 0 | 1 | 4.249194 | 1 |
| 7 | 6 | SITE1 | 0.191141 | 1 | 1 | 0.05485 | 1 |
| 8 | 7 | SITE1 | 0.218275 | 1 | 0 | 15.34294 | 0 |
| 9 | 8 | SITE9 | 0.287513 | 1 | 1 | 2.748406 | 1 |
| 10 | 9 | SITE1 | 0.295866 | 1 | 1 | 10.83515 | 1 |

For your convenience the **Rales_iLook1_subjectdata.csv** has already been created and available in **Samples** subfolder in your East installation folder.

This data file can be imported into East using the **Import** button in the Home ribbon:



Once imported, the file will appear in the **Library** pane as a node in the active workbook, with extension **.cydx**.

67.1.2 Interim Analysis

Click on the node **RALES_iLook1_SubjectData.cydx** and choose the menu item **Analysis>Two Samples>Logrank**. In the resulting dialog box, select the

67 Conditional Simulation

variables as shown in the screen shots below.

Analysis: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank

Data Set: RALES_iLook1_SubjectData.cydx

Main Advanced

Trial Type: Superiority Response Variable: TimeOnStudy Frequency Variable:

Population ID: TreatmentID

Control: 0 Censor Indicator: CensorInd

Treatment: 1 Censored: 0 Complete: 1

Click **OK** to see the following output.

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | | Average Follow-up Time |
|--------------|-----------------|------------|----------------|------------|----------------|------------------------|
| | | Count | % | Count | % | |
| 0 | 610 | 116 | 19.016% | 494 | 80.984% | 6.161 |
| 1 | 595 | 90 | 15.126% | 505 | 84.874% | 6.588 |
| Total | 1205 | 206 | 17.095% | 999 | 82.905% | 6.372 |

Test of Hypothesis:

| Log Rank Score | Std. Error | Standardized Test Statistic | (1-Sided) | | (2-Sided) | |
|----------------|------------|-----------------------------|-----------|---------|-----------|--|
| | | | Tail | p-value | p-value | |
| -15.209 | 7.17 | -2.121 | L.E. | 0.017 | 0.034 | |

Parameter Estimates from Cox Model:

| Hazard Ratio (HR) | ln(Hazard Ratio) (δ) | Std. Error of Estimate of δ | Wald Statistic | 95% Confidence Interval of HR(2-Sided) | |
|-------------------|-------------------------------|------------------------------------|----------------|--|-------------|
| | | | | Lower Limit | Upper Limit |
| 0.743 | -0.297 | 0.141 | -2.113 | 0.564 | 0.979 |

Estimated Hazard Rates:

| | |
|--------------------------------|-------|
| Control (λ_c) | 0.031 |
| Treatment ($\lambda_c * HR$) | 0.023 |

We will use these output values for observed response frequencies to enter into the Test Statistic Calculator.

67.1.3 Simulation

To open the IM design window, select the 'RALES' design (represented by the blue

icon  in the **Library** pane and click . This opens the IM dashboard:

| Look # | Information Fraction | Cumulative Events | Test Statistic | δ | Standard Error | Efficacy | | 95% RCI for δ | | Repeat ... p-value | CP | Predicti ... Power |
|--------|----------------------|-------------------|----------------|----------|----------------|----------|-------|----------------------|-------|-----------------------|----|-----------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | |

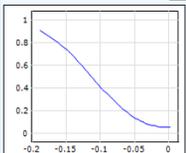
Click the "Enter Interim Data" button to enter the Look # 1 data.

Stopping Boundaries



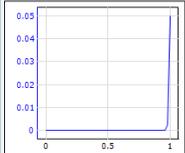
| Events | Efficacy Upper | Efficacy Lower |
|--------|----------------|----------------|
| | | |

Conditional Power



| Eff. Size | CP |
|-----------|-------|
| -0.188 | 0.904 |
| -0.169 | 0.833 |
| -0.145 | 0.714 |
| -0.122 | 0.564 |
| -0.099 | 0.404 |
| -0.076 | 0.259 |
| -0.052 | 0.148 |
| -0.029 | 0.08 |
| -0.006 | 0.051 |
| 0.017 | 0.05 |

Error Spending Function



| Info. Fraction | α |
|----------------|----------|
| | |

Confidence Intervals



| Info. Fraction | RCI Upper |
|----------------|-----------|
| | |

Click on the first blank row in the upper panel corresponding to Look #1 and click on the  button. This invokes the **Test Statistic Calculator** for recalculating the test statistic value based on the interim data. We saw from the interim analysis, the first look was taken at 206 events and the data was as follows: **Estimate of $\delta = \ln(0.743) = -0.29706$** and **Standard Error of Estimate of $\delta = \sqrt{4/206} = 0.139347$** . Enter these results in the relevant fields and then click on **Recalc** to obtain

67 Conditional Simulation

the updated test statistic:

The test statistic is updated to -2.132. After clicking **OK** the table in the dashboard is updated according to the new information:

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | | 95% RCI for HR | | Repeat... p-value | CP | Predicti... Power |
|--------|----------------------|-------------------|----------------|------------------|--------------------------------|----------|--------|----------------|-------|-------------------|----|-------------------|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.166 | 206 | -2.132 | -0.297 | 0.139 | 5.371 | -5.371 | 1.57 | 0.352 | 0.656 | 1 | 0.929 |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | |

The next step is to enter the observed data for the first look. In the IM Dashboard select the first row corresponding to Look #1 and click the **CS** button. This opens an input dialog window.

Specify Subject Info In this pane, use the drop-down menus next to **Select Workbook** and **Select Subject Data** to select the active workbook and the

RALES_iLook1_SubjectData file we imported earlier.

Next, in the **Choose Variables** tab, match the variables names to the corresponding headers in RALES_iLook1_SubjectData using the drop-down menus. In our example the matching would appear as follows:

- Population ID = TreatmentID
- Control = 0
- Treatment = 1
- Status Indicator = CensorIndicator
- Arrival Time = ArrivalTime
- Time on Study = TimeOnStudy

Estimate Parameters from Data...

Specify Subject Info

Select Workbook: RALES

Select Subject Data: RALES_iLook1_SubjectData.cydx

Choose Variables

Population ID: TreatmentID

Control: 0

Treatment: 1

Status Indicator: CensorIndicator

1=Complete
0=Censored
-1 = Dropout

Arrival Time: ArrivalTime

Response Variable: TimeOnStudy

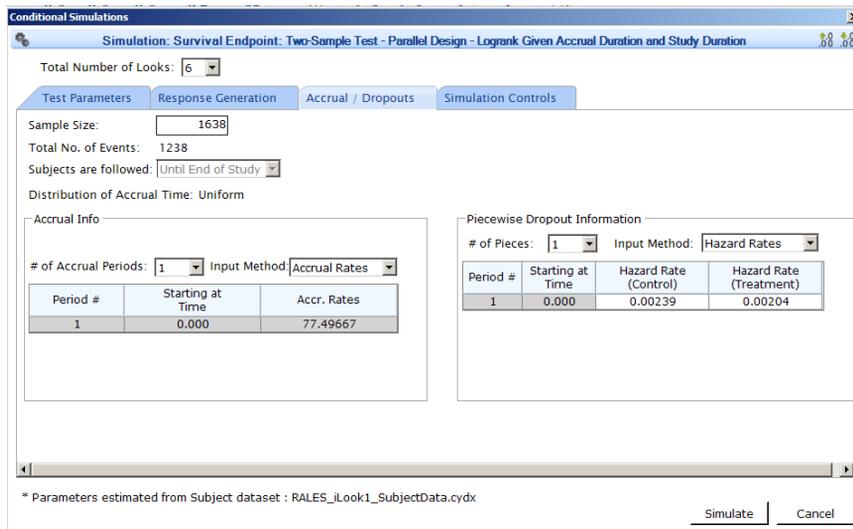
OK Cancel

Click on **OK** to obtain the input dialog window for conditional simulation:

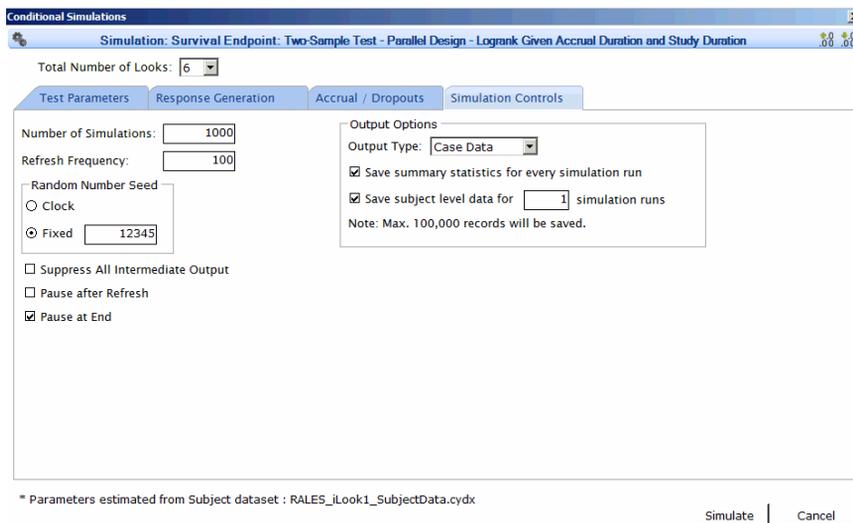
The simulation input dialog window consists of four tabs: **Test Parameters**, **Response Generation**, **Accrual/Dropouts**, and **Simulation Controls**. The first three tabs contain the parameters specified in the previous step. Navigate to the **Accrual/Dropouts** tab. Note that the parameters are estimated from the subject data

67 Conditional Simulation

RALES_iLook1_SubjectData.cydx



Lastly, in the **Simulation Controls** tab set the number of simulations to 1000, select the Fixed Random Seed 12345, check all the output options to save the data



Click **Simulate**. After the simulation is complete the results appear in the **Library** as a sub-node of the RALES design:

Under this sub-node, there is a snapshot of the initial interim data entered (**Snap 1.1**), an output summary of the conditional simulation (**CS:Sim1**) and the updated versions of the output files generated during the initial (unconditional) simulation.

67.1.4 Output

Double-click 'CS:Sim1' in the Library to open a detailed summary of the conditional simulation. The first and third tables, **Actuals: Sample Size and Look Times** and **Actuals: Events and Boundaries**, contain the interim data pertaining to the first look (1205 subjects accrued, 206 events out of which 116 were in the control group and 90 were in the treatment group) and the time of the interim look (15.561 months). The second table named **Conditional Simulation: Average Sample Size and Look Times** displays the projections of these parameters for the remaining five looks. From the fourth table, we can see the boundary crossing probabilities in the remaining five looks. For instance, by the 3rd look, 619 events have been observed and the efficacy boundary has been crossed in 671 out of 1000 simulations.

☰ **Actuals: Sample Size and Look Times**

| Look # | Sample Size | Events | | Dropouts | | Look Time | Average Follow up |
|--------|-------------|---------|-----------|----------|-----------|-----------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 1205 | 116 | 90 | 9 | 8 | 15.561 | 6.372 |

☰ **Conditional Simulation: Average Sample Size and Look Times**

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 2 | 1638 | 232.703 | 180.297 | 18.045 | 16.182 | 22.384 | 9.395 |
| 3 | 1638 | 345.403 | 273.597 | 26.875 | 24.493 | 29.535 | 14.09 |
| 4 | 1638 | 447.48 | 377.52 | 36.497 | 32.689 | 38.733 | 18.882 |
| 5 | 1638 | 547.85 | 484.15 | 44.1 | 39.8 | 51.124 | 23.648 |
| 6 | 1638 | 647 | 591 | 60 | 41 | 68.18 | 28.01 |
| Average | 1638 | 352.172 | 276.324 | 27.343 | 24.87 | 30.336 | 14.318 |

☰ **Actuals: Events and Boundaries**

| Look # | Events | Boundaries | |
|--------|--------|------------|--------|
| | | Efficacy | |
| | | Upper | Lower |
| 1 | 206 | 5.371 | -5.371 |

67 Conditional Simulation

Conditional Simulation: Simulation Boundaries and Boundary Crossing Probabilities (Incremental)

| Look # | Events | Boundaries | | Stopping For | | Total Simulations | |
|--------|--------|------------|--------|----------------|----------------|-------------------|---------|
| | | Efficacy | | Upper Efficacy | Lower Efficacy | Count | % |
| | | Upper | Lower | | | | |
| 2 | 413 | 3.709 | -3.709 | 0 | 152 | 152 | 15.200% |
| 3 | 619 | 2.97 | -2.97 | 0 | 671 | 671 | 67.100% |
| 4 | 825 | 2.539 | -2.539 | 0 | 157 | 157 | 15.700% |
| 5 | 1032 | 2.252 | -2.252 | 0 | 19 | 19 | 1.900% |
| 6 | 1238 | 2.045 | -2.045 | 0 | 1 | 1 | 0.100% |
| Total | | | | 0 | 1000 | 1000 | |
| % | | | | 0.000% | 100.000% | | |

Interim Data Information: Subject Data

Subject Dataset: RALES_iLook1_SubjectData.cyx
 Population ID: TreatmentID (Treatment=1, Control=0)
 Response Variable: TimeOnStudy
 Status Indicator: CensorIndicator
 Arrival Time: ArrivalTime

Simulation Seed and Elapsed Time

Starting Seed: 12345
 Total Number of Simulations: 1000

67.2 Survival Design-Example 2

The ONCOX trial in the previous chapter was designed with a sample size of 460 subjects with an expected accrual period of around 24 months and targeted 374 events within a study period of around 30 months. This example continues from the unconditional simulation performed in the previous chapter and assumes that the workbook **OncoX.cyx** is open in East. You may open it from the Samples folder. In this section we perform a conditional simulation at the first interim look.

67.2.1 Interim Data Preparation

Data preparation for conditional simulation involves compiling the required data from various sources at a certain cut-off point. The data required is **Subject Data** which consist of the following information.

Subject Data Subject data refers to information collected about each subject accrued so far, namely:

- Country: Country ID.
- Arrival time: the time at which the subject arrived at the site.
- Censor information: whether the subject is a completer, a dropout or still in the pipeline.
- Treatment information: whether the subject was randomized to the treatment arm or the control arm.

- Survival information: the survival time of the subject.

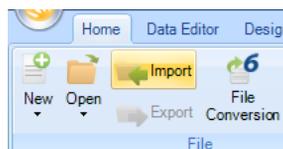
The file ONCOX_iLook1_SubjectData contains a list of subjects accrued so far and the following data for each subject:

- Country: Country ID.
- ArrivalTime: the time at which the subject arrived.
- TreatmentID: a variable indicating which group the subject was randomized to ('1' for treatment, '0' for placebo).
- TimeOnStudy: the length of time the subject has been in the study, corresponding to survival time.
- Status: a variable indicating whether the subject is a completer ('1'), a dropout ('-1') or in the pipeline ('0').
- Censor: a variable indicating whether the subject is a completer ('1') or a non-completer ('0'). A non-completer can be either a dropout or in the pipeline.

A portion of the ONCOX_iLook1_SubjectData file is shown below:

| | A | B | C | D | E | F | G |
|----|-----------|---------|------------|----------|--------|-------|--------|
| 1 | Country | Site_Id | ArrivalTim | TimeOnSt | Status | Trtmt | Censor |
| 2 | United_St | 103 | 0.493421 | 20.23026 | | 0 | 0 |
| 3 | United_St | 103 | 1.085526 | 8.322368 | | 1 | 1 |
| 4 | United_St | 108 | 1.513158 | 6.743421 | | 1 | 0 |
| 5 | United_St | 108 | 1.644737 | 3.75 | | 1 | 1 |
| 6 | United_St | 108 | 1.973684 | 5.953947 | | 1 | 0 |
| 7 | United_St | 103 | 2.5 | 4.934211 | | 1 | 0 |
| 8 | United_St | 103 | 2.565789 | 5 | | 1 | 0 |
| 9 | United_St | 104 | 2.894737 | 9.309211 | | 1 | 0 |
| 10 | United_St | 106 | 2.960526 | 14.70395 | | -1 | 0 |

The data file can be imported into East using the **Import** button in the Home ribbon:



Once imported, the file will appear in the **Library** pane as a node in the active workbook, with extension .cydx.

67.2.2 Simulation

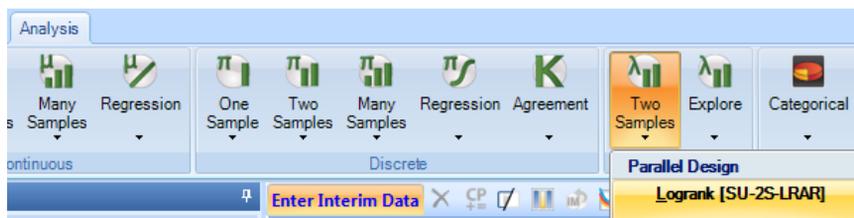
To open the IM design window, select the 'ONCOX' design (represented by the blue

67 Conditional Simulation

icon  in the **Library** pane and click . This opens the IM dashboard:

| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | Futility | 88.421% RCI for HR | | Repeat ... p-value | CP | Predicti... Power |
|--------|----------------------|-------------------|----------------|------------------|--------------------------------|----------|----------|--------------------|-------|--------------------|----|-------------------|
| | | | | | | | | Upper | Lower | | | |
| 1 | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | |

First we need to compute hazard ratio from the interim subject data. For this, click on 'ONCOX_iLook1_SubjectData.cydx' node in the library and then click on **Analysis** > **Two Samples** > **Parallel Design** > **Logrank** menu item.



In the resulting dialog box fill up items as shown below and click OK.

Data Set: ONCOX_iLook1_SubjectData.cydx

Main | Advanced

Superiority | Response Variable: TimeOnStudy | Frequency Variable:

Trtmt | Censor Indicator: Censor | Test Statistic: Logrank

0 | Censored: 0

1 | Complete: 1

Now you will get the following results.

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | |
|--------------|-----------------|--------|---------|----------|---------|
| | | Count | % | Count | % |
| 0 | 203 | 106 | 52.217% | 97 | 47.783% |
| 1 | 199 | 81 | 40.704% | 118 | 59.296% |
| Total | 402 | 187 | 46.517% | 215 | 53.483% |

Parameter Estimates:

| Hazard Ratio (HR) | 95% Confidence Interval(2-Sided) | |
|-------------------|----------------------------------|-------------|
| | Lower Limit | Upper Limit |
| 0.743 | 0.556 | 0.994 |

Test of Hypothesis:

| Log Rank Score | Std. Error | Standardized Test Statistic | (1-Sided) | | (2-Sided) |
|----------------|------------|-----------------------------|-----------|---------|-----------|
| | | | Tail | p-value | p-value |
| -13.782 | 6.819 | -2.021 | L.E. | 0.022 | 0.043 |

Estimated Hazard Rates:

| | |
|--------------------------------|-------|
| Control (λ_c) | 0.108 |
| Treatment ($\lambda_c * HR$) | 0.081 |

The number of events is 187 and the estimated hazard ratio is 0.743. Now go to IM dashboard and click on the first blank row in the upper panel corresponding to Look #1 and click on the **Enter Interim Data** button. This invokes the **Test Statistic Calculator** for recalculating the test statistic values based on the interim data. Enter the cumulative events as 187, Estimate of δ as $\ln(0.743)$, and Standard Error as

67 Conditional Simulation

$\sqrt{4/187}$). Click Recalculate. You will see the Test Statistic computed as -2.031.

Test Statistic Calculator

Editing Look #1

Set Current Look as Last

Read from Analysis Node

Select Workbook: []

Select Analysis Node: []

Cumulative Events: [187]

Input for Survival end point

Estimate of δ : [-0.297]

$\delta = \ln(\lambda_t / \lambda_c)$

Standard Error of Estimate of δ : [0.146]

Output

Test Statistic: [-2.031]

Recalc OK Cancel

Click OK and the results will get posted in the IM dashboard as shown below.

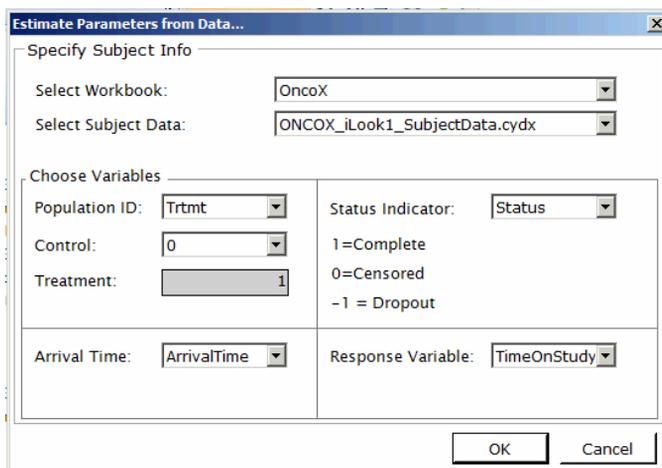
| Look # | Information Fraction | Cumulative Events | Test Statistic | Est. of δ | Std. Error of Est. of δ | Efficacy | Futility | 88.421% RCI for HR | | Repeat ... p-value | CP | Predicti... Power |
|--------|----------------------|-------------------|----------------|------------------|--------------------------------|----------|----------|--------------------|-------|--------------------|-------|-------------------|
| | | | | | | | | Upper | Lower | | | |
| 1 | 0.5 | 187 | -2.031 | -0.297 | 0.146 | -2.895 | 0.121 | 1.135 | 0.518 | 0.278 | 0.899 | 0.816 |
| 2 | | | | | | | | | | | | |

In the IM Dashboard select the first row corresponding to Look #1 and click the **CS** button. This opens an input dialog window.

Specify Subject Info In this pane, use the drop-down menus next to **Select Workbook** and **Select Subject Data** to select the active workbook and the ONCOX_iLook1_SubjectData file we imported earlier.

Next, in the **Choose Variables** tab, match the variables names to the corresponding headers in ONCOX_iLook1_SubjectData using the drop-down menus. In our example the matching would appear as follows:

- Population ID = TreatmentID
- Control = 0
- Treatment = 1
- Status Indicator = Status
- Arrival Time = ArrivalTime
- Time on Study = TimeOnStudy



Click on **OK** to obtain the input dialog window for conditional simulation:

The simulation input dialog window consists of four tabs: **Test Parameters**, **Response Generation**, **Accrual/Dropouts** and **Simulation Controls**. The first three tabs contain the parameters specified in the previous step. Navigate to the

67 Conditional Simulation

Accrual/Dropouts tab. Note that the parameters are estimated from the subject data.

Conditional Simulations
Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

Total Number of Looks: 2

Test Parameters | Response Generation | **Accrual / Dropouts** | Simulation Controls

Sample Size: 460
Total No. of Events: 374
Subjects are followed: Until End of Study
Distribution of Accrual Time: Uniform

Accrual Info

of Accrual Periods: 1 Input Method: Accrual Rates

| Period # | Starting at Time | Accr. Rates |
|----------|------------------|-------------|
| 1 | 0.000 | 19.39810 |

Piecewise Dropout Information

of Pieces: 1 Input Method: Hazard Rates

| Period # | Starting at Time | Hazard Rate (Control) | Hazard Rate (Treatment) |
|----------|------------------|-----------------------|-------------------------|
| 1 | 0.000 | 0.00102 | 0.00201 |

* Parameters estimated from Subject dataset : ONCOX_iLook1_SubjectData.cyx

Simulate Cancel

Lastly, in the **Simulation Controls** tab set the number of simulations to 1000, select the Fixed Random Seed 12345, check all the output options to save the data.

Conditional Simulations
Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration

Total Number of Looks: 2

Test Parameters | Response Generation | Accrual / Dropouts | **Simulation Controls**

Number of Simulations: 1000
Refresh Frequency: 100

Random Number Seed
 Clock
 Fixed 12345

Suppress All Intermediate Output
 Pause after Refresh
 Pause at End

Output Options
Output Type: Case Data
 Save summary statistics for every simulation run
 Save subject level data for 1 simulation runs
Note: Max. 100,000 records will be saved.

* Parameters estimated from Subject dataset : ONCOX_iLook1_SubjectData.cyx

Simulate Cancel

Click **Simulate**. After the simulation is complete the results appear in the **Library** as a sub-node of the ONCOX design:

Under this sub-node, there is a snapshot of the initial interim data entered (**Snap 1.1**), an output summary of the conditional simulation (**CS:Sim1**) and the updated versions of the output files generated during the initial (unconditional) simulation.

67.2.3 Output

Double-click 'CS:Sim1' in the Library to open a detailed summary of the conditional simulation. The first table, **Actuals: Sample Size and Look Times**, contains the interim data pertaining to the first look (402 subjects accrued, 187 events out of which 106 were in the control group and 81 were in the treatment group) and the time of the interim look (20.724 months). The second table named **Conditional Simulation: Average Sample Size and Look Times** displays the projections of these parameters for the second and last look.

Actuals: Sample Size and Look Times

| Look # | Sample Size | Events | | Dropouts | | Look Time | Average Follow up |
|--------|-------------|---------|-----------|----------|-----------|-----------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 402 | 106 | 81 | 1 | 2 | 20.724 | 4.908 |

Conditional Simulation: Average Sample Size and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---------|---------------------|----------------|-----------|------------------|-----------|-------------------|-------------------|
| | | Control | Treatment | Control | Treatment | | |
| 2 | 460 | 199.87 | 174.13 | 1.908 | 4.306 | 34.016 | 8.679 |
| Average | 460 | 199.87 | 174.13 | 1.908 | 4.306 | 34.016 | 8.679 |

Actuals: Events and Boundaries

| Look # | Events | Boundaries | |
|--------|--------|------------|----------|
| | | Efficacy | Futility |
| 1 | 187 | -2.895 | 0.121 |

67 Conditional Simulation

Conditional Simulation: Simulation Boundaries and Boundary Crossing Probabilities (Incremental)

| Look # | Events | Boundaries | | Stopping For | | Total Simulations | |
|--------|--------|------------|----------|--------------|----------|-------------------|----------|
| | | Efficacy | Futility | Efficacy | Futility | Count | % |
| 2 | 374 | -1.972 | -1.972 | 864 | 136 | 1000 | 100.000% |
| Total | | | | 864 | 136 | 1000 | |
| % | | | | 86.400% | 13.600% | | |

Interim Data Information: Subject Data

Subject Dataset: ONCOX_iLook1_SubjectData.cyx
 Population ID: Trtmt (Treatment=1, Control=0)
 Response Variable: TimeOnStudy
 Status Indicator: Status
 Arrival Time: ArrivalTime

Simulation Seed and Elapsed Time

Starting Seed: 12345
 Total Number of Simulations: 1000
 Elapsed Time: 00:00:09

The table **Simulation Boundaries and Boundary Crossing Probabilities** shows the number of simulations in which the efficacy boundary is crossed at each look. For instance, by the second look, 374 events have been observed and the efficacy boundary has been crossed in 864 out of 1000 simulations.

68 *Enrollment/Events Prediction - Analysis*

Prediction is useful even in fixed sample trials, that is, trials in which the user is not interested in stopping early for efficacy or futility. Even in such trials, the user or authorized person(s) may have access to the interim subject and site data or at least the summarized trial data and may want to predict the future enrollment and event milestones in the trial. There may be situations where a group sequential trial might not have been designed using East, or might not possess an access to Interim Monitoring module of East. The investigator is still interested in predicting the Accrual Duration and Study Duration based on an interim subject data. Catering to the needs of all such studies, the **Predict** feature is developed in the current version of East. We make the prediction functionality available through **Analysis** menu.

During the design stage in chapter 66 we used simulation to explore the enrollment timeline and event prediction of four trials: Orlistat (normal design), CAPTURE (binomial design) and RALES and ONCOX (survival designs). The inputs for the design stage simulations were based on estimates of accrual rates and other parameters. Once the trial begins and we obtain data on the realized enrollment, we can use the **Predict** module to update the parameters and generate new predictions about the enrollment and event timelines.

In this chapter, we introduce the Predict feature available in Analysis menu of East 6.4 and demonstrate its use for normal, binomial and survival designs considering data arising from the respective studies. The Predict feature in Analysis can play a vital role in assisting the Data Monitoring Committee (DMC) statistician as well as sponsor statistician in the following manner.

A DMC statistician typically has access to unblinded trial data. With this, she can use Predict feature to forecast how long the subject enrollment is likely to take and how long the study will take to complete by predicting the time by which required number of events would be achieved separately on the treatment and control drug.

The sponsor statistician generally has access to the blinded trial data. She can use Predict feature to forecast enrollment duration as well as study duration based on the available blinded subject and/or events data.

The option of providing summary data as input makes the use of **Predict** feature possible whenever individual subject data are not available. The **Summary Data** may consist of information on number of subjects enrolled, number of events occurred, number of drop outs observed so far etc. In addition to these, estimates of parameters

68 Enrollment/Events Prediction - Analysis

such as hazard rates for events, hazard rates of drop outs also might be available based on the interim or prior data.

In this chapter, we will use four examples for three endpoints - normal (Orlistat trial), binomial (Capture trial), and survival (Rales trial and Oncox trial) to illustrate enrollment/events prediction procedures. The main purpose of these procedures is to predict at any time point of the study, the likely cumulative enrollment for normal, binomial and survival studies and events/dropouts for survival studies.

68.1 Enrollment Only

68.1.1 Subject-level Data

68.1.2 Subject Data with Site Information

68.1.3 Summary Data

Suppose we have a partial data on enrollments of subjects for the Orlistat trial described in chapter 11. The trial is still ongoing and we want to predict the time when the target enrollment would be complete. The enrollment data till the current calendar time are stored in the **ORLISTAT_iLook1_SubjectData.csv** file which is available in the **Samples** folder of East 6.4 installation.

68.1.1 Enrollment Only: Subject-level Data

Data preparation for the **Enrollment Only** menu of **Predict** involves compiling the required data from various sources at a certain cut-off point. The enrollment can be across number of sites or at a single center. For the **Enrollment Only** feature, arrival times of the subjects are required. In this illustration, we assume that there is only Subject data available which comprises of the following variables.

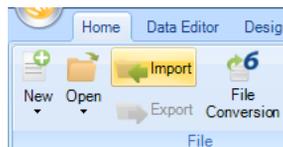
Subject Data Subject data refers to information collected about each subject accrued so far, namely:

- PatientID : Subject ID of the patient.
- Arrival time: the time at which the subject arrived.

For our example we prepare the data on the basis of the subjects enrolled so far. The data in **ORLISTAT_iLook1_SubjectData.csv** contains PatientID and Arrival Time. Note that the data contains some additional variables which are not required for this illustration but will be required later.

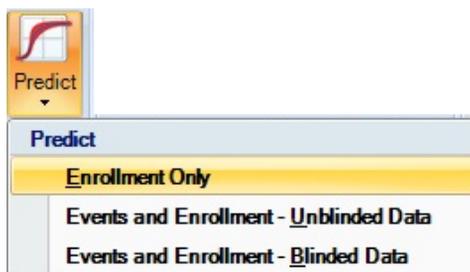
Import the **ORLISTAT_iLook1_SubjectData.csv** file into East using the

Import button in the Home ribbon:

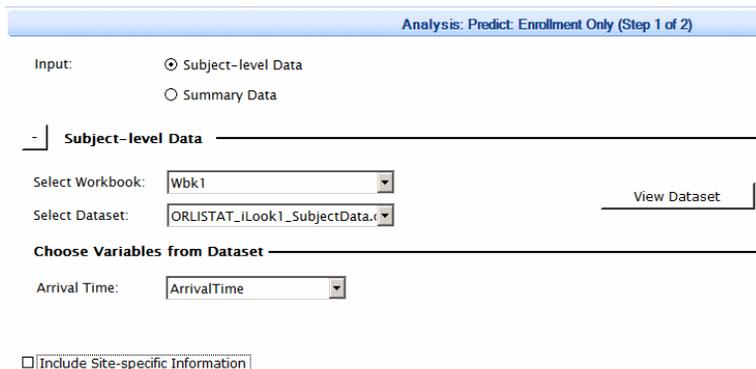


Once imported, the file will appear in the **Library** pane as a node in the active workbook, with extension `.cydx`.

Click on the node `ORLISTAT_iLook1_SubjectData.cydx` and choose the menu item `Analysis>Predict>Enrollment Only`.



In the resulting dialog box, select **Arrival Time** as shown below.



68 Enrollment/Events Prediction - Analysis

Since the default value for Input is **Subject-level Data** we leave it as it is. You may click the **View Dataset** button to view the data.

- Subject-level Data

Selected Dataset: ORLISTAT_iLook1_SubjectData.cydx

| PatientID: 1 Value: 1 | | | | | |
|-----------------------|-----------|--------|-------------|-------------|-------------|
| | PatientID | SiteID | ArrivalTime | TreatmentID | Response |
| 1 | 1 | SITE1 | 0.025987927 | 1 | 13.7118018 |
| 2 | 2 | SITE1 | 0.042236098 | 1 | 5.41314876 |
| 3 | 3 | SITE1 | 0.110120225 | 0 | -3.59692722 |
| 4 | 4 | SITE5 | 0.127380897 | 1 | -10.5383905 |
| 5 | 5 | SITE4 | 0.169874146 | 1 | 2.78465282 |

Click **Hide Dataset** to restore the dialog. Since we are not considering any Site information, leave the check box **Include Site-specific information** blank.

Click **Next**. This will invoke the next input dialog, **Accrual/Dropouts**.

Analysis: Predict: Enrollment Only (Step 1 of 2)

Target Sample Size:

Accrual / Dropouts | Simulation Controls

Current Sample Size: 212
Current Calendar Time: 2.224

Accrual Information

of Accrual Periods: Input Method:

| Period # | Starting at Time | Accrual Rate |
|----------|------------------|--------------|
| 1 | 0.0000 | 95.30998 |
| 2 | 2.2243 | 47.65499 |

You will see some default values already filled in. The **Current Sample Size** is the number of records (number of arrivals) in the data file which is 212 in this case. The **Target Sample Size** default value is 318 which is $1.5 * CurrentSampleSize$. You may change the **Target Sample Size** value. This is the value of targeted enrollment in the trial. The objective is to find out on an average how long will the trial take to enroll these many subjects. The **Current Calendar Time** is accrual time of the last subject in the data which is 2.224. The **Accrual Information** input is meant for simulating the additional, that is

$318 - 212 = 106$ accruals. There are two options for **Input Method**. For the **Accrual Rates** option, East considers the accrual process comprised of two periods. The first period is assumed to be the one presented in the data. Starting time for this period is assumed to be 0 whereas the accrual rate for this period is computed as $(CurrentSampleSize)/(CurrentCalendarTime)$. In this case it is $212/2.2243 = 95.31$. Both the starting time and Accrual Rate fields are uneditable as these are estimated from the data. The second period is the one which starts after the last accrual in the first period. As a result, the **Starting At** time default value is the last arrival time in the data which is also the **Current Calendar Time**. For the second period, the default accrual rate is computed as: $(TargetSampleSize - CurrentSampleSize)/(CurrentCalendarTime)$ which is $(318 - 212)/2.2243 = 47.65499$ for the current example. You can edit both the Starting Time and Accrual Rate for the second period. Accrual may vary over time. To reflect this assumption, one can specify the number of time periods, each having different accrual rates. An alternate way to give accrual input is **Cum Accrual %**. If you choose this option, the input dialog will be

Analysis: Predict: Enrollment Only (Step 1 of 2)

Target Sample Size:

Accrual / Dropouts | Simulation Controls

Current Sample Size: 212
Current Calendar Time: 2.224

Accrual Information

of Accrual Periods: Input Method:

| Period # | By Time | Accr. % |
|----------|---------|-----------|
| 1 | 2.224 | 66.66667 |
| 2 | 4.449 | 100.00000 |

As before, East treats the accruals in two pieces. The default value of **By Time** for the first period is the *CalendarTime* while for the second, it is $2 * CalendarTime$. Default values of **Accr %** for Period 1 and Period 2 are $100 * (CurrentSampleSize/TargetSampleSize)$ and 100% respectively. Both these values are uneditable. If you choose more than two accrual periods, the table expands and allows you to specify the values of **By Time** and **Accr %** fixing the **Accr %** for the last period to 100%.

For this study, let us use the option of **Accrual Rates** and use the default values.

68 Enrollment/Events Prediction - Analysis

Go to the **Simulation Controls** tab. It will show the following screen.

Analysis: Prediction of Enrollment Timeline (Step 2 of 2)

Sample Size:

Accrual / Dropouts | **Simulation Controls**

Number of Simulations:
 Refresh Frequency:

Random Number Seed
 Clock
 Fixed

Suppress All Intermediate Output
 Pause after Refresh
 Stop at End

Output Options
 Output Type:
 Save summary statistics for every simulation run
 Save subject level data for simulation runs
 Note: Max. 100,000 records will be saved.

Output for All Trials

| Percentile (%) |
|----------------|
| 5 |
| 25 |
| 50 |
| 75 |
| 95 |

The default number of simulation runs is 10000. Set the number of simulations to 1000 and the Random Number Seed to 12345. The simulation output can be saved either in a .csv file or as a Case Data. You can save **Summary Statistics** for every simulation run and the **Subject level data** for a few simulation runs. Suppose we want to save the **Summary Statistics** and the **Subject level data** for say, 5 simulation runs. Check both the check boxes and specify 5 simulation runs as indicated in the following screen shot. You can also modify the percentile values available in the **Output for All Trials** table. For now, let us keep them as they are. The **Simulation Controls** dialog will look as shown below:

Sample Size:

Accrual / Dropouts | **Simulation Controls**

Number of Simulations:
 Refresh Frequency:

Random Number Seed
 Clock
 Fixed

Suppress All Intermediate Output
 Pause after Refresh
 Stop at End

Output Options
 Output Type:
 Save summary statistics for every simulation run
 Save subject level data for simulation runs
 Note: Max. 100,000 records will be saved.

Output for All Trials

| Percentile (%) |
|----------------|
| 5 |
| 25 |
| 50 |
| 75 |
| 95 |

Click the **Simulate** button available at the bottom. East simulates the arrival of

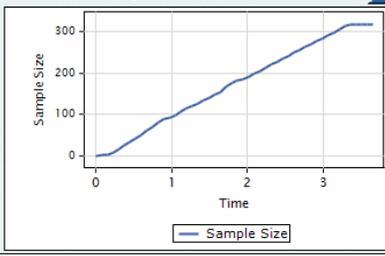
subjects according to the Poisson Arrival process. After a few seconds East will display the message that **Simulations complete. Waiting for User's action.**

Simulations in progress

Input Summary: Accrual: Ongoing, Sample Size = 318

| Current Time | Average Accrual Duration | Average Sample Size |
|--------------|--------------------------|---------------------|
| 2.224 | 3.326 | 318.000 |

Sample Size vs. Time



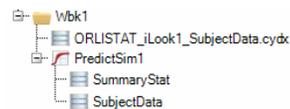
— Sample Size

**Simulations Complete
Waiting for User Action**

Close

Simulation Seed = 12345 Elapsed Time = 00:00:06 1000/1000

Click **Close**. This will save the **Predict Simulation** in the Output Preview window first. Once you save it in the Workbook, it will create a node **PredictSim1** with sub-nodes for **SummaryStat** and **SubjectData** in the **Library**.



Open the **SummaryStat** data by double clicking the sub node. You will see the

68 Enrollment/Events Prediction - Analysis

following display of data.

| ScenarioID | SimIndex | Stage | Status | TotAccruals | TotPendings | TotEvents | AccrDurtn |
|------------|----------|---------|---------|-------------|-------------|-----------|------------|
| 1 | 1 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 1 | Final | SUCCESS | 318 | 0 | 318 | 4.35719694 |
| 1 | 2 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 2 | Final | SUCCESS | 318 | 0 | 318 | 4.7111913 |
| 1 | 3 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 3 | Final | SUCCESS | 318 | 0 | 318 | 4.23758069 |
| 1 | 4 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 4 | Final | SUCCESS | 318 | 0 | 318 | 4.52927409 |
| 1 | 5 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 5 | Final | SUCCESS | 318 | 0 | 318 | 4.42808236 |
| 1 | 6 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 6 | Final | SUCCESS | 318 | 0 | 318 | 4.36122242 |
| 1 | 7 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 7 | Final | SUCCESS | 318 | 0 | 318 | 4.28554977 |
| 1 | 8 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 8 | Final | SUCCESS | 318 | 0 | 318 | 4.58337695 |

Observe that for every simulation East calls the **Current Sample Size** available data as **First Interim** while the **Target Sample Size** as **Final**.

The column **SUCCESS** indicates that the simulation was successful. The variable **TotEvents** is synonymous to Sample Size.

The last column **AccrDurtn** specifies the accrual duration required to enroll the 318 subjects in the respective simulation run. For instance, in the first simulation, the 318th subject arrived at the time epoch 4.35719 and so on. Now double click the

SubjectData sub node in the library. You will see the following display of data.

| ScenarioID | SimulationID | SubjectID | ArrivalTime |
|------------|--------------|-----------|-------------|
| 1 | 1 | 1 | 0.025987927 |
| 1 | 1 | 2 | 0.042236098 |
| 1 | 1 | 3 | 0.110120225 |
| 1 | 1 | 4 | 0.127380897 |
| 1 | 1 | 5 | 0.169874146 |
| 1 | 1 | 6 | 0.174202349 |
| 1 | 1 | 7 | 0.190352212 |
| 1 | 1 | 8 | 0.212313918 |
| 1 | 1 | 9 | 0.223044303 |
| 1 | 1 | 10 | 0.242310384 |
| 1 | 1 | 11 | 0.251971183 |
| 1 | 1 | 12 | 0.263265167 |
| 1 | 1 | 13 | 0.270861527 |
| 1 | 1 | 14 | 0.272797233 |
| 1 | 1 | 15 | 0.291490085 |
| 1 | 1 | 16 | 0.295691405 |
| 1 | 1 | 17 | 0.30714988 |
| 1 | 1 | 18 | 0.310955834 |
| 1 | 1 | 19 | 0.313170244 |
| 1 | 1 | 20 | 0.322496136 |

It shows arrival times of each and every subject in the study for five simulation runs. This is because we have asked to save the data for five runs. The first simulation id is 1. If you scroll down you will be able to see that East has chosen the simulation runs 4, 9,12 and 17 to save the data. This selection is arbitrary on the part of the software. Obviously, if we would have asked for saving data for 1000 runs, East would have chosen all the **SimulationIds** for saving the data (with the restriction that East can store at the most 100,000 records.)

To view the detailed summary output of the simulations, double click the node

68 Enrollment/Events Prediction - Analysis

`PredictSim1` in the Library. The following output is displayed.

Conditional Simulation: Prediction of Enrollment Timelines

| Test Parameters | |
|-------------------------------|--------------------|
| Simulation ID | PredictSim1 |
| Accrual | Ongoing |
| Input | Subject-level Data |
| Include Site Info | No |
| Target Sample Size | 318 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 1000 |

⊖ **Actuals from Interim Trial Data: Sample Size**

| Sample Size | Current Accrual Duration |
|-------------|--------------------------|
| 212 | 2.224 |

⊖ **Conditional Simulation: Average Sample Size**

| Average Sample Size | Average Accrual Duration |
|---------------------|--------------------------|
| 318 | 4.448 |

⊖ **Accrual / Dropouts Parameters**

Accrual Input Method: Accrual Rates

| Period # | Starting at Time | Accrual Rate |
|----------|------------------|--------------|
| 1 | 0 | 95.30998 |
| 2 | 2.224 | 47.65499 |

⊖ **Overall Output**

| Stage | Percentile | No. of Accruals | Accrual Duration |
|--------------|------------|-----------------|------------------|
| Interim | Actuals | 212 | 2.224 |
| | 5% | 318 | 4.103 |
| End of Trial | 25% | 318 | 4.303 |
| | 50% | 318 | 4.447 |
| | 75% | 318 | 4.587 |
| | 95% | 318 | 4.791 |
| | Average | 318 | 4.448 |

Subject-level Data

Subject Dataset: ORLISTAT_iLook1_SubjectData.cydx

Arrival Time: ArrivalTime

Simulation Seed and Elapsed Time

Starting Seed: 12345

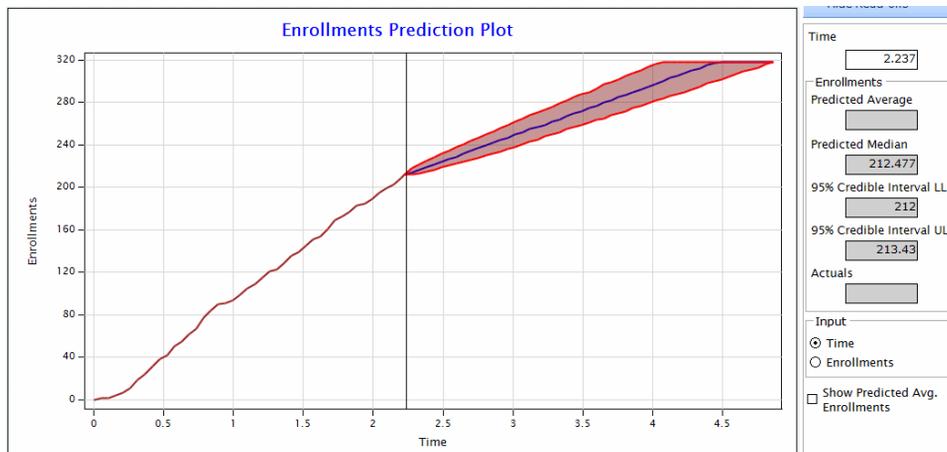
Total Number of Simulations: 1000

Elapsed Time: 00:00:09

The table at the left describes the Simulation scenario. This summary contains an overview of the actual data we observed and the simulated results for the remaining arrivals in the second period. The table **Overall Output** presents the information on the percentiles of the (simulated) total accrual duration. For example, almost 50% simulations have been completed by 4.447 units of time etc. The mean accrual duration of all the simulations is 4.448.

In addition, we can view the enrollment prediction plot using the  tool in the **Library** pane:

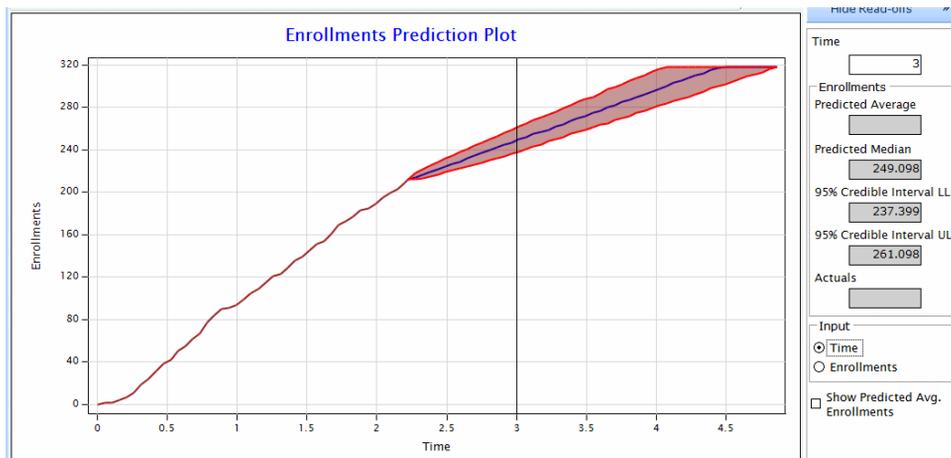
The **Enrollment Prediction Plot** displays the timeline of the observed accruals until 2.234 by which in all 212 subjects have been enrolled. This is as per the observed data.



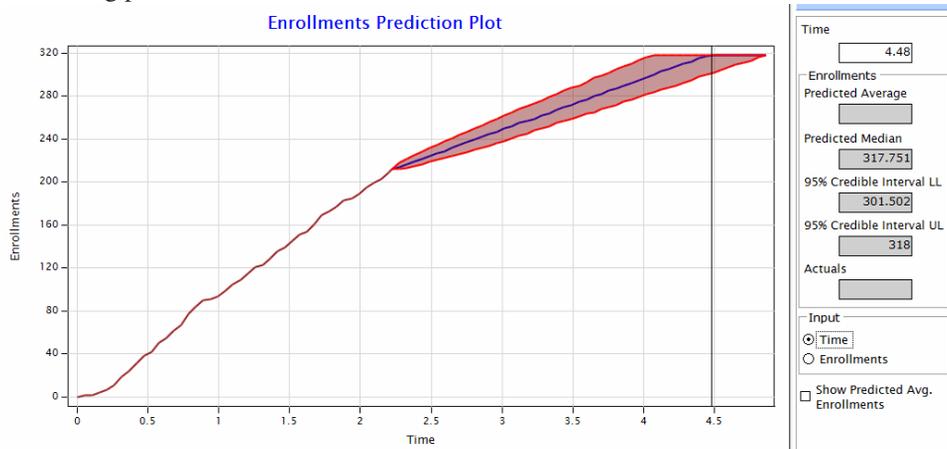
After that point it displays the projected enrollments based on the observed accrual data we specified and the revised **Accrual Rate** in the second period. For example, at year 3 the predicted median enrollments reach the sample size of 249 subjects with

68 Enrollment/Events Prediction - Analysis

95% Confidence Interval as (237.399, 261.098). Please see the plot below.

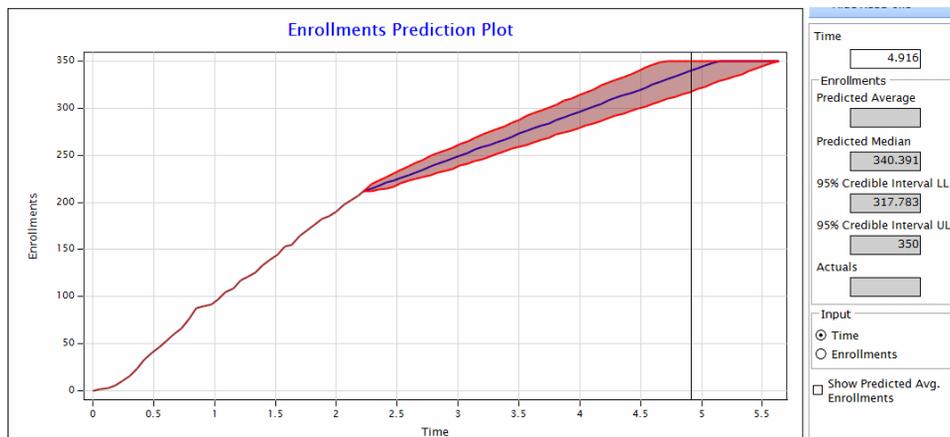


For the targeted 318 subjects it will take 4.48 units of time as is clear from the following plot.



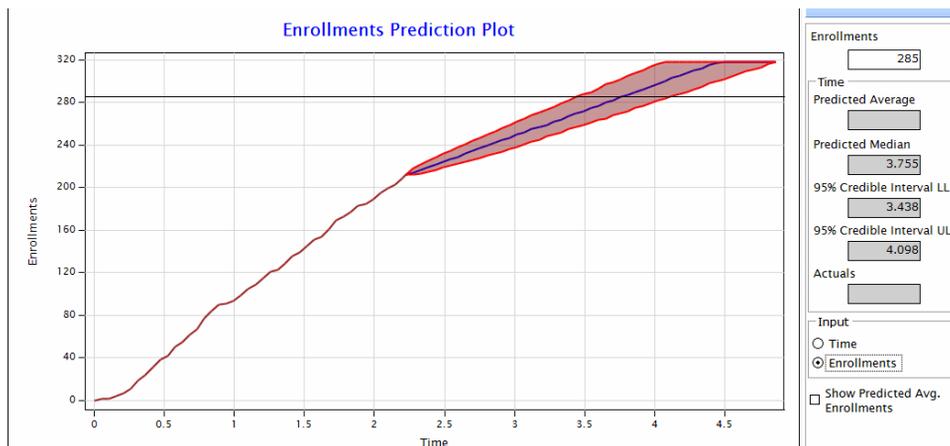
The problem here is that median and upper limit coincide and equal to 318. To envisage the true situation we suggest a workaround to the users. You can rerun the simulations with targeted sample size sufficiently greater than the true targeted sample size. For instance, if you consider the target sample size as say 350, and simulate

keeping rest of the things as previous, you get the **Enrollment Prediction Plot** as shown below:



Try to find out the **Time** so that the 95% lower limit is around 318 as shown in the picture above. This gives the latest time to get 318 events. From the plot, one can say that the latest time by which 318 accruals will happen is 4.916.

In the earlier plot (for 318 targeted sample size), if you want to find out how long will it take to enroll say, 285 subjects. Select the Input > Enrollments option on the plot and type 285 in the **Enrollments** textbox as shown in the following plot.



68 Enrollment/Events Prediction - Analysis

From the read-offs it is clear that the median accrual duration for accruing 285 subjects would be 3.755 with 95% confidence interval (3.438, 4.098).

68.1.2 Enrollment Only: Subject-level Data with Site-specific Information

In the case of a multi-center trial, accrual rates vary across sites. It is necessary to incorporate this information in the study to come up with a better estimate of total accrual duration on a whole. East provides a way to use this information by accepting the following information on different sites. Suppose for the Orlistat trial we also have the site data stored in a .csv file named **ORLISTAT_iLook1_SiteData.csv** which is available in the **Samples** folder of East installation directory. Import this file in East 6. Once imported, the file will appear in the **Library** pane as a node in the active workbook, with extension .cydx. The Site data comprises of the following variables.

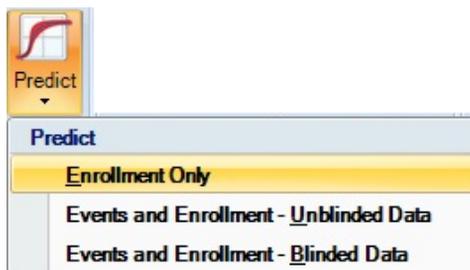
Site Data

Site data refers to information collected about each subject accrued so far, at each of the sites in a multi-center trial.

- Site ID: Site ID of the site.
- Site Accrual Rate: Site specific enrollment rate
- Enrollment Cap: This is the maximum number of subjects the site can enroll.
- Site Initiation: Unopened Sites
 - Start Time: It is the time at which the unopened site will open and start accepting accruals.
 - End Time: It is the time at which the site will stop accepting accruals and close.
- Site Initiation: Opened Sites
 - Site Initiation Time: It is the time at which the site was open and started accepting accruals.

Click on the node **ORLISTAT_iLook1_SubjectData.cydx** and choose the

menu item Analysis>Predict>Enrollment Only.



In the resulting dialog box, select the variable **ArrivalTime** in the drop down for **Arrival Time**. Since we want to include the Site information for this study, check the **Include Site-specific Information** check box. As soon as you check this option, the Input screen enables input for **Site ID** for Subject data as well as some more information about Site data such as workbook, dataset and some variables. Scroll down to see the complete Input dialog. Select **Site ID** for subject-level data and the data set **ORLISTAT_iLook1_SiteData.cyx** for the input of Site-level Data and map the variables from the Site data to the respective inputs as shown in the following screen which shows the necessary part of the input dialog.

Choose Variables from Dataset

Arrival Time: Site ID:

Include Site-specific Information

Site-level Data

Select Workbook:

Select Dataset:

Choose Variables from Dataset

Site ID: **Site Initiation: Unopened Sites**

Accrual Rate/Site: Start Time:

Enrollment Cap: End Time:

Site Initiation: Opened Sites

Site Initiation Time:

68 Enrollment/Events Prediction - Analysis

Click the button **View Dataset** to view the Site Data.

[-] Site-level Data

Selected Dataset: ORLISTAT_iLook1_SiteData.cydx

| SiteOpenFlag: 1 | | Value: 1 | | | |
|-----------------|-------------|----------|-------------|--------------|--------------|
| | SiteOpenFla | SiteID | SiteReadyTi | SiteAccrRate | SubjectsAccr |
| 1 | 1 | SITE1 | 0 | 10 | 22 |
| 2 | 1 | SITE2 | 0.186156561 | 10 | 26 |
| 3 | 1 | SITE3 | 0.230253726 | 10 | 25 |
| 4 | 1 | SITE4 | 0.154611662 | 10 | 19 |
| 5 | 1 | SITE5 | 0.122660027 | 10 | 17 |

Click **Hide Dataset**. Click **Next**. This will invoke the **Accrual/Dropouts Information** dialog. The **Current Sample Size** is 212 which is equal to the number of subjects arrived as per the subject data. East gives two options for generating arrivals either following Poisson process or Uniform. Let us select the option of **Poisson** arrivals.

Target Sample Size:

Accrual / Dropouts | Simulation Controls

Current Sample Size: 212
Current Calendar Time: 2.224

Total Number of Sites: Accrual Model: **Poisson**

Number of Sites Opened: 18

| Site ID | Site Initiation Period | | Accrual Rate/Site | Enrollment Cap | Planned Accrual Rate/Site | Site Initiation Time | No. of Accruals |
|---------|------------------------|-----|-------------------|----------------|---------------------------|----------------------|-----------------|
| | Start | End | | | | | |
| SITE1 | NA | NA | 9.89066 | 368 | 10 | 0 | 22 |
| SITE2 | NA | NA | 12.75658 | 368 | 10 | 0.186 | 26 |
| SITE3 | NA | NA | 12.53719 | 368 | 10 | 0.23 | 25 |
| SITE4 | NA | NA | 9.18003 | 368 | 10 | 0.155 | 19 |
| SITE5 | NA | NA | 8.08884 | 368 | 10 | 0.123 | 17 |
| SITE6 | NA | NA | 4.05073 | 368 | 5 | 0.249 | 8 |
| SITE7 | NA | NA | 4.72907 | 368 | 5 | 0.11 | 10 |
| SITE8 | NA | NA | 4.31522 | 368 | 5 | 0.139 | 9 |
| SITE9 | NA | NA | 5.49709 | 368 | 5 | 0.223 | 11 |
| SITE10 | NA | NA | 4.93886 | 368 | 5 | 0.2 | 10 |

Subject-level Data: ORLISTAT_iLook1_SubjectData.cydx
Site-level Data: ORLISTAT_iLook1_SiteData.cydx

Go to the **Simulation Controls** tab. Choose the **Random Number Seed** as

Fixed with its value 12345. Check the options for saving simulation outputs in files.

Target Sample Size:

Accrual / Dropouts | Simulation Controls

Number of Simulations:
 Refresh Frequency:

Random Number Seed
 Clock
 Fixed

Suppress All Intermediate Output
 Pause after Refresh
 Stop at End

Output Options
 Output Type:
 Save summary statistics for every simulation run
 Save subject level data for simulation runs
 Save site-wise summary for every simulation run
 Save site parameters data for simulation runs
 Note: Max. 100,000 records will be saved.

Output for All Trials

| Percentile (%) |
|----------------|
| 5.000 |
| 25.000 |
| 50.000 |
| 75.000 |
| 95.000 |

Subject-level Data: ORLISTAT_iLook1_SubjectData.cydx
 Site-level Data: ORLISTAT_iLook1_SiteData.cydx

Click **Simulate**. After a few seconds East will display the message **Simulations complete. Waiting for User's action**. Click **Close**. This will save the **Predict Simulation** in the Output Preview window first.

| ID | Average Sample Size | Average Accrual Duration |
|-------------|---------------------|--------------------------|
| PredictSim3 | 318 | 3.265 |

Once you save it in the Workbook, it will create a node **PredictSim3** with subnodes for the simulation outputs. To view the detailed output, double click the node

68 Enrollment/Events Prediction - Analysis

PredictSim3 in **Library**. You will see the following output.

Conditional Simulation: Prediction of Enrollment Timelines

| Test Parameters | |
|-------------------------------|--------------------|
| Simulation ID | PredictSim3 |
| Accrual | Ongoing |
| Input | Subject-level Data |
| Include Site Info | Yes |
| Target Sample Size | 318 |
| Accrual / Dropouts Parameters | |
| Accrual Model | Poisson |
| Total No. of Sites | 18 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 1000 |

Actuals from Interim Trial Data: Sample Size

| Sample Size | Current Accrual Duration |
|-------------|--------------------------|
| 212 | 2.224 |

Conditional Simulation: Average Sample Size

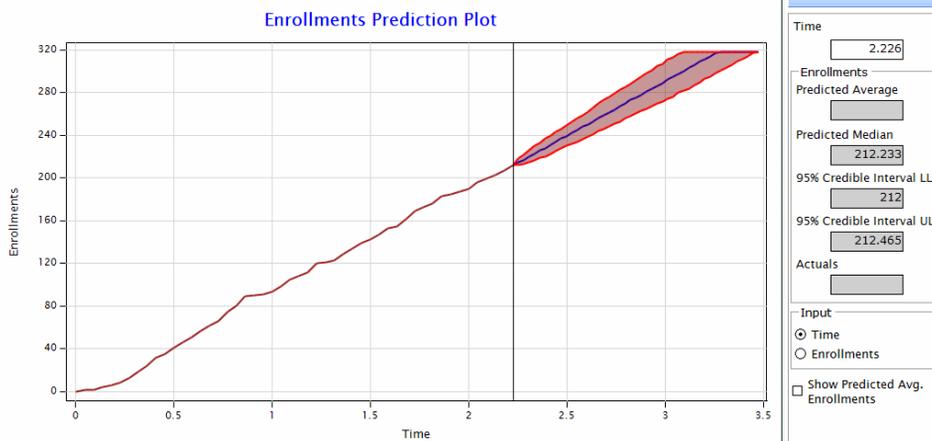
| Average Sample Size | Average Accrual Duration |
|---------------------|--------------------------|
| 318 | 3.265 |

Overall Output

| | Stage | Percentile | No. of Sites Opened | No. of Accruals | Accrual Duration |
|---|--------------|------------|---------------------|-----------------|------------------|
| | Interim | Actuals | 0 | 212 | 2.224 |
| - | End of Trial | 5% | 18 | 318 | 3.107 |
| | | 25% | 18 | 318 | 3.2 |
| | | 50% | 18 | 318 | 3.261 |
| | | 75% | 18 | 318 | 3.335 |
| | | 95% | 18 | 318 | 3.423 |
| | | Average | 18 | 318 | 3.265 |

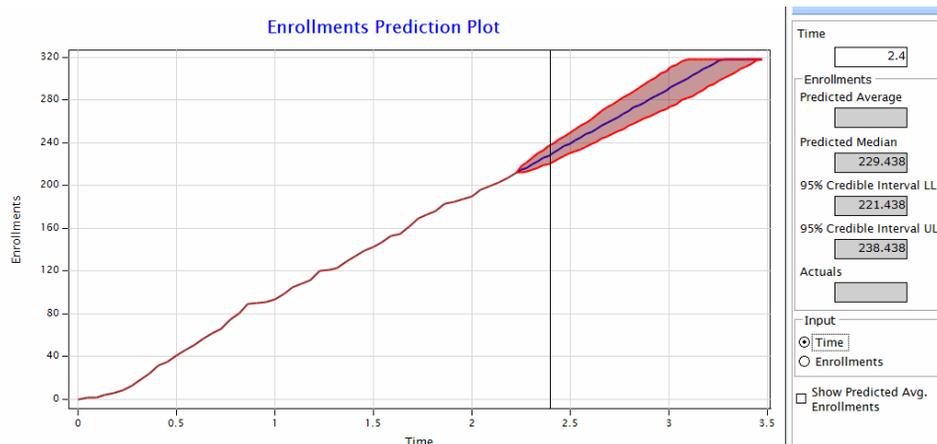
The **Average Accrual Duration** across all the simulations is 3.265. The **Accrual Duration** column in the **Overall Summary** table indicates the frequency distribution of **Accrual Duration**. Accordingly, median accrual duration is 3.261, whereas 75% of the simulations have total Accrual Duration 3.335.

To view the **Enrollments Simulation Plot** click the **PredictSim1** node in the **Library**; use the  tool in the **Library** pane.



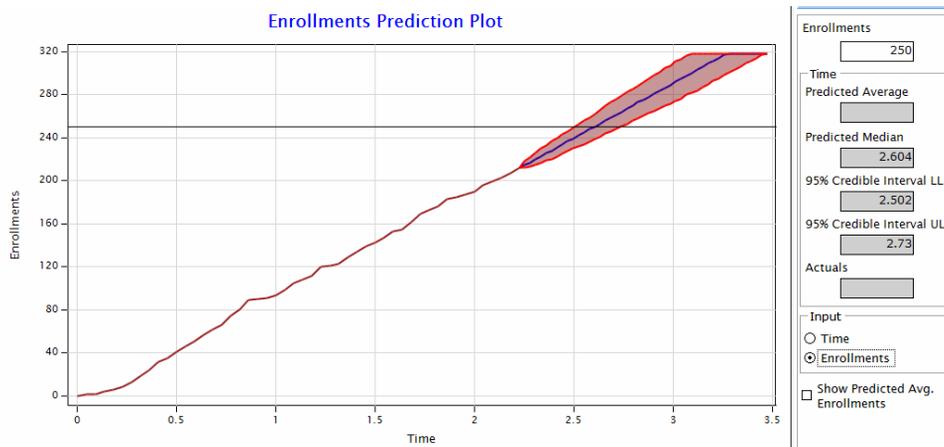
The **Enrollment Prediction Plot** displays the timeline of the observed accruals until 2.226 by which in all 212 subjects have been enrolled. This is as per the observed data.

After that point it displays the projected enrollments based on the observed accrual data we specified and the revised **Accrual Rate** in the second period. For example, at year 2.4 the predicted median enrollments reach the sample size of 229 subjects with the 95% Confidence Interval as (221.438, 238.438). Please see the plot below.



68 Enrollment/Events Prediction - Analysis

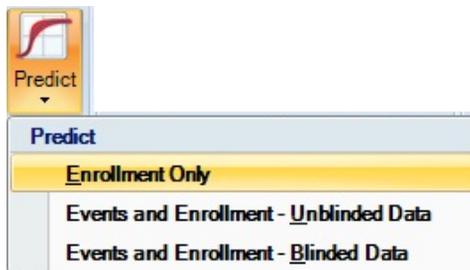
On similar lines, if you want to find out how long will it take to enroll say, 250 subjects. Select the Input> Enrollments option on the plot and type 250 in the **Enrollments** textbox as shown in the following plot.



From the read-offs it is clear that the median accrual duration for accruing 250 subjects would be 2.604 with 95% confidence interval (2.502, 2.73).

The **Predict** feature of East can also handle situations where the sites are initially closed, but would open later and start accruing the subjects subsequently. We will illustrate this feature now.

Import the **ORLISTAT_EnrollmentOnly_SubjectData.csv** and **ORLISTAT_EnrOnly_SiteData.csv** which will create data nodes in the library. As before, choose the menu item Analysis>Predict>Enrollment Only.



In the resulting dialog box, give the inputs as shown below:

Input: Subject-level Data
 Summary Data

- | **Subject-level Data** _____

Select Workbook:

Select Dataset:

Choose Variables from Dataset _____

Arrival Time: Site ID:

Include Site-specific Information

- | **Site-level Data** _____

Select Workbook:

Select Dataset:

- | **Site-level Data** _____

Select Workbook:

Select Dataset:

Choose Variables from Dataset _____

Site ID:

Accrual Rate/Site:

Enrollment Cap:

Site Initiation: Unopened Sites

Start Time:

End Time:

Site Initiation: Opened Sites

Site Initiation Time:

68 Enrollment/Events Prediction - Analysis

Click **View Site Dataset**

Selected Dataset: ORLISTAT_EnrOnly_SiteData.cydx

| SiteOpenFlag: 1 | | Value: 1 | | | | |
|-----------------|-------------|----------|-------------|--------------|-----------|--|
| | SiteOpenFla | SiteID | SiteReadyTi | SiteAccrRate | SIP_Start | |
| 1 | 1 | SITE1 | 0 | 10 | 0 | |
| 2 | 1 | SITE2 | 0.186156561 | 10 | 0 | |
| 3 | 1 | SITE3 | 0.230253726 | 10 | 0 | |
| 4 | 0 | SITE4 | . | 10 | 2 | |
| 5 | 1 | SITE5 | 0.122660027 | 10 | 0 | |

Note that the **Site Ready Time** for sites 4 and 10 are missing. Note that the two sites can be opened anytime during the time interval (2, 9). Accordingly, the **SIP_Start** and **SIP_End** values are 2 and 9 respectively.

Click **Hide Dataset**. Click **Next**. This will invoke the **Accrual/Dropouts Information** dialog. The **Current Sample Size** is 183 as the subjects at the sites 4 and 10 have not yet been enrolled. East gives two options for generating arrivals: Poisson or Uniform. Select the option of **Poisson** arrivals. Scroll down a little for the middle table to view the **Unopened Sites** information. The values of **Site Initiation Period Start** and **End** are **NA** as these sites are already open whereas for the sites 4 and 10, **Start** and periods are specified which will be used to generate the **Site Initiation Times** for these two sites. The column **Accrual Rate/Site** depicts the accrual rates calculated from the existing data. These will be used to simulate the remaining $274 - 183 = 91$ accruals. You may change the values of **Accrual rate/Site**. Suppose henceforth the sites 13 and 17 are expected to enroll the subjects pretty fast. We want to change the accrual rates for the sites 13 and 17 to 20 and 40 respectively. Change the corresponding values. The **Planned Accrual Rate** are the values read from the data in the variable **SiteAcrrRate**. These values can't be edited. Now the input screen showing the lower part of the table scrolled down would look as shown below:

Target Sample Size:

Current Sample Size: 183
 Current Calendar Time: 2.224

Total Number of Sites: Accrual Model:

Number of Sites Opened: 16

| Site ID | Site Initiation Period | | Accrual Rate/Site | Enrollment Cap | Planned Accrual Rate/Site | Site Initiation Time | No. of Accruals |
|---------|------------------------|-----|-------------------|----------------|---------------------------|----------------------|-----------------|
| | Start | End | | | | | |
| SITE11 | NA | NA | 5.5078 | 368 | 5 | 0.046 | 12 |
| SITE12 | NA | NA | 4.20469 | 368 | 5 | 0.084 | 9 |
| SITE13 | NA | NA | 20 | 368 | 5 | 0.104 | 11 |
| SITE14 | NA | NA | 2.36288 | 368 | 2 | 0.108 | 5 |
| SITE15 | NA | NA | 2.32135 | 368 | 2 | 0.07 | 5 |
| SITE16 | NA | NA | 2.70295 | 368 | 2 | 0.005 | 6 |
| SITE17 | NA | NA | 40 | 368 | 2 | 0.209 | 3 |
| SITE18 | NA | NA | 1.93816 | 368 | 2 | 0.161 | 4 |
| SITE4 | 2 | 9 | 10 | 368 | 10 | NA | NA |
| SITE10 | 2 | 9 | 5 | 368 | 5 | NA | NA |

Subject-level Data: ORLISTAT_EnrollmentOnly_SubjectData.cydx
 Site-level Data: ORLISTAT_EnrOnly_SiteData.cydx

Go to the **Simulation Controls** tab. You can change the simulation parameters here as per your wish. Suppose you want to have the outputs stored in .csv format. Select the **Output Type** as CSV file. Either you can already create the files in local folders and select them using the **Browse** button or you may create the files while you are browsing.

Suppose the **Summary.csv**, **SubjectData.csv**, **SitewiseSummary.csv** and **SitewisePara.csv** are the files which would store the Summary statistics for every simulation run, Subject level data for 1 simulation run, Sitewise summary for every simulation run and Sitewise parameter data for 1 simulation run respectively. All these files are to be stored on say the local drive *G*. Choose the **Random Number Seed** as **Fixed** with its value 12345. The input screen for simulation will be seen as

68 Enrollment/Events Prediction - Analysis

shown below:

Target Sample Size:

Accrual / Dropouts | Simulation Controls

Number of Simulations:
 Refresh Frequency:

Random Number Seed
 Clock
 Fixed

Suppress All Intermediate Output
 Pause after Refresh
 Stop at End

Output Options
 Output Type:

Save summary statistics for every simulation run

Save subject level data for simulation runs

Save site-wise summary for every simulation run

Save site parameters data for simulation runs

Output for All Trials

Subject-level Data: ORLISTAT_EnrollmentOnly_SubjectData.cydx
 Site-level Data: ORLISTAT_EnrOnly_SiteData.cydx

Click **Simulate**. After a few seconds East will display the message **Simulations complete. Waiting for User's action**. Click **Close**. This will save the **Predict Simulation** in the Output Preview window first.

| ID | Average Sample Size | Average Accrual Duration |
|-------------|---------------------|--------------------------|
| PredictSim4 | 274 | 2.866 |

Once you save it in the Workbook, it will create a node **PredictSim1**. This time no sub nodes will be created as we have asked the output to be saved in .CSV files at the specified locations on the machine. To view the detailed output, double click the node

PredictSim1 in **Library**. You will see the following output.

Conditional Simulation: Prediction of Enrollment Timelines

| Test Parameters | |
|-------------------------------|--------------------|
| Simulation ID | PredictSim4 |
| Accrual | Ongoing |
| Input | Subject-level Data |
| Include Site Info | Yes |
| Target Sample Size | 274 |
| Accrual / Dropouts Parameters | |
| Accrual Model | Poisson |
| Total No. of Sites | 18 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 1000 |

Actuals from Interim Trial Data: Sample Size

| Sample Size | Current Accrual Duration |
|-------------|--------------------------|
| 183 | 2.224 |

Conditional Simulation: Average Sample Size

| Average Sample Size | Average Accrual Duration |
|---------------------|--------------------------|
| 274 | 2.866 |

Overall Output

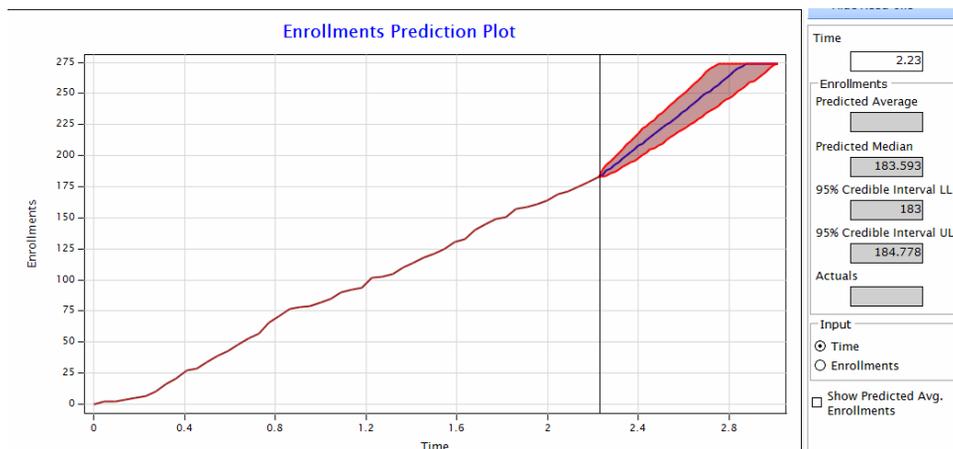
| | Stage | Percentile | No. of Sites Opened | No. of Accruals | Accrual Duration |
|--|--------------|------------|---------------------|-----------------|------------------|
| | Interim | Actuals | 0 | 183 | 2.224 |
| | End of Trial | 5% | 16 | 274 | 2.759 |
| | | 25% | 16 | 274 | 2.818 |
| | | 50% | 16 | 274 | 2.864 |
| | | 75% | 16 | 274 | 2.91 |
| | | 95% | 17 | 274 | 2.979 |
| | | Average | 16.194 | 274 | 2.866 |

The **Average Accrual Duration** across all the simulations is 2.866. The **Accrual Duration** column in the **Overall Summary** table indicates the frequency distribution of **Accrual Duration**. Accordingly, median accrual duration is 2.864, whereas 75% of the simulations have total Accrual Duration 2.91.

To view the **Enrollments Simulation Plot** click the **PredictSim1** node in

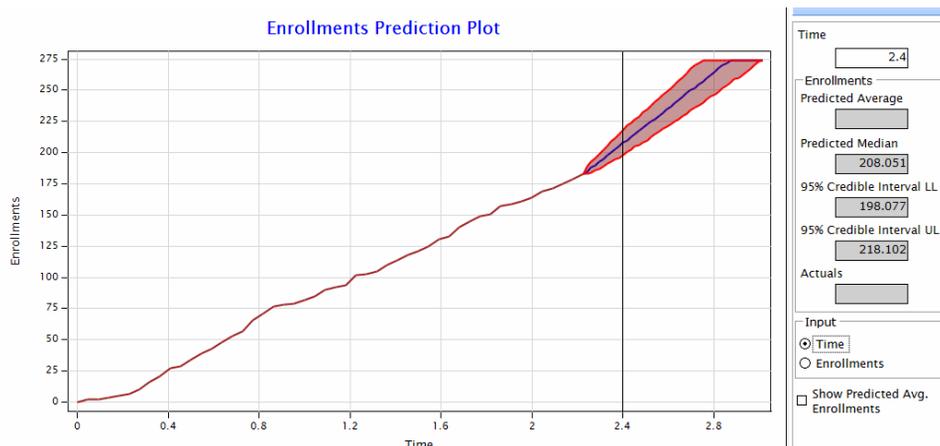
68 Enrollment/Events Prediction - Analysis

the Library; use the  tool in the **Library** pane.



The **Enrollment Prediction Plot** displays the timeline of the observed accruals until 2.23 by which in all 183 subjects have been enrolled. This is as per the observed data.

After that point it displays the projected enrollments based on the observed accrual data we specified and the revised **Accrual Rate** in the second period. For example, at year 2.4 the predicted median enrollments reach the sample size of 208 subjects with the 95% Confidence Interval as (198.077, 218.102). Please see the plot below.



Let us investigate further the output of the simulations run. Below is repeated the **Overall Output** from the detail output.

Overall Output

| | Stage | Percentile | No. of Sites Opened | No. of Accruals | Accrual Duration |
|--------------|---------|------------|---------------------|-----------------|------------------|
| | Interim | Actuals | 0 | 183 | 2.224 |
| End of Trial | | 5% | 16 | 274 | 2.759 |
| | | 25% | 16 | 274 | 2.818 |
| | | 50% | 16 | 274 | 2.864 |
| | | 75% | 16 | 274 | 2.91 |
| | | 95% | 17 | 274 | 2.979 |
| | | Average | | 16.194 | 274 |

Subject-level Data

Subject Dataset: ORLISTAT_EnrollmentOnly_SubjectData.cydx
 Arrival Time: ArrivalTime
 Site ID: SiteID

Site-level Data

Dataset: ORLISTAT_EnrOnly_SiteData.cydx
 Site ID: SiteID
 Accrual Rate / Site: SiteAccrRate
 Enrollment Cap: Ecap
 Unopened Sites: Start SIP_Start
 Unopened Sites: End SIP_End
 Opened Sites: Site Initiation Time: SiteReadyTime

If you have a close look at the column **No of Sites Opened**, you will notice that till the 75th percentile of the Accrual duration, that is by 2.91, only 16 sites were open as was the situation in the beginning. One more site has got opened during 2.91 and 2.979. Let us see what all has happened during this time period.

We have saved the outputs in .CSV files. Open the file **Summary.csv** which stores summary statistics for every simulation run. You will see the data as shown below:

68 Enrollment/Events Prediction - Analysis

| ScenarioID | SimIndex | Stage | Status | TotAccruals | TotPendings | TotEvents | AccrDurtn |
|------------|----------|---------|---------|-------------|-------------|-----------|-----------|
| 1 | 1 | Interim | SUCCESS | 183 | 0 | 183 | |
| 1 | 1 | Final | SUCCESS | 274 | 0 | 274 | 2.804084 |
| 1 | 2 | Interim | SUCCESS | 183 | 0 | 183 | |
| 1 | 2 | Final | SUCCESS | 274 | 0 | 274 | 3.015385 |
| 1 | 3 | Interim | SUCCESS | 183 | 0 | 183 | |
| 1 | 3 | Final | SUCCESS | 274 | 0 | 274 | 2.732081 |
| 1 | 4 | Interim | SUCCESS | 183 | 0 | 183 | |
| 1 | 4 | Final | SUCCESS | 274 | 0 | 274 | 2.860316 |
| 1 | 5 | Interim | SUCCESS | 183 | 0 | 183 | |
| 1 | 5 | Final | SUCCESS | 274 | 0 | 274 | 2.91049 |
| 1 | 6 | Interim | SUCCESS | 183 | 0 | 183 | |
| 1 | 6 | Final | SUCCESS | 274 | 0 | 274 | 2.809429 |
| 1 | 7 | Interim | SUCCESS | 183 | 0 | 183 | |
| 1 | 7 | Final | SUCCESS | 274 | 0 | 274 | 2.834545 |
| 1 | 8 | Interim | SUCCESS | 183 | 0 | 183 | |
| 1 | 8 | Final | SUCCESS | 274 | 0 | 274 | 2.883579 |
| 1 | 9 | Interim | SUCCESS | 183 | 0 | 183 | |
| 1 | 9 | Final | SUCCESS | 274 | 0 | 274 | 2.912459 |
| 1 | 10 | Interim | SUCCESS | 183 | 0 | 183 | |
| 1 | 10 | Final | SUCCESS | 274 | 0 | 274 | 2.882488 |
| 1 | 11 | Interim | SUCCESS | 183 | 0 | 183 | |
| 1 | 11 | Final | SUCCESS | 274 | 0 | 274 | 2.815699 |
| 1 | 12 | Interim | SUCCESS | 183 | 0 | 183 | |
| 1 | 12 | Final | SUCCESS | 274 | 0 | 274 | 2.97942 |

Observe that for every simulation East calls the **Current Sample Size** available data as **Interim** while the **Target Sample Size** as **Final**.

The column **SUCCESS** indicates that the simulation was successful. The variable **TotEvents** is synonymous to Sample Size.

The last column **AccrDurtn** specifies the accrual duration required to enroll the 274 subjects in the respective simulation run. For instance, in the first simulation, the 274th subject arrived at the time epoch 2.80408 and so on. Now open the **Subject.csv** file which stores arrival times of each subject for one simulation. You will see the

following display of data.

| ScenarioID | Simulation | SubjectID | SiteID | ArrivalTime |
|------------|------------|-----------|--------|-------------|
| 1 | 4 | 1 | SITE1 | 0.025988 |
| 1 | 4 | 2 | SITE1 | 0.042236 |
| 1 | 4 | 3 | SITE1 | 0.11012 |
| 1 | 4 | 4 | SITE5 | 0.127381 |
| 1 | 4 | 5 | SITE1 | 0.174202 |
| 1 | 4 | 6 | SITE11 | 0.190352 |
| 1 | 4 | 7 | SITE18 | 0.223044 |
| 1 | 4 | 8 | SITE1 | 0.24231 |
| 1 | 4 | 9 | SITE14 | 0.251971 |
| 1 | 4 | 10 | SITE13 | 0.270862 |
| 1 | 4 | 11 | SITE9 | 0.272797 |
| 1 | 4 | 12 | SITE6 | 0.29149 |
| 1 | 4 | 13 | SITE1 | 0.295691 |
| 1 | 4 | 14 | SITE8 | 0.30715 |
| 1 | 4 | 15 | SITE16 | 0.310956 |
| 1 | 4 | 16 | SITE11 | 0.31317 |
| 1 | 4 | 17 | SITE3 | 0.322496 |
| 1 | 4 | 18 | SITE5 | 0.323448 |
| 1 | 4 | 19 | SITE1 | 0.330773 |
| 1 | 4 | 20 | SITE7 | 0.343222 |
| 1 | 4 | 21 | SITE9 | 0.354224 |
| 1 | 4 | 22 | SITE3 | 0.366509 |
| 1 | 4 | 23 | SITE8 | 0.368021 |
| 1 | 4 | 24 | SITE14 | 0.377406 |

Note that the data are ordered according to the arrival times of all the subjects across all the sites. The first three subjects arrive at Site 1 in succession, whereas the fourth subject arrives at Site 5 and so on. You can sort the data on sites and see that the sites 4 and 10 are not opened in this particular simulation, that is simulation 4. The last subject arrived in Simulation 1 at time point 2.8603 at Site 12. The question is when the sites 4 and 10 finally opened and started accepting subjects. Since these sites don't occur in the **Summary** data, they must have got opened after the last arrival. To verify this, open the file **SitewisePara.csv** which stores the site parameter data for one simulation run.

68 Enrollment/Events Prediction - Analysis

The data are as follows:

| SimulationID | SiteOpenFlag | SiteAlreadyOpened | SiteID | SiteInitiationTime | SiteAccrRate | SubjectsAccrued | LastSubjectRand | AccrualDuration | ObsrvdAccrualRate |
|--------------|--------------|-------------------|--------|--------------------|--------------|-----------------|-----------------|-----------------|-------------------|
| 4 | 1 | 1 | SITE1 | 0 | 9.890658354 | 30 | 2.800099846 | 2.860315512 | 10.48835343 |
| 4 | 1 | 1 | SITE2 | 0.186156561 | 12.75657563 | 32 | 2.837161558 | 2.674158951 | 11.96637918 |
| 4 | 1 | 1 | SITE3 | 0.230253726 | 12.53718927 | 37 | 2.821524552 | 2.630061786 | 14.06811057 |
| 4 | 1 | 1 | SITE5 | 0.122660027 | 8.088839954 | 25 | 2.853708425 | 2.737655485 | 9.131901417 |
| 4 | 1 | 1 | SITE6 | 0.249368303 | 4.050729739 | 9 | 2.24351113 | 2.610947209 | 3.447024884 |
| 4 | 1 | 1 | SITE7 | 0.109742456 | 4.729074542 | 13 | 2.53453285 | 2.750573056 | 4.726287845 |
| 4 | 1 | 1 | SITE8 | 0.138679635 | 4.315219169 | 12 | 2.804154539 | 2.721635877 | 4.409112954 |
| 4 | 1 | 1 | SITE9 | 0.2232636 | 5.497093431 | 13 | 2.826070896 | 2.637051912 | 4.929747474 |
| 4 | 1 | 1 | SITE11 | 0.045993743 | 5.507802525 | 18 | 2.760596596 | 2.814721769 | 6.394948232 |
| 4 | 1 | 1 | SITE12 | 0.083856203 | 4.204694056 | 11 | 2.860315512 | 2.776459309 | 3.961880502 |
| 4 | 1 | 1 | SITE13 | 0.104353084 | 20 | 21 | 2.856303581 | 2.755962428 | 7.619842632 |
| 4 | 1 | 1 | SITE14 | 0.108255554 | 2.362875775 | 7 | 2.581823687 | 2.752059958 | 2.543549235 |
| 4 | 1 | 1 | SITE15 | 0.070404542 | 2.321352699 | 5 | 2.039847442 | 2.78991097 | 1.792171884 |
| 4 | 1 | 1 | SITE16 | 0.004520343 | 2.702945298 | 7 | 2.242340779 | 2.855795169 | 2.451156188 |
| 4 | 1 | 1 | SITE17 | 0.209314545 | 40 | 30 | 2.852858541 | 2.651000967 | 11.31648022 |
| 4 | 1 | 1 | SITE18 | 0.160503158 | 1.93815546 | 4 | 0.693544141 | 2.699812354 | 1.48158445 |
| 4 | 0 | 0 | SITE4 | 3.568568796 | 10 | 0 | | | |
| 4 | 0 | 0 | SITE10 | 7.957575908 | 5 | 0 | | | |

This file gives site wise details of the accrual process. The columns **SiteInitiationTime** and **SubjectsAccrued** specify the time at which the site was opened and the number of subjects it accrued in that simulation. The column **LastSubjectRand** shows the time at which the last subject arrived at that site. Since the accrual rates of different sites are different, some sites would accrue more subjects than the ones having low accrual rates. Also the **SiteInitiation time** matters for accruing a few or more subjects. Note that from the last two rows, the **SiteInitiationTime** for Site 4 is 3.5685 and for Site 10, it is 7.9575. However, the last subject arrived in the study at 2.8603 at Site 12. As a result, both the sites 4 and 10 got opened after the accruals in the study on a whole were complete.

The columns **Accrual Duration** and **ObsrvdAccrualRate** specify the site wise accrual duration and the rate at which the site accrued subjects. Now open the

SitewiseSummary.csv file. It will display the following information.

| SiteID | AvgInitiationTime | AvgLastSubjArrTi | AvgNumOfSubj | AvgAccrualDuration | AvgAccrualRate | SiteOpenedSimCount |
|--------|-------------------|------------------|--------------|--------------------|----------------|--------------------|
| SITE1 | 0 | 2.771076386 | 28.425 | 2.86578114 | 9.924168948 | 1000 |
| SITE2 | 0.186156561 | 2.790011152 | 34.102 | 2.679624579 | 12.73493555 | 1000 |
| SITE3 | 0.230253726 | 2.793034357 | 33.121 | 2.635527414 | 12.57435624 | 1000 |
| SITE5 | 0.122660027 | 2.75273273 | 22.25 | 2.743121113 | 8.117166776 | 1000 |
| SITE6 | 0.249368303 | 2.594265724 | 10.581 | 2.616412837 | 4.04622245 | 1000 |
| SITE7 | 0.109742456 | 2.66879964 | 13.107 | 2.756038684 | 4.758469916 | 1000 |
| SITE8 | 0.138679635 | 2.615900021 | 11.727 | 2.727101505 | 4.302407839 | 1000 |
| SITE9 | 0.2232636 | 2.69809095 | 14.595 | 2.64251754 | 5.527265536 | 1000 |
| SITE11 | 0.045593743 | 2.701926042 | 15.583 | 2.820187397 | 5.528758013 | 1000 |
| SITE12 | 0.083856203 | 2.649938265 | 11.692 | 2.781924937 | 4.204762602 | 1000 |
| SITE13 | 0.104353084 | 2.827229069 | 23.787 | 2.761428056 | 8.61788689 | 1000 |
| SITE14 | 0.108255554 | 2.377496429 | 6.512 | 2.757525586 | 2.362744397 | 1000 |
| SITE15 | 0.070404542 | 2.491420976 | 6.469 | 2.795376598 | 2.315971655 | 1000 |
| SITE16 | 0.004520343 | 2.530278341 | 7.756 | 2.861260797 | 2.711734131 | 1000 |
| SITE17 | 0.209314545 | 2.847897825 | 28.556 | 2.656466595 | 10.75760759 | 1000 |
| SITE18 | 0.160503158 | 2.085822219 | 5.264 | 2.705277982 | 1.946455293 | 1000 |
| SITE4 | 2.53903862 | 2.788917193 | 3.381443299 | 0.352922115 | 13.95909324 | 97 |
| SITE10 | 2.54878504 | 2.755377017 | 1.494845361 | 0.377991763 | 6.523872886 | 97 |

Averages across all the simulations for the quantities, **Initiation Time, last Subject Arrival Time, Number of Subjects Enrolled, Accrual Duration, Accrual Rate** are provided for individual sites. From the last two rows, it should be noted that only in 97 simulations out of 1000, the two sites 4 and 10 have got opened during the accrual duration.

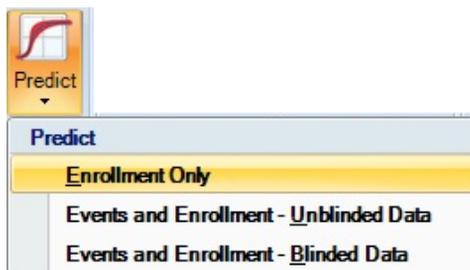
We suggest you to try with different input subject data and/ or site data with varying accrual rates, site initiation intervals, enrollment cap to develop better insight of the accrual process.

68.1.3 Enrollment Only: Summary Data

In the earlier section we saw how to simulate the accruals and estimate the average accrual duration when an interim subject-level data is available. However, sometimes, the subject-level data may not be available. What can be available is the summary of the accruals that have happened till date. For example, in the case of Orlistat trial considered above, the DMC statistician may have the information that there have been 212 subjects accrued so far and the last subject arrived at time 2.224. The DMC statistician is interested in knowing the total accrual duration for say 318 accruals. East through its Predict feature makes it possible to still come up with an estimate of average accrual duration based on arrival simulations for these additional $318 - 212 = 106$ arrivals. To see this, choose the menu item

68 Enrollment/Events Prediction - Analysis

Analysis>Predict>Enrollment Only.



In the ensuing dialog box, select the **Input** option, **Summary Data**. Fill in the Sample Size as 212 and the **Current Calendar Time** as 2.224. The screen will look as shown below.

Analysis: Predict: Enrollment Only (Step 1 of 2)

Input: Subject-level Data
 Summary Data

Summary Data

Sample Size:

Current Calendar Time:

* Click 'Next' to specify other parameters.

Click **Next**. The next input dialog appears as shown here:

Analysis: Predict: Enrollment Only (Step 2 of 2)

Target Sample Size:

Accrual / Dropouts

Simulation Controls

Current Sample Size: 212

Current Calendar Time: 2.224 Accrual Model: Poisson

Accrual Information

of Accrual Periods: Input Method:

| Period # | Starting at Time | Accrual Rate |
|----------|------------------|--------------|
| 1 | 0.000 | 95.324 |
| 2 | 2.224 | 47.662 |

As before, default values for **Current Sample Size** and the **Target Sample Size** are 212 and 318 respectively, default for the **target Sample Size** being $1.5 * CurrentSampleSize$. You may change the **Target Sample Size** value. This is the value of targeted enrollment in the trial. The objective is to find out on an average how long will the trial take to enroll these many subjects. The **Current Calendar Time** is accrual time of the last subject in the data which is 2.224. The **Accrual Information** input is meant for simulating the additional, that is $318 - 212 = 106$ accruals. There are two options for **Input Method**. For the **Accrual Rates** option, East considers the accrual process comprised of two periods. The first period is assumed to be the one presented in the data. Starting time for this period is assumed to be 0 whereas the accrual rate for this period is computed as $(CurrentSampleSize)/(CurrentCalendarTime)$. In this case it is $212/2.2243 = 95.31$. Both the starting time and Accrual Rate fields are uneditable as these are estimated from the data. The second period is the one which starts after the last subject in the first period has arrived. As a result, the **Starting At** time default value is the last arrival time in the data which is also the **Current Calendar Time**. For the second period, the default accrual rate is computed as: $(TargetSampleSize - CurrentSampleSize)/(CurrentCalendarTime)$ which is $(318 - 212)/2.2243 = 47.65499$ for the current example. You can edit both the Starting Time and Accrual Rate for the second period. Accrual may vary over time. To reflect this assumption, one can specify the number of time periods, each having different accrual rates. Go to the **Simulation Controls** tab. Let us fix the Random Number Seed to 12345. Check the Output options, **Save the summary statistics for every simulation run** and **Save subject-level**

68 Enrollment/Events Prediction - Analysis

data for 1 simulation run. The input screen will be as shown below:

Target Sample Size:

Accrual / Dropouts | Simulation Controls

Number of Simulations:

Refresh Frequency:

Random Number Seed

Clock

Fixed

Suppress All Intermediate Output

Pause after Refresh

Stop at End

Output Options

Output Type:

Save summary statistics for every simulation run

Save subject level data for simulation runs

Note: Max. 100,000 records will be saved.

Output for All Trials

| Percentile (%) |
|----------------|
| 5.000 |
| 25.000 |
| 50.000 |
| 75.000 |
| 95.000 |

Click the button **Simulate**. East simulates the arrival of subjects according to the Uniform Arrival process. After a few seconds East will display the message that **Simulations complete. Waiting for User's action**. Click **Close**. This will save the **Predict Simulation** in the Output Preview window first. Once you save it in the Workbook, it will create a node **PredictSim1** with sub-nodes for **SummaryStat** and **SubjectData** in the **Library**.



To view the detailed summary output of the simulations, double click the node **PredictSim1** in the Library. The following output is displayed.

Conditional Simulation: Prediction of Enrollment Timelines

| Test Parameters | |
|-------------------------------|--------------|
| Simulation ID | PredictSim1 |
| Accrual | Ongoing |
| Input | Summary Data |
| Target Sample Size | 318 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 1000 |

⊖ **Actuals from Interim Trial Data: Sample Size**

| Sample Size | Current Time |
|-------------|--------------|
| 212 | 2.224 |

⊖ **Conditional Simulation: Average Sample Size**

| Average Sample Size | Average Accrual Duration |
|---------------------|--------------------------|
| 318 | 4.427 |

⊖ **Accrual / Dropouts Parameters**

Accrual Input Method: Accrual Rates

| Period # | Starting at Time | Accrual Rate |
|----------|------------------|--------------|
| 1 | 0 | 95.32374 |
| 2 | 2.224 | 47.66187 |

⊖ **Overall Output**

| | Stage | Percentile | No. of Accruals | Accrual Duration |
|---|--------------|------------|-----------------|------------------|
| | Interim | Actuals | 212 | 2.224 |
| ⊖ | End of Trial | 5% | 318 | 4.386 |
| | | 25% | 318 | 4.42 |
| | | 50% | 318 | 4.433 |
| | | 75% | 318 | 4.442 |
| | | 95% | 318 | 4.447 |
| | | Average | 318 | 4.427 |

Simulation Seed and Elapsed Time

Starting Seed: 12345
 Total Number of Simulations: 1000
 Elapsed Time: 00:00:09

The table at the left describes the Simulation scenario. This summary contains an

68 Enrollment/Events Prediction - Analysis

overview of the actual data we observed and the simulated results for the remaining arrivals in the second period. The table **Overall Output** presents the information on the percentiles of the (simulated) total accrual duration. For example, almost 50% simulations have completed by 4.433 units of time etc. The mean accrual duration of all the simulations is 4.427.

Open the **SummaryStat** data by double clicking the sub node. You will see the following display of data.

| ScenarioID | SimIndex | Stage | Status | TotAccruals | TotPendings | TotEvents | AccrDurtn |
|------------|----------|---------|---------|-------------|-------------|-----------|------------|
| 1 | 1 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 1 | Final | SUCCESS | 318 | 0 | 318 | 4.34938575 |
| 1 | 2 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 2 | Final | SUCCESS | 318 | 0 | 318 | 4.43545849 |
| 1 | 3 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 3 | Final | SUCCESS | 318 | 0 | 318 | 4.43374255 |
| 1 | 4 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 4 | Final | SUCCESS | 318 | 0 | 318 | 4.43808629 |
| 1 | 5 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 5 | Final | SUCCESS | 318 | 0 | 318 | 4.42559999 |
| 1 | 6 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 6 | Final | SUCCESS | 318 | 0 | 318 | 4.44474304 |
| 1 | 7 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 7 | Final | SUCCESS | 318 | 0 | 318 | 4.40053653 |
| 1 | 8 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 8 | Final | SUCCESS | 318 | 0 | 318 | 4.43956756 |
| 1 | 9 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 9 | Final | SUCCESS | 318 | 0 | 318 | 4.44045153 |
| 1 | 10 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 10 | Final | SUCCESS | 318 | 0 | 318 | 4.43619724 |
| 1 | 11 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 11 | Final | SUCCESS | 318 | 0 | 318 | 4.41786203 |

Observe that for every simulation East calls the **Current Sample Size** available data as **First Interim** while the **Target Sample Size** as **Final**.

The column **SUCCESS** indicates that the simulation was successful. The variable **TotEvents** is synonymous to Sample Size.

The last column **AccrDurtn** specifies the accrual duration required to enroll the 318 subjects in the respective simulation run. For instance, in the first simulation, the 318th subject arrived at the time epoch 4.34938 and so on. Now double click the **SubjectData** sub node in the library. You will see the following display of data.

| ScenarioID | SimulationID | SubjectID | ArrivalTime |
|------------|--------------|-----------|-------------|
| 1 | 1 | 213 | 2.23154614 |
| 1 | 1 | 214 | 2.23830539 |
| 1 | 1 | 215 | 2.24938829 |
| 1 | 1 | 216 | 2.25249941 |
| 1 | 1 | 217 | 2.28054916 |
| 1 | 1 | 218 | 2.30184938 |
| 1 | 1 | 219 | 2.3080157 |
| 1 | 1 | 220 | 2.31911506 |
| 1 | 1 | 221 | 2.34185009 |
| 1 | 1 | 222 | 2.35918468 |
| 1 | 1 | 223 | 2.3594013 |
| 1 | 1 | 224 | 2.37968859 |
| 1 | 1 | 225 | 2.39830855 |
| 1 | 1 | 226 | 2.40411586 |
| 1 | 1 | 227 | 2.45398704 |
| 1 | 1 | 228 | 2.47648728 |
| 1 | 1 | 229 | 2.48733738 |
| 1 | 1 | 230 | 2.51316958 |
| 1 | 1 | 231 | 2.5820502 |
| 1 | 1 | 232 | 2.62757609 |
| 1 | 1 | 233 | 2.62920135 |
| 1 | 1 | 234 | 2.64162236 |

It shows arrival times of each and every subject in the study for one simulation. This is because we have asked to save the data for one run. To view the detailed summary output of the simulations, Open the **SummaryStat** data by double clicking the sub

68 Enrollment/Events Prediction - Analysis

node. You will see the following display of data.

| ScenarioID | SimIndex | Stage | Status | TotAccruals | TotPendings | TotEvents | AccrDurtn |
|------------|----------|---------|---------|-------------|-------------|-----------|------------|
| 1 | 1 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 1 | Final | SUCCESS | 318 | 0 | 318 | 4.35719694 |
| 1 | 2 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 2 | Final | SUCCESS | 318 | 0 | 318 | 4.7111913 |
| 1 | 3 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 3 | Final | SUCCESS | 318 | 0 | 318 | 4.23758069 |
| 1 | 4 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 4 | Final | SUCCESS | 318 | 0 | 318 | 4.52927409 |
| 1 | 5 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 5 | Final | SUCCESS | 318 | 0 | 318 | 4.42808236 |
| 1 | 6 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 6 | Final | SUCCESS | 318 | 0 | 318 | 4.36122242 |
| 1 | 7 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 7 | Final | SUCCESS | 318 | 0 | 318 | 4.28554977 |
| 1 | 8 | Interim | SUCCESS | 212 | 0 | 212 | . |
| 1 | 8 | Final | SUCCESS | 318 | 0 | 318 | 4.58337695 |

Observe that for every simulation East calls the **Current Sample Size** available data as **First Interim** while the **Target Sample Size** as **Final**.

The column **SUCCESS** indicates that the simulation was successful. The variable **TotEvents** is synonymous to Sample Size.

The last column **AccrDurtn** specifies the accrual duration required to enroll the 318 subjects in the respective simulation run. For instance, in the first simulation, the 318th subject arrived at the time epoch 4.35001 and so on. Now double click the **SubjectData** sub node in the library. You will see the following display of data.

| ScenarioID | SimulationID | SubjectID | ArrivalTime |
|------------|--------------|-----------|-------------|
| 1 | 1 | 213 | 2.23154614 |
| 1 | 1 | 214 | 2.23830539 |
| 1 | 1 | 215 | 2.24938829 |
| 1 | 1 | 216 | 2.25249941 |
| 1 | 1 | 217 | 2.28054916 |
| 1 | 1 | 218 | 2.30184938 |
| 1 | 1 | 219 | 2.3080157 |
| 1 | 1 | 220 | 2.31911506 |
| 1 | 1 | 221 | 2.34185009 |
| 1 | 1 | 222 | 2.35918468 |
| 1 | 1 | 223 | 2.3594013 |
| 1 | 1 | 224 | 2.37968859 |
| 1 | 1 | 225 | 2.39830855 |
| 1 | 1 | 226 | 2.40411586 |
| 1 | 1 | 227 | 2.45398704 |
| 1 | 1 | 228 | 2.47648728 |
| 1 | 1 | 229 | 2.48733738 |
| 1 | 1 | 230 | 2.51316958 |
| 1 | 1 | 231 | 2.5820502 |
| 1 | 1 | 232 | 2.62757609 |
| 1 | 1 | 233 | 2.62920135 |
| 1 | 1 | 234 | 2.64162236 |

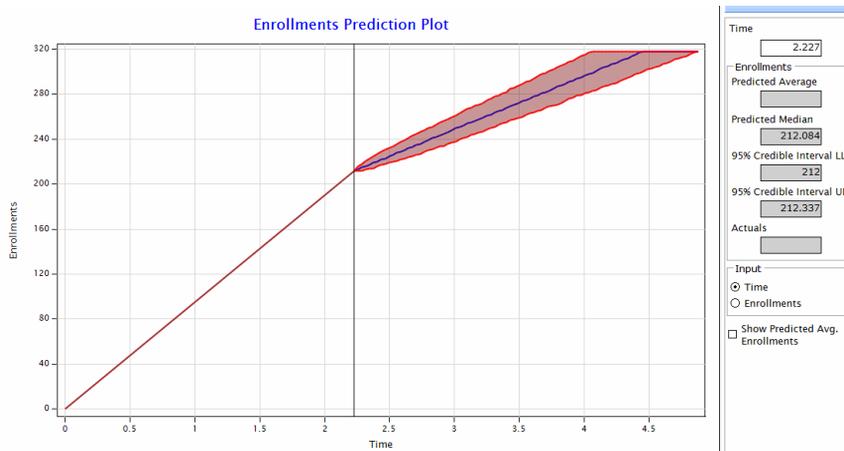
It shows arrival times for the subjects in the study for one simulation run. Note that the first subject id is 213 as the summary data input was for 212 subjects. If you scroll down, you can see that the last subject id is 318 which is the target sample size. This

68 Enrollment/Events Prediction - Analysis

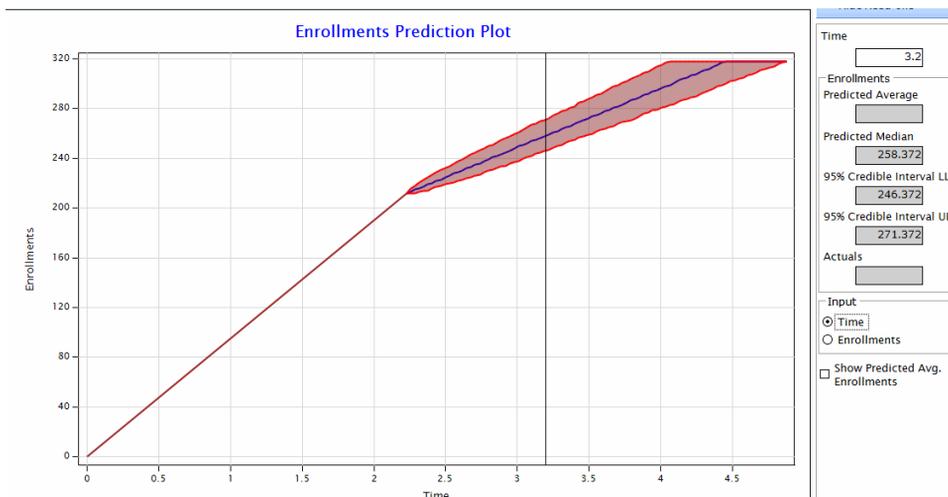
subject was accrued at 4.34938.

In addition, you can view the enrollment prediction plot using the  tool in the **Library** pane:

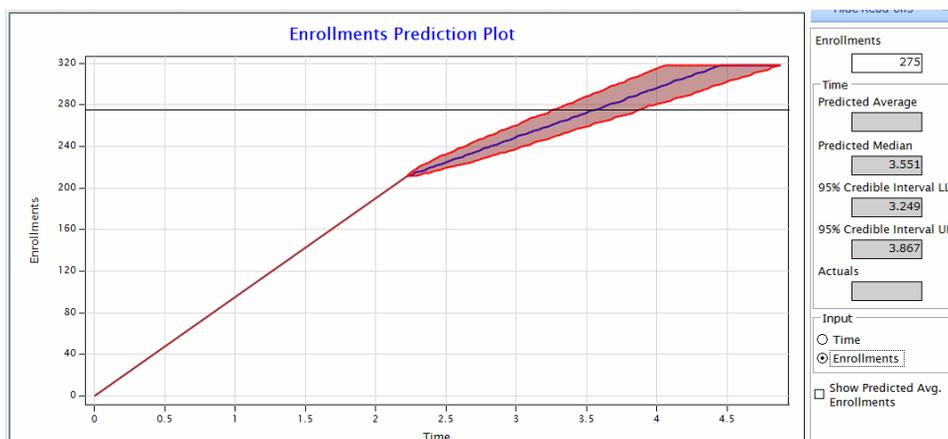
The **Enrollment Prediction Plot** displays the timeline of the observed accruals until 2.234 by which in all 212 subjects have been enrolled. This is as per the observed data.



After that point it displays the projected enrollments based on the observed accrual data we specified and the revised **Accrual Rate** in the second period. For example, at year 3.2 the predicted median enrollments reach the sample size of 258 subjects with the 95% Confidence Interval as (246.372, 271.372). Please see the plot below.



On similar lines, if you want to find out how long will it take to enroll say, 275 subjects. Select the Input > Enrollments option on the plot and type 275 in the **Enrollments** textbox as shown in the following plot.



From the read-offs it is clear that the median accrual duration for accruing 275 subjects would be 3.551 with 95% confidence interval (3.249, 3.867).

68 Enrollment/Events Prediction - Analysis

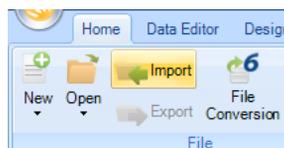
68.2 Events and Enrollment- Unblinded Data

- 68.2.1 Accrual Complete
- 68.2.2 Accrual Ongoing

A DMC statistician typically has access to unblinded trial data. In survival studies, the prediction of enrollments as well as of events is of interest. As it is important to know how long the accruals will take to complete, it is equally important to know how much time it will take to get the required number of events. With the use of predict feature, providing the inputs such as accrual rate, hazard rates, drop-out rates for the treatment and control arms you can forecast how long the subject enrolment is likely to take, and how long the trial is likely to take to complete. With PREDICT, one is able to simulate the accrual process and the follow up time, so as to predict the average accrual duration, average follow up time and average study duration (by predicting when the required number of events are likely to be achieved). We will treat the cases unblinded and blinded data separately. In unblinded situation, the user is expected to know the subject data or summary data for both the control and treatment arms separately. For instance, the control and treatment have different hazard and drop out rates and this information can be provided to East by giving different inputs for the two arms. In the case of blinded, the user is supposed to know the common hazard rate which is utilized to generate the events for control and treatment both. We will illustrate the feature with the help of Oncox (for unblinded) and Rales (for blinded) trials explained in chapter 44.

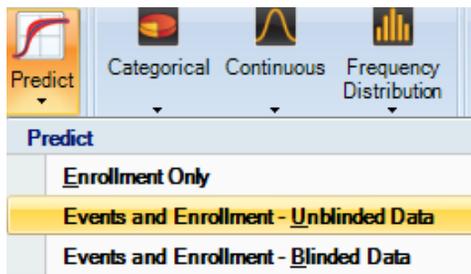
68.2.1 Events and Enrollment- Unblinded Data: Accrual Complete

Subject-level Data Assume that the study has already accrued all the subjects and we are interested in forecasting only the follow up time and study duration. The trial has accrued in all 402 subjects and stopped accruing anymore. The Subject data are available in the file **ONCOX_iLook1_SubjectData.csv** in the Samples folder of East installation directory. The data file can be imported into East using the **Import** button in the Home ribbon:



Once imported, the file will appear in the **Library** pane as nodes in the active workbook, with extension **.cydx**. Choose the menu item

Analysis>Predict>Events and Enrollment-Unblinded Data.



In the ensuing dialog box, select the **Accrual** option, **Complete**. Select data set **ONCOX_iLook1_SubjectData.cyx** . Map the variables from the data to the ones shown below:

Analysis: Predict: Events and Enrollment - Unblinded Data (Step 1 of 2)

Input: Subject-level Data Accrual: Complete
 Summary Data Ongoing

Subject-level Data

Select Workbook:

Select Dataset:

Choose Variables from Dataset

Arrival Time: Time on Study Variable:

Population ID: Status Indicator:

Control: (1 = complete, 0 = censored, -1 = dropout)

Treatment: 1

Include Site-specific Information

68 Enrollment/Events Prediction - Analysis

Click **Next**. The next input dialog appears as shown here:

Analysis: Predict: Events and Enrollment - Unblinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation **Accrual / Dropouts** Simulation Controls

Survival Information

of Hazard Pieces: Input Method:

Hazard Ratio:

Hazard Rate - Control (λ_c):

Hazard Rate - Treatment (λ_t):

The default values for **Hazard Rate - Control** and **Hazard Rate - Treatment** are estimated from the subject data. You can verify this by running the LogRank Test from **Analysis > Events > Two Samples > LogRank [SU-2S-LR]**. The input dialog for the same would be

Data Set: Oncox_EventsOnly_SubjectData.cyx

Main **Advanced**

Trial Type: Response Variable: Frequency Variable:

Population ID: Censor Indicator: Test Statistic:

Control: Censored:

Treatment: Complete:

Choose the variables as shown in the dialog. Click **OK**. A partial output is shown

below:

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | | Average Follow-up Time |
|--------------|-----------------|--------|---------|----------|---------|------------------------|
| | | Count | % | Count | % | |
| 0 | 203 | 106 | 52.217% | 97 | 47.783% | 4.814 |
| 1 | 199 | 81 | 40.704% | 118 | 59.296% | 5.004 |
| Total | 402 | 187 | 46.517% | 215 | 53.483% | 4.908 |

Test of Hypothesis:

| Log Rank Score | Std. Error | Standardized Test Statistic | (1-Sided) | | (2-Sided) |
|----------------|------------|-----------------------------|-----------|---------|-----------|
| | | | Tail | p-value | p-value |
| -13.782 | 6.819 | -2.021 | L.E. | 0.022 | 0.043 |

Parameter Estimates from Cox Model:

| Hazard Ratio (HR) | ln(Hazard Ratio) (δ) | Std. Error of Estimate of δ | Wald Statistic | 95% Confidence Interval(2-Sided) | |
|-------------------|-------------------------------|------------------------------------|----------------|----------------------------------|-------------|
| | | | | Lower Limit | Upper Limit |
| 0.743 | -0.297 | 0.148 | -2.01 | 0.556 | 0.993 |

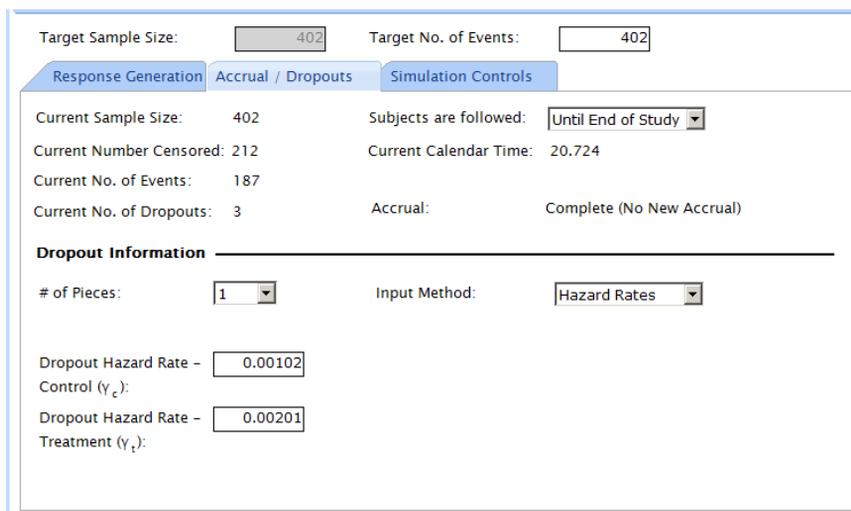
Estimated Hazard Rates:

| | |
|--------------------------------|---------|
| Control (λ_c) | 0.10847 |
| Treatment ($\lambda_c * HR$) | 0.08062 |

Observe that the **Hazard Ratio** is 0.743 and the **Estimated Hazard Rates** table gives the Hazard Rate for Control 0.10847 and that for treatment 0.08062. These are the same as the ones East chose while predicting the events. Please refer to the second input dialog for **Predict**. Continuing the **Predict** for the Oncox subject data, go to the **Accrual/ DropOuts** tab. In the ensuing dialog, you will see

68 Enrollment/Events Prediction - Analysis

almost all values filled in.



The screenshot shows a simulation control interface with the following parameters:

| | | | | | |
|--|---------|------------------------|---------------------------|---------------------|--|
| Target Sample Size: | 402 | Target No. of Events: | 402 | | |
| Response Generation | | Accrual / Dropouts | | Simulation Controls | |
| Current Sample Size: | 402 | Subjects are followed: | Until End of Study | | |
| Current Number Censored: | 212 | Current Calendar Time: | 20.724 | | |
| Current No. of Events: | 187 | Accrual: | Complete (No New Accrual) | | |
| Current No. of Dropouts: | 3 | | | | |
| Dropout Information | | | | | |
| # of Pieces: | 1 | Input Method: | Hazard Rates | | |
| Dropout Hazard Rate - Control (y_c): | 0.00102 | | | | |
| Dropout Hazard Rate - Treatment (y_t): | 0.00201 | | | | |

The **Current Calendar Time** is accrual time of the last subject in the data which is 20.724.

The drop out hazard rates for Control and Treatment are estimated from the subject data. The **Target Sample Size** is disabled as we have chosen the option **Accruals complete**. However, the **Target Number of Events** default value is 402, the same as in the data. This value can be edited. You can edit the values of hazard rates, target number of events etc. You can choose a specific follow up period as well by selecting **For Fixed Period** in the **Subjects are followed** textbox. The number of hazard pieces in the Drop out information also can be increased to specify different hazard rates for different time periods. The **Number of pieces** equal to 0 will assume that there aren't going to be any drop outs. For now, let us proceed further with all the default values. Go to **Simulation Controls**

tab. Check the Output Options for saving the outputs.

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | Simulation Controls

Number of Simulations:
 Refresh Frequency:

Random Number Seed
 Clock
 Fixed

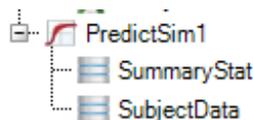
Suppress All Intermediate Output
 Pause after Refresh
 Stop at End

Output Options
 Output Type:
 Save summary statistics for every simulation run
 Save subject level data for simulation runs
 Note: Max. 100,000 records will be saved.

Output for All Trials

| Percentile (%) |
|----------------|
| 5.000 |
| 25.000 |
| 50.000 |
| 75.000 |
| 95.000 |

Click **Simulate**. East simulates the arrival of subjects according to Poisson Arrival process with inter-arrival times following exponential distribution. After a few seconds East will display the message **Simulations complete. Waiting for User's action**. Click **Close**. This will save the **Predict Simulation** in the Output Preview window first. Once you save it in the Workbook, it will create a node **PredictSim1** with sub-nodes for **SummaryStat** and **SubjectData** in the **Library**.



To view the detailed summary output of the simulations, double click the node

68 Enrollment/Events Prediction - Analysis

`PredictSim1` in the Library. The following output is displayed.

Conditional Simulation: Unblinded Prediction of Event Timelines

| Test Parameters | |
|---|--------------------|
| Simulation ID | PredictSim1 |
| Accrual | Complete |
| Input | Subject-level Data |
| Target Sample Size | 402 |
| Target No. of Events | 402 |
| Response Generation Parameters | |
| HR = λ_1/λ_2 | 0.743 |
| Hazard Rate - Control (λ_c) | 0.10847 |
| Hazard Rate - Treatment (λ_t) | 0.08062 |
| Accrual / Dropouts Parameters | |
| Subject are followed | Until End of Study |
| Dropout Hazard Rate Control | 0.00102 |
| Dropout Hazard Rate Treatment | 0.00201 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 1000 |

Actuals from Interim Trial Data: Sample Size and Events

| Sample Size | Events | | Dropouts | | Pipeline | Average Follow up | Accrual Duration |
|-------------|---------|-----------|----------|-----------|----------|-------------------|------------------|
| | Control | Treatment | Control | Treatment | | | |
| 402 | 106 | 81 | 1 | 2 | 212 | 4.908 | 20.724 |

Conditional Simulation: Average Sample Size and Events

| Average Sample Size | Average Events | | Average Dropouts | | Average Follow up | Average Study Duration |
|---------------------|----------------|-----------|------------------|-----------|-------------------|------------------------|
| | Control | Treatment | Control | Treatment | | |
| 402 | 201.111 | 194.234 | 1.889 | 4.766 | 10.609 | 86.765 |

Overall Output

| Stage | Percentile | No. of Events | No. of Accruals | No. of Dropouts | Study Duration |
|--------------|------------|---------------|-----------------|-----------------|----------------|
| Interim | Actuals | 187 | 402 | 3 | 20.724 |
| | 5% | 392 | 402 | 4 | 67.981 |
| End of Trial | 25% | 394 | 402 | 5 | 76.641 |
| | 50% | 396 | 402 | 6 | 84.38 |
| | 75% | 397 | 402 | 8 | 94.738 |
| | 95% | 398 | 402 | 10 | 111.235 |
| | Average | 395.345 | 402 | 6.655 | 86.765 |

The table at the left describes the Simulation scenario. This summary contains an overview of the actual data we observed. Since the accruals were complete, the **Target Sample Size** and the **Target Number of Events** are the same and equal to 402. The table **Actuals from Interim Trial Data: Sample Size and Events** presents the detailed information of the subject data such as events on control and treatment arms, drop outs, average follow up etc. Observe that at the end of the current time, the subjects in pipeline are 212. These are followed till the end of the study. The study is complete when all the subjects in pipeline either experience events or drop out. The table **Average Sample Size and Events** provide information about the average study duration, average number of events on control and treatment, average drop outs, average follow up time etc. From this table it should be noted that it will take on an average 86.765 units of time to complete the study. The number of events on control arm and treatment arm would be around 201 and 194. Average follow up time for an individual is 10.609. The **Overall Output** table describes the details of the distribution of **Average Study Duration** across 1000 simulations. Note that the column **No of Accruals** has all values 402 since the accruals were complete and only events are being forecasted. Since there are a few drop outs, we expect lesser, say around 395 events to occur out of 402 subjects. It is worth noting that the 5th percentile of the Average Study Duration is going to give 392 events pretty early, by the time 67.981. The 95th percentile is 111.235 which is the maximum duration the study can take. You could have changed the percentiles input, if you want to be more specific. For instance

you can input 100% to get the value of maximum study duration. Let us have a look at the individual files stored in the **Library**.

Open the **SummaryStat** data by double clicking the sub node. You will see the following display of data (shown in parts).

| ScenarioID | SimIndex | Stage | Status | Accruals0 | DropOuts0 | Pendings0 | Events0 | Accruals1 | DropOuts1 | Pendings1 | Events1 | TotAccruals |
|------------|----------|---------|---------|-----------|-----------|-----------|---------|-----------|-----------|-----------|---------|-------------|
| 1 | 1 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 | 81 | 402 |
| 1 | 1 | Final | SUCCESS | 203 | 1 | 0 | 202 | 199 | 3 | 0 | 196 | 402 |
| 1 | 2 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 | 81 | 402 |
| 1 | 2 | Final | SUCCESS | 203 | 2 | 0 | 201 | 199 | 5 | 0 | 194 | 402 |
| 1 | 3 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 | 81 | 402 |
| 1 | 3 | Final | SUCCESS | 203 | 2 | 0 | 201 | 199 | 4 | 0 | 195 | 402 |
| 1 | 4 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 | 81 | 402 |
| 1 | 4 | Final | SUCCESS | 203 | 1 | 0 | 202 | 199 | 4 | 0 | 195 | 402 |
| 1 | 5 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 | 81 | 402 |
| 1 | 5 | Final | SUCCESS | 203 | 3 | 0 | 200 | 199 | 3 | 0 | 196 | 402 |

| TotDropOuts | TotPendings | TotEvents | LookTime | AvgFollowUp | AccrDurtn |
|-------------|-------------|-----------|------------|-------------|------------|
| 3 | 212 | 187 | 20.7236842 | 4.90802566 | . |
| 4 | 0 | 398 | 118.437015 | 11.2260304 | 20.7236842 |
| 3 | 212 | 187 | 20.7236842 | 4.90802566 | . |
| 7 | 0 | 395 | 78.2068965 | 10.681094 | 20.7236842 |
| 3 | 212 | 187 | 20.7236842 | 4.90802566 | . |
| 6 | 0 | 396 | 88.3005566 | 10.7578699 | 20.7236842 |
| 3 | 212 | 187 | 20.7236842 | 4.90802566 | . |
| 5 | 0 | 397 | 94.3448636 | 10.7870712 | 20.7236842 |
| 3 | 212 | 187 | 20.7236842 | 4.90802566 | . |
| 6 | 0 | 396 | 80.5364877 | 10.6705112 | 20.7236842 |
| 3 | 212 | 187 | 20.7236842 | 4.90802566 | . |
| 7 | 0 | 395 | 74.8973821 | 10.7676946 | 20.7236842 |
| 3 | 212 | 187 | 20.7236842 | 4.90802566 | . |
| 6 | 0 | 396 | 79.3085397 | 10.3374261 | 20.7236842 |
| 3 | 212 | 187 | 20.7236842 | 4.90802566 | . |
| 11 | 0 | 391 | 88.0893703 | 10.6308622 | 20.7236842 |

Accruals0 and **Accruals1** specify the total subjects accrued on Control and Treatment arms respectively. Similar convention is used for naming the various

68 Enrollment/Events Prediction - Analysis

quantities for Control and Treatment. As before, **Interim** refers to the available data whereas **Final** refers to the simulated data. Observe that although the **TotAccruals** for every simulation is 402, the **TotEvents** may differ from simulation to simulation. Total number of events occurred in simulations 1, 2 and 3 are 398, 395 and 396 respectively. This is because of varying number of drop outs which are 4, 7 and 6 respectively for these three runs. The column **AvgFollowUp** indicates the average follow-up time of subjects by this stage, interim or Final. It is worth noting that the **LookTime** corresponding to the Final stage is essentially the study duration observed in that particular simulation. In other words, all the 402 subjects were accrued and followed in a period of 118.44 time units in simulation 1, while 78.207 for simulation 2 and so on.

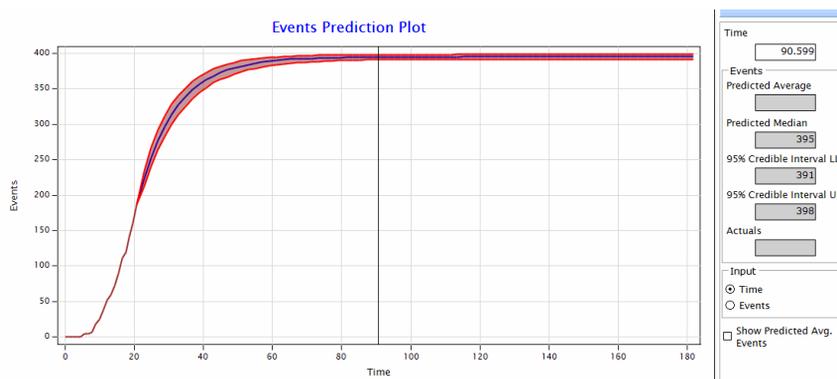
Open the **Subject Data** file which stores detailed information about one simulation.

| ScenarioID | SimulationID | SubjectID | ArrivalTime | TreatmentID | SurvivalTime | DropOutTime | CensorInd_1 |
|------------|--------------|-----------|-------------|-------------|--------------|-------------|-------------|
| 1 | 1 | 1 | 0.493421053 | 0 | 28.1516334 | 237.981976 | 1 |
| 1 | 1 | 2 | 1.08552632 | 1 | 8.32236842 | . | 1 |
| 1 | 1 | 3 | 1.5131579 | 0 | 6.74342105 | . | 1 |
| 1 | 1 | 4 | 1.64473684 | 1 | 3.75 | . | 1 |
| 1 | 1 | 5 | 1.97368421 | 0 | 5.95394737 | . | 1 |
| 1 | 1 | 6 | 2.5 | 0 | 4.93421053 | . | 1 |
| 1 | 1 | 7 | 2.56578947 | 0 | 5 | . | 1 |
| 1 | 1 | 8 | 2.89473684 | 0 | 9.30921053 | . | 1 |
| 1 | 1 | 9 | 2.96052632 | 0 | . | 14.7039474 | 0 |
| 1 | 1 | 10 | 3.125 | 1 | 11.8421053 | . | 1 |
| 1 | 1 | 11 | 3.35526316 | 1 | 21.4230398 | 136.499015 | 1 |
| 1 | 1 | 12 | 3.3881579 | 0 | 1.05263158 | . | 1 |
| 1 | 1 | 13 | 3.91447368 | 0 | 6.64473684 | . | 1 |
| 1 | 1 | 14 | 4.04605263 | 0 | 1.34868421 | . | 1 |
| 1 | 1 | 15 | 4.17763158 | 0 | 29.8693205 | 2764.51356 | 1 |
| 1 | 1 | 16 | 4.17763158 | 1 | 3.55263158 | . | 1 |

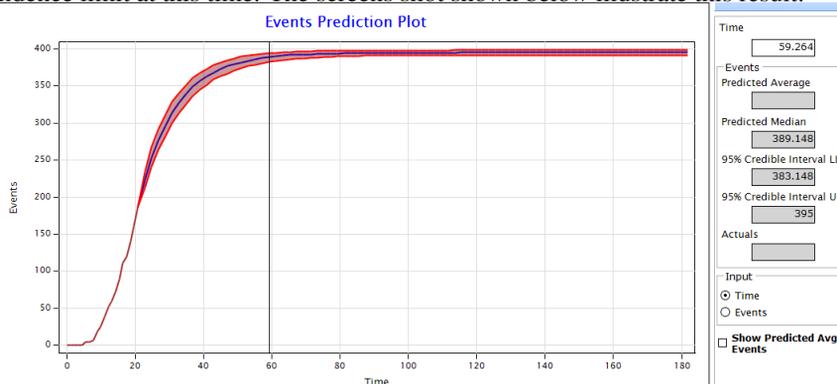
TreatmentID equal to 0 means the subject is on Control while 1 indicates Treatment. **Arrival Time** is the calendar time, **Survival Time** is the duration for which the subject was alive in the study. **DropOutTime** is the duration of time the subject was present in the study before dropping out. These are generated using the specified drop out hazard rates for control and treatment. For the first subject in the data, the **DropOutTime** is 237.9819 which is greater than the survival time. The time on study is the time subject was present in the study, which is **Accrual Time** plus **Survival Time** or **Accrual Time** plus **Drop Out Time** whichever is

minimum. In this case it is, 28.645. This means that the subject will not drop out till the study is completed. Accordingly, the value of **CensorInd_1** is 1 as it results in a complete observation for survival. On the other hand, observe that for the SubjectID 9, the **DropOutTime** is 14.7039, and survival time is not displayed. It means that the generated survival time was more than the drop out time. As a result, the drop out will happen before the event. The subject drops out resulting into a censored observation(**CensorInd_1** =0).

The **Events Prediction Plot** (invoked using the  tool in the **Library** pane) shows that the median number of events 395 are reached in a duration of 90.599 units of time.

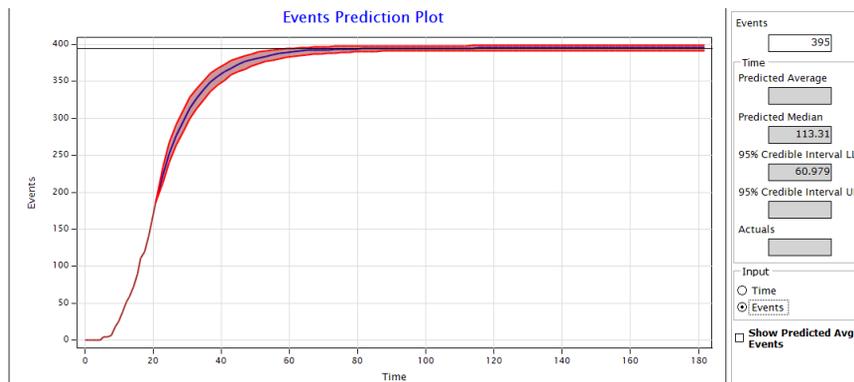


The earliest time to reach this target may be by 59.264 by looking at the upper 95% confidence limit at this time. The screens shot shown below illustrate this result.



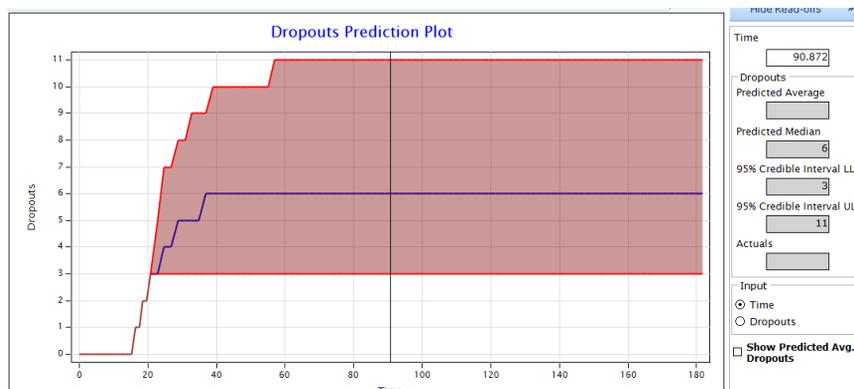
68 Enrollment/Events Prediction - Analysis

In order to find the average study duration for getting the median number of events, select the Input option **Events** in the plot. Enter 395 for Events.

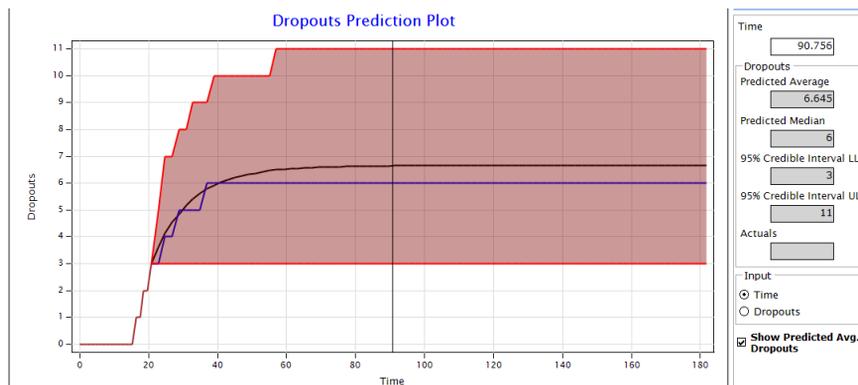


The median study duration is 113.31. Since the 95% upper limit does not exist, one can not forecast the latest time to reach the target.

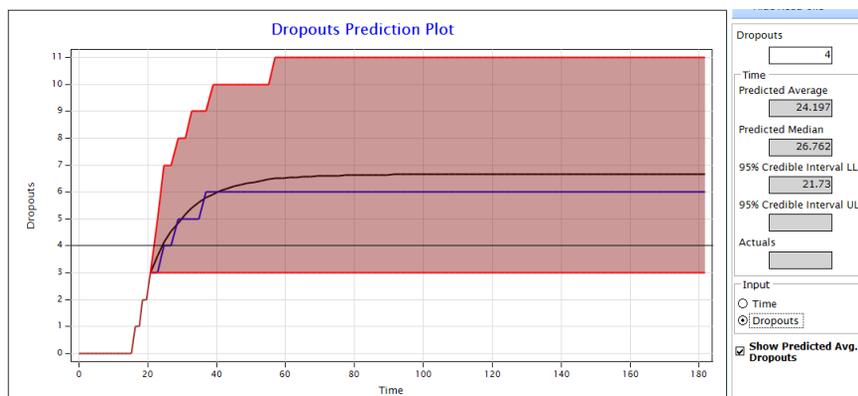
Invoke the **(Dropout Prediction Plot)** using the  tool in the **Library** pane.



In this plot, the simulation results indicate that by the end of 90.87, the median number of dropouts in both the control and treatment arms would be 6 in a 95% confidence interval of 3 to 11. If you select the **Show Predicted Avg. Dropouts** the predicted dropouts will be added to the plot.



To find out the duration by which there will be specified number of drop outs, select the **Input** option **Dropouts**. Suppose we are interested in knowing by what time there will be 4 drop outs, give this input for **Events**. Click **Enter**.



Note that the predicted median time for getting 4 dropouts is 26.762.

Summary Data In the earlier section we saw how to generate the events and follow all subjects till they experience either events or drop out. We estimated the study duration with the help of Predict feature in East. For this to use, we assumed that an interim subject-level data was available which had information on individual arrival time, status etc. However, many a times, the subject-level data may not be available. What can be available is the summary of the accruals that have happened till date. For

Click **Next**. The next input dialog appears as shown here:

Analysis: Predict: Events and Enrollment - Unblinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | Simulation Controls

Survival Information

of Hazard Pieces: Input Method:

Hazard Ratio:

Hazard Rate - Control (λ_c):

Hazard Rate - Treatment (λ_t):

The default values for **Hazard Rate - Control** and **Hazard Rate - Treatment** are shown in the dialog. Note that these are not estimated from any data as we don't have the individual subject data as input. Nonetheless, use the same hazard rates 0.10847 for Control and 0.08062 for treatment, input as in the previous section. Can we consider Target No of events as 500? If you give this input, you will get an error **Value range for target no. of events should be [188, 402]**. This is because the accruals are complete and it won't accept any further accruals. The 187 events are already occurred. Suppose you are interested in 385 events. Input this value for **Target No. of Events**

Analysis: Predict: Events and Enrollment - Unblinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | Simulation Controls

Survival Information

of Hazard Pieces: Input Method:

Hazard Ratio:

Hazard Rate - Control (λ_c):

Hazard Rate - Treatment (λ_t):

Go to the **Accruals/Dropouts** tab. Suppose instead of drop out hazard rates, the information is available on the probabilities of drop out. Suppose the probability of drop out for a subject receiving Control is 0.5% and Treatment is 0.6% and these are

68 Enrollment/Events Prediction - Analysis

applicable from the current calendar time onwards which is 20.724. Give all these inputs.

Analysis: Predict: Events and Enrollment - Unblinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | **Simulation Controls**

Current Sample Size: 402 Subjects are followed:

Current Number Censored: 212 Current Calendar Time: 20.724

Current No. of Events: 187

Current No. of Dropouts: 3 Accrual: Complete (No New Accrual)

Dropout Information

of Pieces: Input Method:

By Time:

Probability of Dropout - Control:

Probability of Dropout - Treatment:

Note: Period 1 hazard rates apply after time 20.724.

Go to **Simulation Controls** tab. Give a fixed seed 12345. Save the outputs for Summary and Subject data.

Analysis: Predict: Events and Enrollment - Unblinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | **Simulation Controls**

Number of Simulations:

Refresh Frequency:

Random Number Seed

Clock

Fixed

Suppress All Intermediate Output

Pause after Refresh

Stop at End

Output Options

Output Type:

Save summary statistics for every simulation run

Save subject level data for simulation runs

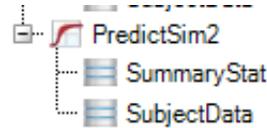
Note: Max. 100,000 records will be saved.

Output for All Trials

| Percentile (%) |
|----------------|
| 5.000 |
| 25.000 |
| 50.000 |
| 75.000 |
| 95.000 |

Click on **Simulate**. East simulates the events according to Poisson Arrival process with inter-arrival times following exponential distribution. The parameters are derived from the specified hazard rates for Control and Treatment. For details refer to the Appendix M.

After a few seconds East will display the message **Simulations complete. Waiting for User's action.** Click **Close**. This will save the **Predict Simulation** in the Output Preview window first. Once you save it in the Workbook, it will create a node **PredictSim2** with sub-nodes for **SummaryStat** and **SubjectData** in the **Library**.



To view the detailed summary output of the simulations, double click the node **PredictSim2** in the Library. The following output is displayed.

Conditional Simulation: Unblinded Prediction of Event Timelines

| Test Parameters | |
|---|--------------------|
| Simulation ID | PredictSim1 |
| Accrual | Complete |
| Input | Summary Data |
| Target Sample Size | 402 |
| Target No. of Events | 385 |
| Response Generation Parameters | |
| HR = λ_t/λ_c | 0.743 |
| Hazard Rate - Control (λ_c) | 0.108 |
| Hazard Rate - Treatment (λ_t) | 0.081 |
| Accrual / Dropouts Parameters | |
| Subject are followed | Until End of Study |
| Dropout Hazard Rate Control | 0.005 |
| Dropout Hazard Rate Treatment | 0.006 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 1000 |

⊖ **Actuals from Interim Trial Data: Sample Size and Events**

| Sample Size | Events | | Dropouts | | Pipeline | Current Time |
|-------------|---------|-----------|----------|-----------|----------|--------------|
| | Control | Treatment | Control | Treatment | | |
| 402 | 106 | 81 | 1 | 2 | 212 | 20.724 |

⊖ **Conditional Simulation: Average Sample Size and Events**

| Average Sample Size | Average Events | | Average Dropouts | | Average Study Duration |
|---------------------|----------------|-----------|------------------|-----------|------------------------|
| | Control | Treatment | Control | Treatment | |
| 402 | 197.194 | 187.097 | 5.202 | 9.92 | 70.37 |

⊖ **Overall Output**

| | Stage | Percentile | No. of Events | No. of Accruals | No. of Dropouts | Study Duration |
|---|---------|------------|---------------|-----------------|-----------------|----------------|
| ⊖ | Interim | Actuals | 187 | 402 | 3 | 20.724 |
| | | 5% | 381 | 402 | 9 | 52.755 |
| | 25% | 384 | 402 | 13 | 59.574 | |
| | 50% | 385 | 402 | 15 | 66.732 | |
| | 75% | 385 | 402 | 18 | 77.774 | |
| | 95% | 385 | 402 | 21 | 98.407 | |
| | Average | 384.291 | 402 | 15.122 | 70.37 | |

Simulation Seed and Elapsed Time

Starting Seed: 12345

The table at the left describes the Simulation scenario. This summary contains an overview of the actual data we observed. Since the accruals were complete, the **Target Sample Size** and the **Target Number of Events** are 402 and 385 respectively. The table **Actuals from Interim Trial Data: Sample**

68 Enrollment/Events Prediction - Analysis

Size and Events presents the detailed information of the subject data such as events on control and treatment arms, drop outs, average follow up etc. Observe that at the end of the current time, the subjects in pipeline are 212. These are followed till the end of the study. The study is complete when in all 385 events occur. A few simulations may give lesser number of events as there can be more drop outs. The table **Average Sample Size and Events** provide information about the average study duration, average number of events on control and treatment, average drop outs, average follow up time etc. From this table it should be noted that it will take on an average 50.946 units of time to complete the study. The number of events on control arm and treatment arm would be around 197 and 187. The **Overall Output** table describes the details of the distribution of **Average Study Duration** across 1000 simulations. Note that the column **No of Accruals** has all values 402 since the accruals were complete and only events are being forecasted. Since the targeted number of events was 385, the **No of Events** column show the value 385 in almost all the cases. It is worth noting that the 5th percentile of the Average Study Duration is going to give 381 events pretty early, by the time 52.755. The 95th percentile is 98.407 by which almost in all cases the target would be achieved. Let us have a look at the individual files stored in the **Library**.

Open the **SummaryStat** data by double clicking the sub node. You will see the

following display of data (shown in parts).

| ScenarioID | SimIndex | Stage | Status | Accruals0 | DropOuts0 | Pendings0 | Events0 | Accruals1 | DropOuts1 | Pendings1 |
|------------|----------|---------|---------|-----------|-----------|-----------|---------|-----------|-----------|-----------|
| 1 | 1 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 |
| 1 | 1 | Final | SUCCESS | 203 | 6 | 1 | 196 | 199 | 7 | 3 |
| 1 | 2 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 |
| 1 | 2 | Final | SUCCESS | 203 | 6 | 0 | 197 | 199 | 14 | 0 |
| 1 | 3 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 |
| 1 | 3 | Final | SUCCESS | 203 | 8 | 1 | 194 | 199 | 6 | 2 |
| 1 | 4 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 |
| 1 | 4 | Final | SUCCESS | 203 | 4 | 2 | 197 | 199 | 8 | 3 |
| 1 | 5 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 |
| 1 | 5 | Final | SUCCESS | 203 | 4 | 0 | 199 | 199 | 17 | 0 |
| 1 | 6 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 |
| 1 | 6 | Final | SUCCESS | 203 | 8 | 0 | 195 | 199 | 6 | 3 |
| 1 | 7 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 |
| 1 | 7 | Final | SUCCESS | 203 | 5 | 1 | 197 | 199 | 9 | 2 |
| 1 | 8 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 |
| 1 | 8 | Final | SUCCESS | 203 | 6 | 0 | 197 | 199 | 13 | 0 |
| 1 | 9 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 |
| 1 | 9 | Final | SUCCESS | 203 | 8 | 0 | 195 | 199 | 11 | 0 |
| 1 | 10 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 116 |

| | | | | | |
|-----|-----|----|-----|-----|------------|
| 81 | 402 | 3 | 212 | 187 | 20.7236 |
| 189 | 402 | 13 | 4 | 385 | 63.0960116 |
| 81 | 402 | 3 | 212 | 187 | 20.7236 |
| 185 | 402 | 20 | 0 | 382 | 78.2034628 |
| 81 | 402 | 3 | 212 | 187 | 20.7236 |
| 191 | 402 | 14 | 3 | 385 | 59.8274803 |
| 81 | 402 | 3 | 212 | 187 | 20.7236 |
| 188 | 402 | 12 | 5 | 385 | 66.2361613 |
| 81 | 402 | 3 | 212 | 187 | 20.7236 |
| 182 | 402 | 21 | 0 | 381 | 80.5329182 |
| 81 | 402 | 3 | 212 | 187 | 20.7236 |
| 190 | 402 | 14 | 3 | 385 | 66.0449292 |
| 81 | 402 | 3 | 212 | 187 | 20.7236 |
| 188 | 402 | 14 | 3 | 385 | 65.8378117 |
| 81 | 402 | 3 | 212 | 187 | 20.7236 |
| 186 | 402 | 19 | 0 | 383 | 88.0853607 |

Accruals0 and **Accruals1** specify the total subjects accrued on Control and Treatment arms respectively. Similar convention is used for naming the various quantities for Control and Treatment. As before, **Interim** refers to the available data whereas **Final** refers to the simulated data. Observe that although the **TotAccruals** for every simulation is 402, and the **TotEvents** is 385. The **TotPending** values for **Final Look** are the subjects which have neither experienced events nor have dropped out till the end of the study. This is because the study is concluded after getting 385 events and does not proceed till all the subjects experience the event as was the case in the previous section. It is worth noting that the **LookTime**

68 Enrollment/Events Prediction - Analysis

corresponding to the Final stage is essentially the study duration observed in that particular simulation.

Open the **Subject Data** file which stores detailed information about one simulation.

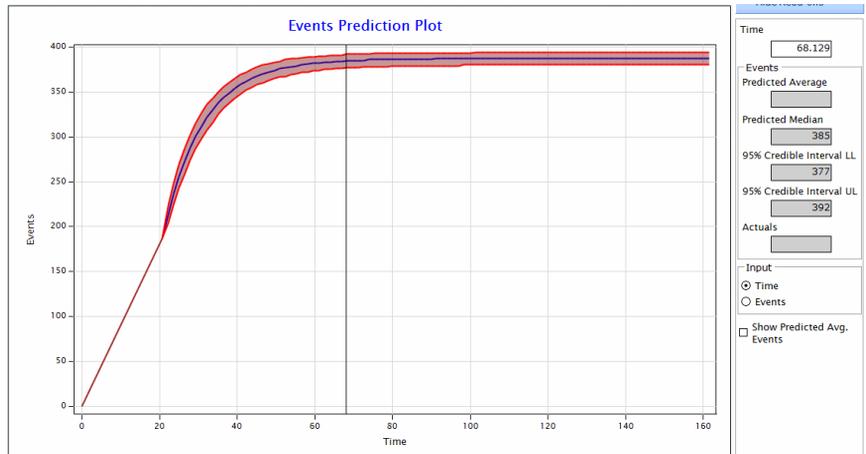
| ScenarioID | SimulationID | SubjectID | TreatmentID | SurvivalTime | DropOutTime | CensorInd_1 |
|------------|--------------|-----------|-------------|--------------|-------------|-------------|
| 1 | 1 | 191 | 1 | 0.0554131358 | 32.3599823 | 1 |
| 1 | 1 | 192 | 1 | 0.0740037415 | 177.77468 | 1 |
| 1 | 1 | 193 | 0 | 0.2077833 | 37.4806914 | 1 |
| 1 | 1 | 194 | 0 | 0.371328902 | 759.453399 | 1 |
| 1 | 1 | 195 | 1 | 0.434490088 | 165.92292 | 1 |
| 1 | 1 | 196 | 0 | 0.439958267 | 7.93860821 | 1 |
| 1 | 1 | 197 | 1 | 0.480387907 | 139.12904 | 1 |
| 1 | 1 | 198 | 1 | 0.501063968 | 62.9622212 | 1 |
| 1 | 1 | 199 | 1 | 0.588871877 | 318.580163 | 1 |
| 1 | 1 | 200 | 0 | 0.637448719 | 269.040631 | 1 |
| 1 | 1 | 201 | 0 | 0.686826927 | 240.852533 | 1 |
| 1 | 1 | 202 | 1 | 0.728121285 | 3.20707008 | 1 |
| 1 | 1 | 203 | 1 | 0.800459118 | 117.999072 | 1 |
| 1 | 1 | 204 | 0 | 0.818479618 | 111.963745 | 1 |
| 1 | 1 | 205 | 0 | 0.910754891 | 506.907131 | 1 |
| 1 | 1 | 206 | 1 | 1.04244909 | 518.242629 | 1 |

The **SubjectID** starts from 191 as there were 187 events occurred earlier and 3 had dropped out. For the initial 190 subjects the detail information such as arrival time, drop out time etc is not available. **TreatmentID** 0 means the subject is on Control while 1 indicates Treatment. For the 191 subject onwards, the survival times and drop out times are generated. **Survival Time** is the duration for which the subject was alive in the study. **DropOutTime** is the duration of time the subject was present in the study before dropping out. These are generated using the specified drop out probabilities for control and treatment. Note that the data is sorted on Survival Times. Key points to observe:

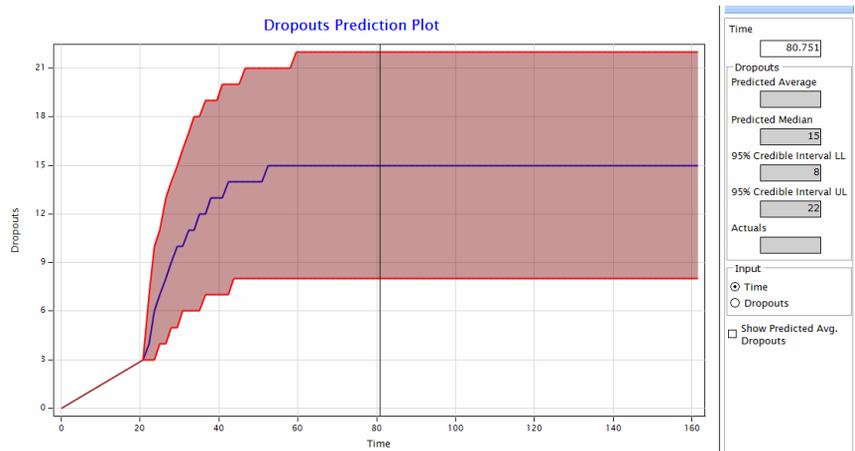
- Since out of targeted 385 events, 187 were observed earlier, the required number of events is essentially 198.
- Subject 289 drops out as its generated survival time is greater than its drop out time.
- The subjects having SubjectID 401 and 402 are not followed as the requirement of 385 events has been satisfied. They just form the group of **pending** subjects which are 4 in number.
- For the subjects which are either dropped out or form a pending observation, the

value of **CensorInd_1** is 0.

The **Events Prediction Plot** (invoked using the  tool in the **Library** pane) shows that the median number of events 385 are reached in a duration of 68.129 units of time.



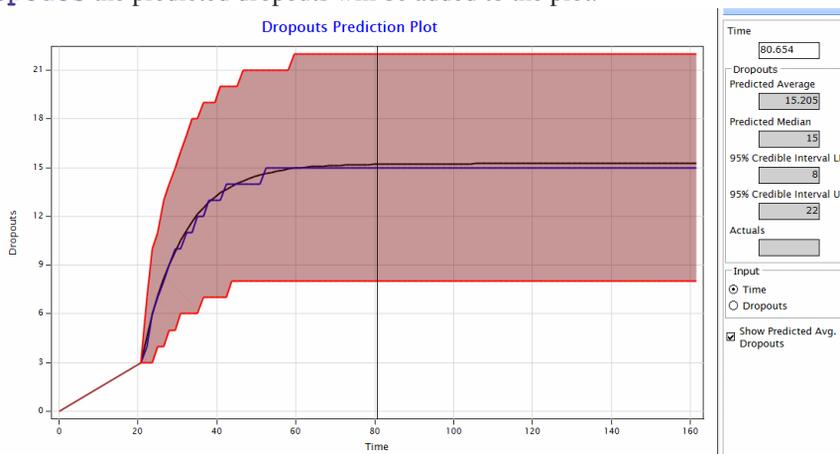
Invoke the **(Dropout Prediction Plot)** using the  tool in the **Library** pane.



In this plot, the simulation results indicate that by the end of 80.751, the median

68 Enrollment/Events Prediction - Analysis

number of dropouts in both the control and treatment arms would be 15 in a 95% confidence interval of 8 to 22. If you select the **Show Predicted Avg. Dropouts** the predicted dropouts will be added to the plot.



68.2.2 Events and Enrollment- Unblinded Data: Accrual Ongoing

Subject-level Data The ONCOX trial in the earlier chapters was designed with a sample size of 460 subjects with an expected accrual period of around 24 months and targeted 374 events within a study period of around 30 months. Assume that an interim look has been taken and subject data are available at this time point. The trial is still accruing subjects and we are interested in forecasting Accrual duration as well as the Study Duration. The trial has accrued in all 402 subjects so far. The Subject data are available in the file **ONCOX_iLook1_SubjectData.csv** in the Samples folder of East installation directory. The file ONCOX_iLook1_SubjectData contains a list of subjects accrued so far and the following data for each subject:

- Country: Country ID.
- SiteID: the site at which the subject arrived.
- ArrivalTime: the time at which the subject arrived.
- TreatmentID: a variable indicating which group the subject was randomized to ('1' for treatment, '0' for placebo).
- TimeOnStudy: the length of time the subject has been in the study, corresponding to survival time.
- Status: a variable indicating whether the subject is a completer ('1'), a dropout

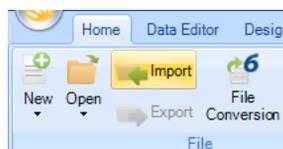
(‘-1’) or in the pipeline (‘0’).

- Censor: a variable indicating whether the subject is a completer (‘1’) or a non-completer (‘0’). A non-completer can be either a dropout or in the pipeline.

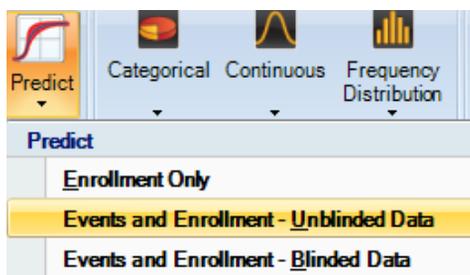
The trial is assumed to enroll subjects from several sites, the information about which is provided in the file **ONCOX_iLook1_SiteData** also available in **Samples** folder. The file contains the following data for each site:

- Country: Country ID.
- SiteID: the identification number of the site.
- SiteReadyTime: the time at which the site was initiated.
- SiteAccrRate: the site accrual rate specified in the enrollment plan.
- SubjectsAccrued: the number of subjects accrued at the site.
- LastSubjectRand: the randomization time of the last subject arriving at the site.
- ObsrvdAccrualRate: the observed accrual rate at the site.
- PosteriorAccrualRate: the updated site accrual rate.
- SIP_Start: the start of the initiation period of the site.
- SIP_End: the end of the initiation period of the site.
- Ecap: the enrollment cap, representing the maximum number of subjects that can be enrolled at the site.

Both these files can be imported into East using the **Import** button in the Home ribbon:



Once imported, the files will appear in the **Library** pane as nodes in the active workbook, with extension .cydx. Choose the menu item **Analysis>Predict>Events and Enrollment-Unblinded Data**.



68 Enrollment/Events Prediction - Analysis

In the ensuing dialog box, select the **Accrual** option, **Ongoing**. Select data set **ONCOX_iLook1_SubjectData.cydx** . Tick the check box **Include Site-specific Information**. Map the variables from the data to the ones shown below:

Analysis: Predict: Events and Enrollment - Unblinded Data (Step 1 of 2)
? 00

| | | | |
|--------|---|----------|--|
| Input: | <input checked="" type="radio"/> Subject-level Data | Accrual: | <input type="radio"/> Complete |
| | <input type="radio"/> Summary Data | | <input checked="" type="radio"/> Ongoing |

- | Subject-level Data

| | | |
|------------------|--|--------------|
| Select Workbook: | <input type="text" value="Wbk1"/> | View Dataset |
| Select Dataset: | <input type="text" value="ONCOX_iLook1_SubjectData.cydx"/> | |

Choose Variables from Dataset

| | | | |
|----------------|--|--|--|
| Arrival Time: | <input type="text" value="ArrivalTime"/> | Time on Study Variable: | <input type="text" value="TimeOnStudy"/> |
| Population ID: | <input type="text" value="Trtmt"/> | Status Indicator: | <input type="text" value="Status"/> |
| Control: | <input type="text" value="0"/> | (1 = complete, 0 = censored, -1 = dropout) | |
| Treatment: | 1 | Site ID: | <input type="text" value="Site_Id"/> |

Include Site-specific Information

Include Site-specific Information

- | Site-level Data

| | | |
|------------------|---|--------------|
| Select Workbook: | <input type="text" value="Wbk1"/> | View Dataset |
| Select Dataset: | <input type="text" value="ONCOX_iLook1_SiteData.cydx"/> | |

Choose Variables from Dataset

| | | | |
|----------------------|--|--|---|
| Site ID: | <input type="text" value="Site_Id"/> | Site Initiation: Unopened Sites | |
| Accrual Rate / Site: | <input type="text" value="ObservedEnrolRate"/> | Start Time: | <input type="text" value="SIP_Start"/> |
| Enrollment Cap: | <input type="text" value="EnrolCap"/> | End Time: | <input type="text" value="SIP_End"/> |
| | | Site Initiation: Opened Sites | |
| | | Site Initiation Time: | <input type="text" value="ActivationTime"/> |

Next >>

1738

68.2 Events and Enrollment- Unblinded Data – 68.2.2 Accrual Ongoing

Click **Next**. The next input dialog appears as shown here:

Analysis: Predict: Events and Enrollment - Unblinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | Simulation Controls

Survival Information

of Hazard Pieces: Input Method:

Hazard Ratio:

Hazard Rate - Control (λ_c):

Hazard Rate - Treatment (λ_t):

Subject-level Data: ONCOX_iLook1_SubjectData.cydx
 Site-level Data: ONCOX_iLook1_SiteData.cydx

<< Back Simulate

The default values for **Hazard Rate - Control** and **Hazard Rate -Treatment** are estimated from the subject data. You can verify this by running the LogRank Test from **Analysis > Events > Two Samples > LogRank [SU-2S-LR]**. The input dialog for the same would be

Data Set: Oncox_EventsOnly_SubjectData.cydx

Main | **Advanced**

Trial Type: Response Variable: Frequency Variable:

Population ID: Censor Indicator: Test Statistic:

Control: Censored:

Treatment: Complete:

Choose the variables as shown in the dialog. Click **OK**. A partial output is shown

68 Enrollment/Events Prediction - Analysis

below:

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | | Average Follow-up Time |
|--------------|-----------------|--------|---------|----------|---------|------------------------|
| | | Count | % | Count | % | |
| 0 | 203 | 106 | 52.217% | 97 | 47.783% | 4.814 |
| 1 | 199 | 81 | 40.704% | 118 | 59.296% | 5.004 |
| Total | 402 | 187 | 46.517% | 215 | 53.483% | 4.908 |

Test of Hypothesis:

| Log Rank Score | Std. Error | Standardized Test Statistic | (1-Sided) | | (2-Sided) |
|----------------|------------|-----------------------------|-----------|---------|-----------|
| | | | Tail | p-value | p-value |
| -13.782 | 6.819 | -2.021 | L.E. | 0.022 | 0.043 |

Parameter Estimates from Cox Model:

| Hazard Ratio (HR) | ln(Hazard Ratio) (δ) | Std. Error of Estimate of δ | Wald Statistic | 95% Confidence Interval(2-Sided) | |
|-------------------|-------------------------------|------------------------------------|----------------|----------------------------------|-------------|
| | | | | Lower Limit | Upper Limit |
| 0.743 | -0.297 | 0.148 | -2.01 | 0.556 | 0.993 |

Estimated Hazard Rates:

| | |
|--------------------------------|---------|
| Control (λ_c) | 0.10847 |
| Treatment ($\lambda_c * HR$) | 0.08062 |

Observe that the **Hazard Ratio** is 0.743 and the **Estimated Hazard Rates** table gives the Hazard Rate for Control 0.10847 and that for treatment 0.08062. These are the same as the ones East chose while predicting the events. Please refer to the second input dialog for **Predict**. Continuing the **Predict** for the Oncox subject data, go to the **Accrual/ Dropouts** tab. In the ensuing dialog, you will see almost all values filled in. Change the **Accrual Model** to **Poisson**.

Analysis: Predict: Events and Enrollment - Unblinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | **Simulation Controls**

Current Sample Size: 402 Subjects are followed:

Current Number Censored: 212 Current Calendar Time: 20.724

Current No. of Events: 187

Current No. of Dropouts: 3 Accrual Model:

Total Number of Sites: Number of Sites Opened: 118

| Site ID | Site Initiation Period | | Accrual Rate/Site | Enrollment Cap | Planned Accrual Rate/Site | Site Initiation Time | No. of Accruals |
|---------|------------------------|-----|-------------------|----------------|---------------------------|----------------------|-----------------|
| | Start | End | | | | | |
| 601 | NA | NA | 0.15315 | 7 | 0.15315 | 7.664 | 2 |
| 602 | NA | NA | 0.12614 | 27 | 0.12614 | 4.868 | 2 |
| 603 | NA | NA | 0.32513 | 45 | 0.32513 | 8.421 | 4 |
| 604 | NA | NA | 0.18131 | 9 | 0.18131 | 4.178 | 3 |
| 605 | NA | NA | 0.06468 | 7 | 0.06468 | 5.263 | 1 |
| 606 | NA | NA | 0.35349 | 30 | 0.35349 | 6.579 | 5 |
| 607 | NA | NA | 0.47377 | 10 | 0.47377 | 8.059 | 6 |
| 608 | NA | NA | 0.56296 | 9 | 0.56296 | 10.066 | 0 |
| 609 | NA | NA | 4.44878 | 12 | 4.44878 | 19.375 | 0 |

Dropout Information

of Pieces: Input Method:

Dropout Hazard Rate - Control (γ_c):

Dropout Hazard Rate - Treatment (γ_t):

Subject-level Data: ONCOX_ILook1_SubjectData.cydx
Site-level Data: ONCOX_ILook1_SiteData.cydx

The **Current Calendar Time** is accrual time of the last subject in the data which is 20.724.

The drop out hazard rates for Control and Treatment are estimated from the subject data. The **Target Sample Size** is 603 which is $1.5 * SampleSize$. You are free to change the **Target Sample Size** as we are assuming that the study is still accepting enrolments. The **Target Number of Events** default value is 402, the same as in the data. This value can be edited. You can edit the values of hazard rates, target number of events etc. You can choose a specific follow up period as well by selecting **For Fixed Period** in the **Subjects are followed** textbox. The number of hazard pieces in the Drop out information also can be increased to specify different hazard rates for different time periods. The **Number of pieces** equal to

68 Enrollment/Events Prediction - Analysis

0 will assume that there aren't going to be any drop outs. For now, let us proceed further with all the default values. Go to **Simulation Controls** tab. Give a fixed seed 12345. Check the Output Options for saving the outputs.

Target Sample Size: Target No. of Events:

Response Generation | Accrual / Dropouts | **Simulation Controls**

Number of Simulations:
 Refresh Frequency:

Random Number Seed
 Clock
 Fixed

Suppress All Intermediate Output
 Pause after Refresh
 Stop at End

Output Options
 Output Type:
 Save summary statistics for every simulation run
 Save subject level data for simulation runs
 Save site-wise summary for every simulation run
 Save site parameters data for simulation runs
 Note: Max. 100,000 records will be saved.

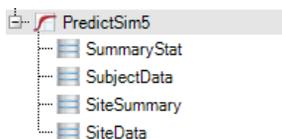
Output for All Trials

| Percentile (%) |
|----------------|
| 5.000 |
| 25.000 |
| 50.000 |
| 75.000 |
| 95.000 |

Subject-level Data: ONCOX_iLook1_SubjectData.cydx
 Site-level Data: ONCOX_iLook1_SiteData.cydx

<< Back Simulate

Click **Simulate**. East simulates the arrival of subjects according to Poisson Arrival process with inter-arrival times following exponential distribution. After a few seconds East will display the message **Simulations complete. Waiting for User's action**. Click **Close**. This will save the **Predict Simulation** in the Output Preview window first. Once you save it in the Workbook, it will create a node **PredictSim1** with sub-nodes for **SummaryStat**, **SubjectData**, **SiteSummary** and **SitePara** in the **Library**.



To view the detailed summary output of the simulations, double click the node **PredictSim1** in the Library. The following output is displayed.

Conditional Simulation: Unblinded Prediction of Enrollment and Event Timelines

| Test Parameters | |
|---|--------------------|
| Simulation ID | PredictSim5 |
| Accrual | Ongoing |
| Input | Subject-level Data |
| Include Site Info | Yes |
| Target Sample Size | 603 |
| Target No. of Events | 402 |
| Response Generation Parameters | |
| HR = λ_c/λ_t | 0.743 |
| Hazard Rate - Control (λ_c) | 0.10847 |
| Hazard Rate - Treatment (λ_t) | 0.08062 |
| Accrual / Dropouts Parameters | |
| Subject are followed | Until End of Study |
| Accrual Model | Poisson |
| Total No. of Sites | 127 |
| Dropout Hazard Rate Control | 0.00102 |
| Dropout Hazard Rate Treatment | 0.00201 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 1000 |

Actuals from Interim Trial Data: Sample Size and Events

| Sample Size | Events | | Dropouts | | Average Follow up | Accrual Duration | Current Time |
|-------------|---------|-----------|----------|-----------|-------------------|------------------|--------------|
| | Control | Treatment | Control | Treatment | | | |
| 402 | 106 | 61 | 1 | 2 | 4.908 | 20.724 | 20.724 |

Conditional Simulation: Average Sample Size and Events

| Average Sample Size | Average Events | | Average Dropouts | | Average Follow up | Average Accrual Duration | Average Study Duration |
|---------------------|----------------|-----------|------------------|-----------|-------------------|--------------------------|------------------------|
| | Control | Treatment | Control | Treatment | | | |
| 603 | 218.803 | 183.197 | 2.101 | 4.547 | 7.102 | 26.009 | 30.027 |

Overall Output

| Stage | Percentile | No. of Events | No. of Sites Opened | No. of Accruals | No. of Dropouts | Accrual Duration | Study Duration |
|--------------|------------|---------------|---------------------|-----------------|-----------------|------------------|----------------|
| Interim | Actuals | 187 | 0 | 402 | 3 | 20.724 | 20.724 |
| | 5% | 402 | 127 | 603 | 4 | 25.384 | 29.109 |
| End of Trial | 25% | 402 | 127 | 603 | 5 | 25.744 | 29.632 |
| | 50% | 402 | 127 | 603 | 7 | 25.987 | 30.012 |
| | 75% | 402 | 127 | 603 | 8 | 26.259 | 30.424 |
| | 95% | 402 | 127 | 603 | 10 | 26.678 | 30.932 |
| | Average | 402 | 127 | 603 | 6.648 | 26.009 | 30.027 |

The table at the left describes the Simulation scenario. This summary contains an overview of the actual data we observed. The table **Actuals from Interim Trial Data: Sample Size and Events** presents the detailed information of the subject data such as events on control and treatment arms, drop outs, average follow up etc. The table **Average Sample Size and Events** provide information about the average study duration, average number of events on control and treatment, average drop outs, average follow up time etc. From this table it should be noted that it will take on an average 30 units of time to complete the study. The number of events on control arm and treatment arm would be around 219 and 183. Average follow up time for an individual is 7.102. The **Average Accrual Duration** is 26. The **Overall Output** table describes the details of the distribution of **Accrual Duration** and **Study Duration** across 1000 simulations. It is worth noting that the 5th percentile of the Average Study Duration is going to give 402 events, by the time 29.109. The 95th percentile is 30.931 which is the maximum duration the study can take. You could have changed the percentiles input, if you want to be more specific. For instance you can input 100% to get the value of maximum study duration. Let us have a look at the individual files stored in the **Library**. Open the **SummaryStat** data by double clicking the sub node. You

68 Enrollment/Events Prediction - Analysis

will see the following display of data (shown in parts).

| ScenarioID | SimIndex | Stage | Status | Accruals0 | DropOuts0 | Pendings0 | Events0 | Accruals1 | DropOuts1 | Pendings |
|------------|----------|---------|---------|-----------|-----------|-----------|---------|-----------|-----------|----------|
| 1 | 1 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 1 |
| 1 | 1 | Final | SUCCESS | 304 | 1 | 95 | 208 | 299 | 4 | 1 |
| 1 | 2 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 1 |
| 1 | 2 | Final | SUCCESS | 304 | 3 | 82 | 219 | 299 | 5 | 1 |
| 1 | 3 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 1 |
| 1 | 3 | Final | SUCCESS | 304 | 1 | 80 | 223 | 299 | 3 | 1 |
| 1 | 4 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 1 |
| 1 | 4 | Final | SUCCESS | 304 | 2 | 90 | 212 | 299 | 3 | 1 |
| 1 | 5 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 1 |
| 1 | 5 | Final | SUCCESS | 304 | 1 | 84 | 219 | 299 | 4 | 1 |
| 1 | 6 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 1 |
| 1 | 6 | Final | SUCCESS | 304 | 2 | 89 | 213 | 299 | 5 | 1 |
| 1 | 7 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 1 |
| 1 | 7 | Final | SUCCESS | 304 | 2 | 84 | 218 | 299 | 7 | 1 |
| 1 | 8 | Interim | SUCCESS | 203 | 1 | 96 | 106 | 199 | 2 | 1 |
| 1 | 8 | Final | SUCCESS | 304 | 3 | 88 | 213 | 299 | 6 | 1 |

| Events1 | TotAccruals | TotDropOuts | TotPendings | TotEvents | LookTime | AvgFollowUp | AccrDurn |
|---------|-------------|-------------|-------------|-----------|------------|-------------|------------|
| 81 | 402 | 3 | 212 | 187 | 20.7236842 | 4.90802566 | .. |
| 194 | 603 | 5 | 196 | 402 | 30.0946311 | 7.18652332 | 25.6111911 |
| 81 | 402 | 3 | 212 | 187 | 20.7236842 | 4.90802566 | .. |
| 183 | 603 | 8 | 193 | 402 | 30.7871927 | 7.56158411 | 26.106917 |
| 81 | 402 | 3 | 212 | 187 | 20.7236842 | 4.90802566 | .. |
| 179 | 603 | 4 | 197 | 402 | 29.6080451 | 6.99869051 | 25.7137251 |
| 81 | 402 | 3 | 212 | 187 | 20.7236842 | 4.90802566 | .. |
| 190 | 603 | 5 | 196 | 402 | 30.8617163 | 7.33841465 | 27.1143374 |
| 81 | 402 | 3 | 212 | 187 | 20.7236842 | 4.90802566 | .. |
| 183 | 603 | 5 | 196 | 402 | 29.9861206 | 7.1440437 | 25.7681219 |
| 81 | 402 | 3 | 212 | 187 | 20.7236842 | 4.90802566 | .. |
| 189 | 603 | 7 | 194 | 402 | 29.3962969 | 6.82361746 | 25.8851163 |
| 81 | 402 | 3 | 212 | 187 | 20.7236842 | 4.90802566 | .. |
| 184 | 603 | 9 | 192 | 402 | 30.5350883 | 7.23252928 | 26.1944495 |
| 81 | 402 | 3 | 212 | 187 | 20.7236842 | 4.90802566 | .. |
| 189 | 603 | 9 | 192 | 402 | 30.5046809 | 7.28711001 | 26.0107641 |

Accruals0 and **Accruals1** specify the total subjects accrued on Control and Treatment arms respectively. Similar convention is used for naming the various quantities for Control and Treatment. As before, **Interim** refers to the available data whereas **Final** refers to the simulated data. Observe that although the **TotAccruals** for every simulation is 603, and the **TotEvents** is 402. The **TotPending** values for **Final Look** are the subjects which have neither experienced events nor have dropped out till the end of the study. This is because the study is concluded after getting 402 events and does not proceed till all the subjects experience the event. Note that the **LookTime** corresponding to the Final stage is essentially the study duration observed in that particular simulation.

Open the **Subject Data** file which stores detailed information about one

simulation.

| ScenarioID | SimulationID | SubjectID | SiteID | ArrivalTime | TreatmentID | SurvivalTime | DropOutTime | CensorInd_1 |
|------------|--------------|-----------|--------|-------------|-------------|--------------|-------------|-------------|
| 1 | 1 | 1 | 103 | 0.493421053 | 0 | 39.5568914 | 2784.09113 | 0 |
| 1 | 1 | 2 | 103 | 1.08552632 | 1 | 8.32236842 | . | 1 |
| 1 | 1 | 3 | 108 | 1.5131579 | 0 | 6.74342105 | . | 1 |
| 1 | 1 | 4 | 108 | 1.64473684 | 1 | 3.75 | . | 1 |
| 1 | 1 | 5 | 108 | 1.97368421 | 0 | 5.95394737 | . | 1 |
| 1 | 1 | 6 | 103 | 2.5 | 0 | 4.93421053 | . | 1 |
| 1 | 1 | 7 | 103 | 2.56578947 | 0 | 5 | . | 1 |
| 1 | 1 | 8 | 104 | 2.89473684 | 0 | 9.30921053 | . | 1 |
| 1 | 1 | 9 | 106 | 2.96052632 | 0 | . | 14.7039474 | 0 |
| 1 | 1 | 10 | 103 | 3.125 | 1 | 11.8421053 | . | 1 |
| 1 | 1 | 11 | 103 | 3.35526316 | 1 | 17.6680865 | 709.599668 | 1 |
| 1 | 1 | 12 | 106 | 3.3881579 | 0 | 1.05263158 | . | 1 |
| 1 | 1 | 13 | 121 | 3.91447368 | 0 | 6.64473684 | . | 1 |
| 1 | 1 | 14 | 121 | 4.04605263 | 0 | 1.34868421 | . | 1 |
| 1 | 1 | 15 | 104 | 4.17763158 | 0 | 16.6742692 | 60.7361324 | 1 |
| 1 | 1 | 16 | 106 | 4.17763158 | 1 | 3.55263158 | . | 1 |

TreatmentID 0 means the subject is on Control while 1 indicates Treatment. For all the subjects, the survival times and drop out times are generated. **Survival Time** is the duration for which the subject was alive in the study. **DropOutTime** is the duration of time the subject was present in the study before dropping out. For the existing data, the **Drop out times** are generated. These are generated using the specified drop out probabilities for control and treatment. For the new arrivals, accrual times, Survival time as well as drop out time are generated. Open the **SiteSummary** file which stores detailed information about one simulation.

| SiteID | AvgInitiationTim | AvgLastSubjArrTime | AvgNumOfSubj | AvgAccrualDuration | AvgAccrualRate | SiteOpenedSimCount |
|--------|------------------|--------------------|--------------|--------------------|----------------|--------------------|
| 601 | 7.66447368 | 17.7082393 | 2.79 | 18.3450151 | 0.152202001 | 1000 |
| 602 | 4.86842105 | 20.6665723 | 2.682 | 21.1410678 | 0.126935936 | 1000 |
| 603 | 8.42105263 | 22.9385107 | 5.771 | 17.5884362 | 0.328308278 | 1000 |
| 604 | 4.17763158 | 20.9721167 | 3.893 | 21.831746 | 0.178354487 | 1000 |
| 605 | 5.2631579 | 10.7241936 | 1.341 | 20.7463309 | 0.0646568571 | 1000 |
| 606 | 6.57894737 | 23.3447195 | 6.888 | 19.4305414 | 0.354671177 | 1000 |
| 607 | 8.05921053 | 23.396438 | 8.252 | 17.5849062 | 0.472997353 | 1000 |
| 608 | 10.0657895 | 24.5780332 | 3.109 | 15.9419988 | 0.204104198 | 1000 |
| 609 | 19.375 | 23.4331492 | 11.994 | 4.06166241 | 3.0580163 | 1000 |
| 610 | 8.68421053 | 21.5765675 | 4.324 | 17.3252783 | 0.24979278 | 1000 |
| 611 | 7.66447368 | 22.9316773 | 4.206 | 18.3438616 | 0.229378542 | 1000 |
| 401 | 13.7171053 | 23.7030069 | 0.728 | 12.3119235 | 0.113301797 | 1000 |
| 402 | 13.125 | 20.5909601 | 1.725 | 12.8844888 | 0.134044109 | 1000 |
| 403 | 8.91447368 | 20.0872371 | 2.896 | 17.0950151 | 0.169532166 | 1000 |
| 404 | 7.96052632 | 23.7034588 | 0.82 | 18.0292352 | 0.0792590528 | 1000 |
| 301 | 6.57894737 | 23.6131371 | 11.46 | 17.9602484 | 0.644874352 | 1000 |

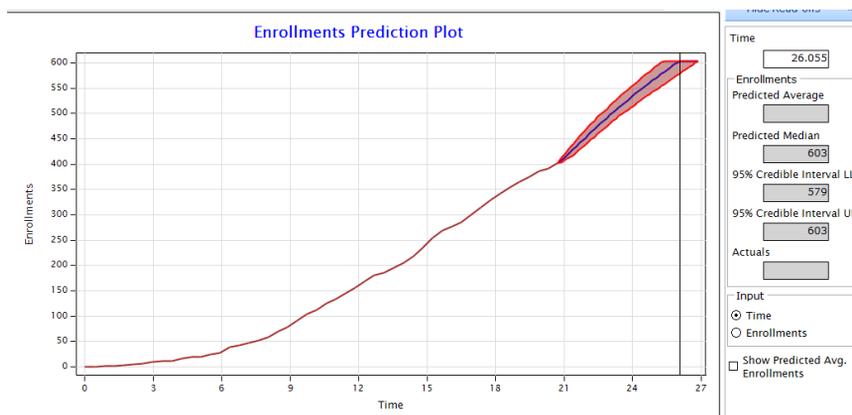
68 Enrollment/Events Prediction - Analysis

The file contains averages across all simulations for each site of the quantities such as initiation times, accrual duration, number of subjects enrolled, accrual rate, number of sites opened etc.

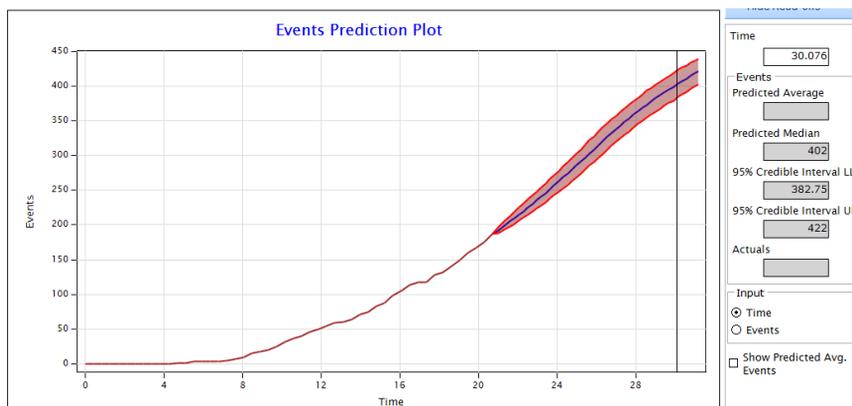
Open the **SiteData** file which stores detailed information about one simulation.

| SimulationID | SiteOpenFlag | SiteAlreadyO | SiteID | SiteInitiation | SiteAccrRate | SubjectsAccr | LastSubjectR | AccrualDurat | ObsrvdAccru |
|--------------|--------------|--------------|--------|----------------|--------------|--------------|--------------|--------------|--------------|
| 1 | 1 | 1 | 601 | 7.66447368 | 0.153148615 | 2 | 10.8552632 | 17.9467174 | 0.111440992 |
| 1 | 1 | 1 | 602 | 4.86842105 | 0.126141079 | 3 | 22.6138459 | 20.7427701 | 0.144628706 |
| 1 | 1 | 1 | 603 | 8.42105263 | 0.32513369 | 4 | 16.5789474 | 17.1901385 | 0.232691552 |
| 1 | 1 | 1 | 604 | 4.17763158 | 0.181312127 | 3 | 16.8092105 | 21.4335595 | 0.139967419 |
| 1 | 1 | 1 | 605 | 5.2631579 | 0.0646808511 | 1 | 5.42763158 | 20.3480332 | 0.0491447989 |
| 1 | 1 | 1 | 606 | 6.57894737 | 0.353488372 | 8 | 25.2097628 | 19.0322437 | 0.420339299 |
| 1 | 1 | 1 | 607 | 8.05921053 | 0.473766234 | 9 | 25.3557568 | 17.5519806 | 0.512762646 |
| 1 | 1 | 1 | 608 | 10.0657895 | 0.562962963 | 3 | 24.8026916 | 15.5454016 | 0.192983113 |
| 1 | 1 | 1 | 609 | 19.375 | 4.44878049 | 12 | 24.7529693 | 5.37796932 | 2.23132549 |
| 1 | 1 | 1 | 610 | 8.68421053 | 0.249180328 | 5 | 21.6852469 | 16.9269806 | 0.295386408 |
| 1 | 1 | 1 | 611 | 7.66447368 | 0.229722922 | 6 | 24.1768435 | 17.9467174 | 0.334322977 |
| 1 | 1 | 1 | 401 | 13.7171053 | 0.142723005 | 0 | - | - | - |
| 1 | 1 | 1 | 402 | 13.125 | 0.131601732 | 1 | 17.1710526 | 12.4861911 | 0.0800884747 |
| 1 | 1 | 1 | 403 | 8.91447368 | 0.169359331 | 2 | 14.5394737 | 16.6967174 | 0.119784024 |
| 1 | 1 | 1 | 404 | 7.96052632 | 0.156701031 | 1 | 22.9683474 | 17.6506648 | 0.0566550899 |
| 1 | 1 | 1 | 301 | 6.57894737 | 0.63627907 | 10 | 22.8659793 | 19.0322437 | 0.525424124 |

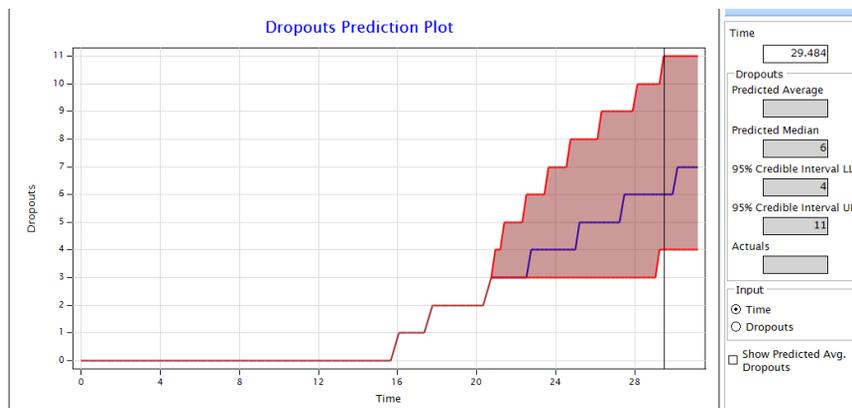
Click on the **PredictSim** node. The **Enrollments Prediction Plot** (invoked using the  tool in the **Library** pane) shows that the median number of enrollments 603 are reached in a duration of 26.055 units of time.



Invoke the **(Events Prediction Plot)** using the  tool in the **Library** pane.



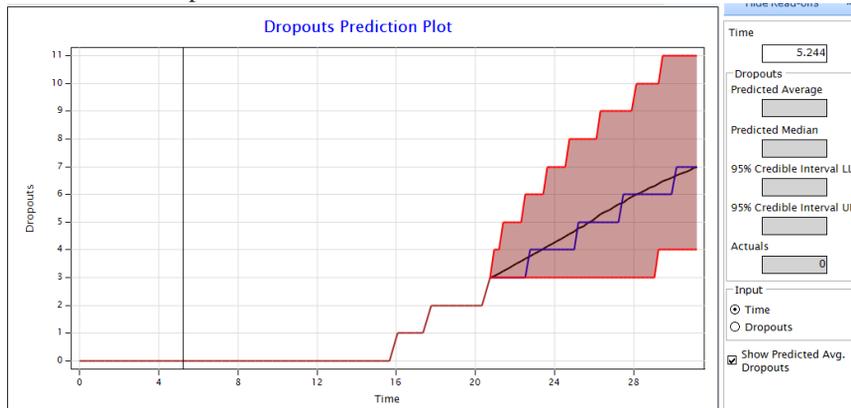
In this plot, the simulation results indicate that by the end of 30.076, the median number of events in both the control and treatment arms would be 402 in a 95% confidence interval of 382.75 to 422. Invoke the **(Dropouts Prediction Plot)** using the  tool in the **Library** pane.



In this plot, the simulation results indicate that by the end of 29.48, the median number of drop outs in both the control and treatment arms would be 6 in a 95% confidence interval of 4 to 11.

68 Enrollment/Events Prediction - Analysis

If you select the **Show Predicted Avg. Dropouts** the predicted dropouts will be added to the plot.



Summary Data In the earlier section we saw how to generate the enrollment and follow them till the required number of events occur or the target sample size is reached. We estimated the study duration with the help of Predict feature in East. For this to use, we assumed that an interim subject-level data was available which had information on individual arrival time, status etc. However, many a times, the subject-level data may not be available. What can be available is the summary of the accruals that have happened till date. For example, in the case of Oncox trial considered above, the DMC statistician may have the information that there have been 402 subjects accrued so far, 203 on Control and 199 on Treatment. The number of events occurred so far on Control and Treatment are 106 and 81 respectively. The subjects dropped out are 1 on Control and 2 on Treatment arm. The last subject arrived at time 20.7236. The DMC statistician is interested in knowing the total study duration when all the accrued 402 subjects are followed till end. East through its Predict feature makes it possible to still come up with an estimate of average study duration based on simulating events from Poisson process based on the specified or default hazard rates. Accruals are simulated till the target sample size is reached or the target number of events are observed. To see this, choose the menu item

68 Enrollment/Events Prediction - Analysis

Click **Next**. The next input dialog appears as shown here:

Analysis: Predict: Events and Enrollment - Unblinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | Simulation Controls

Survival Information

of Hazard Pieces: Input Method:

Hazard Ratio:

Hazard Rate - Control (λ_c):

Hazard Rate - Treatment (λ_t):

The default values for **Hazard Rate - Control** and **Hazard Rate - Treatment** are shown in the dialog. Note that these are not estimated from any data as we don't have the individual subject data as input. Let us use the same hazard rates input as in the previous section. Hazard Rate for Control 0.10847 and that for treatment 0.08062 are these values. The default Target No of events is 402. The 188 events are already occurred. This means that the accrual will continue till we get 402 events in all. After filling all these values, the input dialog looks as shown below:

Analysis: Predict: Events and Enrollment - Unblinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | Simulation Controls

Survival Information

of Hazard Pieces: Input Method:

Hazard Ratio:

Hazard Rate - Control (λ_c):

Hazard Rate - Treatment (λ_t):

Go to the **Accruals/Dropouts** tab. Suppose instead of drop out hazard rates, the information is available on the probabilities of drop out. Suppose the probability of drop out for a subject receiving Control is 0.5% and Treatment is 0.6% and these are applicable from the current calendar time onwards which is 20. 724. Give all these

inputs.

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | Simulation Controls

Current Number Censored: 212 Current Calendar Time: 20.724
 Current No. of Events: 187
 Current No. of Dropouts: 3 Accrual Model: Poisson

Accrual Information

of Accrual Periods: Input Method:

| Period # | Starting at Time | Accrual Rate |
|----------|------------------|--------------|
| 1 | 0.0000 | 19.39817 |
| 2 | 20.7236 | 9.69909 |

Dropout Information

of Pieces: Input Method:

By Time:
 Probability of Dropout -
 Control:
 Probability of Dropout -

Go to **Simulation Controls** tab. Give a fixed seed 12345. Save the outputs for Summary and Subject data.

Target Sample Size: Target No. of Events:

Response Generation | Accrual / Dropouts | **Simulation Controls**

Number of Simulations:
 Refresh Frequency:

Random Number Seed
 Clock
 Fixed

Suppress All Intermediate Output
 Pause after Refresh
 Stop at End

Output Options
 Output Type:
 Save summary statistics for every simulation run
 Save subject level data for simulation runs
 Note: Max. 100,000 records will be saved.

Output for All Trials

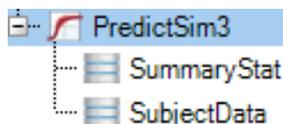
| Percentile (%) |
|----------------|
| 5.000 |
| 25.000 |
| 50.000 |
| 75.000 |
| 95.000 |

Click on **Simulate**. East simulates the events according to Poisson Arrival process with inter-arrival times following exponential distribution. The parameters are derived from the specified hazard rates for Control and Treatment. For details refer to the

68 Enrollment/Events Prediction - Analysis

Appendix M.

After a few seconds East will display the message **Simulations complete. Waiting for User's action.** Click **Close**. This will save the **Predict Simulation** in the Output Preview window first. Once you save it in the Workbook, it will create a node **PredictSim3** with sub-nodes for **SummaryStat** and **SubjectData** in the **Library**.



To view the detailed summary output of the simulations, double click the node **PredictSim3** in the Library. The following output is displayed.

Conditional Simulation: Unblinded Prediction of Enrollment and Event Timelines

| Test Parameters | |
|---|--------------------|
| Simulation ID | PredictSim2 |
| Accrual | Ongoing |
| Input | Summary Data |
| Target Sample Size | 603 |
| Target No. of Events | 402 |
| Response Generation Parameters | |
| HR = λ_c/λ_t | 0.743 |
| Hazard Rate - Control (λ_c) | 0.108 |
| Hazard Rate - Treatment (λ_t) | 0.081 |
| Accrual / Dropouts Parameters | |
| Subject are followed | Until End of Study |
| Dropout Hazard Rate Control | 0.005 |
| Dropout Hazard Rate Treatment | 0.006 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 1000 |

Actuals from Interim Trial Data: Sample Size and Events

| Sample Size | Events | | Dropouts | | Current Time |
|-------------|---------|-----------|----------|-----------|--------------|
| | Control | Treatment | Control | Treatment | |
| 402 | 106 | 81 | 1 | 2 | 20.724 |

Conditional Simulation: Average Sample Size and Events

| Average Sample Size | Average Events | | Average Dropouts | | Average Accrual Duration | Average Study Duration |
|---------------------|----------------|-----------|------------------|-----------|--------------------------|------------------------|
| | Control | Treatment | Control | Treatment | | |
| 543.577 | 214.891 | 187.109 | 6.161 | 9.869 | 35.206 | 35.31 |

Accrual / Dropouts Parameters

Accrual Input Method: Accrual Rates

| Period # | Starting at Time | Accrual Rate |
|----------|------------------|--------------|
| 1 | 0 | 19.398 |
| 2 | 20.724 | 9.699 |

Overall Output

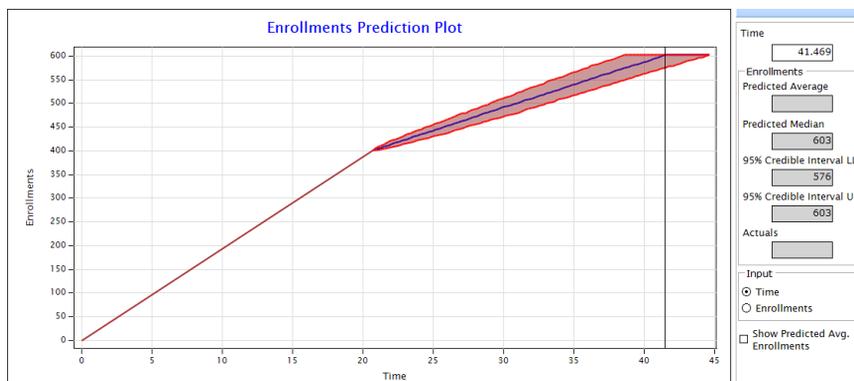
| Stage | Percentile | No. of Actuals | No. of Events | No. of Accruals | No. of Dropouts | Accrual Duration | Study Duration |
|--------------|------------|----------------|---------------|-----------------|-----------------|------------------|----------------|
| Interim | Actuals | 187 | 402 | 3 | 20.724 | 20.724 | |
| | 5% | 402 | 526 | 10 | 33.781 | 33.872 | |
| | 25% | 402 | 536 | 13 | 34.609 | 34.699 | |
| | 50% | 402 | 543 | 16 | 35.171 | 35.295 | |
| End of Trial | 75% | 402 | 551 | 18 | 35.796 | 35.899 | |
| | 95% | 402 | 562 | 22 | 36.685 | 36.815 | |
| | Average | 402 | 543.577 | 16.03 | 35.206 | 35.31 | |

Simulation Seed and Elapsed Time

Starting Seed: 12345
 Total Number of Simulations: 1000
 Elapsed Time: 00:00:13

The table at the left describes the Simulation scenario. This summary contains an overview of the actual data we observed. The table **Actuals from Interim Trial Data: Sample Size and Events** presents the summary data input. The study is complete when in all 402 events occur. The table **Average Sample Size and Events** provide information about the average study duration, average number of events on control and treatment, average drop outs, average follow up time etc. From this table it should be noted that it will take on an average 35.31 units of time to complete the study. The number of events on control arm and treatment arm would be around 215 and 187. The **Overall Output** table describes the details of the distribution of **Average Study Duration** across 1000 simulations. Note that the column **No of Events** has all values 402 (except for actuals) meaning thereby the target sample size of 603 was adequate in giving the required number of events. The values in the column **Number of Accruals** vary since any simulated study concludes as soon as the target number of events are achieved. It is worth noting that the 5th percentile of the Average Study Duration is going to give 402 events pretty early, by the time 33.872. The 95th percentile is 36.815 by which almost in all cases the target would be achieved.

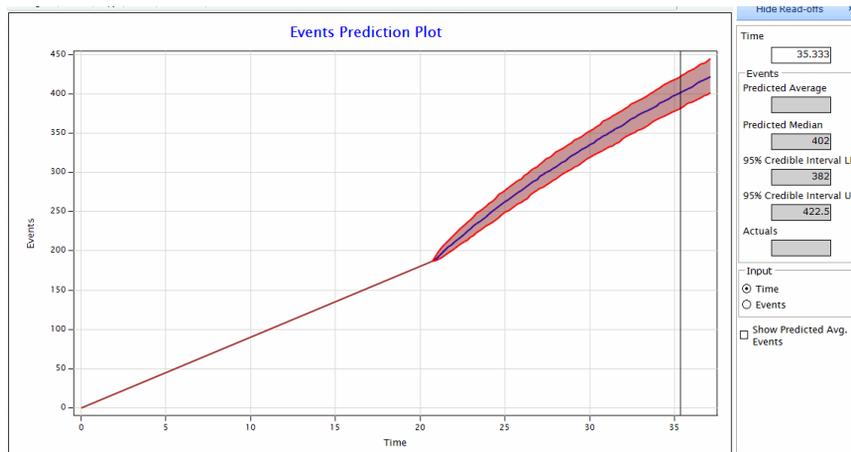
The **Enrollments Prediction Plot** (invoked using the  tool in the **Library** pane) shows that the median number of accruals 603 are reached in a duration of 41.469 units of time.



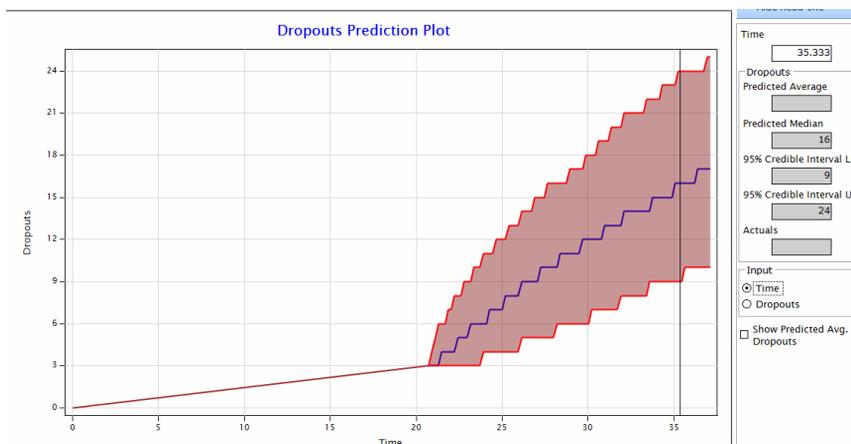
The **Events Prediction Plot** (invoked using the  tool in the **Library** pane) shows that the median number of events 402 are reached in a duration of 35.333 units

68 Enrollment/Events Prediction - Analysis

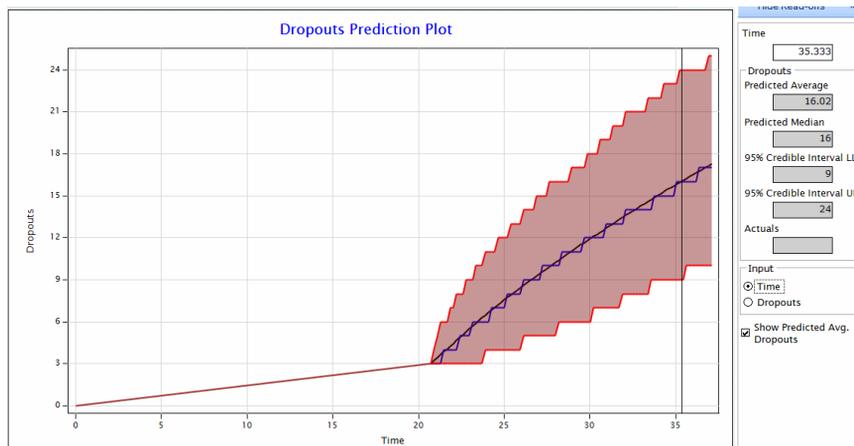
of time.



Invoke the **(Dropout Prediction Plot)** using the  tool in the **Library** pane.



In this plot, the simulation results indicate that by the end of 35.333, the median number of dropouts in both the control and treatment arms would be 16 in a 95% confidence interval of 9 to 24. If you select the **Show Predicted Avg. Dropouts** the predicted dropouts will be added to the plot.



68.3 Events and Enrollment-Blinded Data

68.3.1 Events and Enrollment-Blinded Data: Accrual Complete

In the case of blinded data, information on the individual responses on the treatment and control arms is not available. Instead, common hazard rate and common dropout rate are available. We will explain the Predict feature for blinded data with the help of RALES Trial for Time to Event end point.

Subject-level Data The simulation of the RALES trial in Chapter 66 indicated a required sample size of 1638 subjects with an expected accrual period of around 20 months. The total duration of the study was around 72 months. Assume that an interim look has been taken and subject data are available at this time point. The trial is still accruing subjects and we are interested in forecasting Accrual Duration as well as the Study Duration. The trial has accrued in all 1205 subjects so far. The Subject data are available in the file **RALES_iLook1_SubjectData.csv** in the **Samples** folder of East installation directory. The file **RALES_iLook1_SubjectData.csv** contains a list of subjects accrued so far and the following data for each subject:

- SiteID: the site at which the subject arrived.
- ArrivalTime: the time at which the subject arrived.
- TreatmentID: a variable indicating which group the subject was randomized to ('1' for treatment, '0' for placebo).
- TimeOnStudy: the length of time the subject has been in the study, corresponding to survival time.

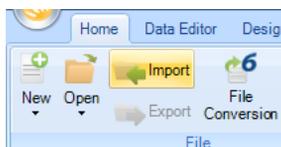
68 Enrollment/Events Prediction - Analysis

- **CensorIndicator**: a variable indicating whether the subject is a completer ('1'), a dropout ('-1') or in the pipeline ('0').
- **CensorInd**: a variable indicating whether the subject is a completer ('1') or a non-completer ('0'). A non-completer can be either a dropout or in the pipeline.

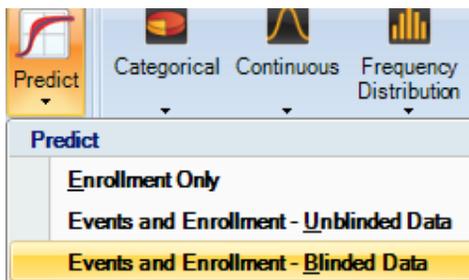
A portion of the **RALES_iLook1_SubjectData.csv** file is shown below:

| | A | B | C | D | E | F | G |
|----|-----------|--------|------------|----------|-----------------|----------|-----------|
| 1 | PatientID | SiteID | ArrivalTim | Treatmen | CensorIndicator | TimeOnSt | CensorInd |
| 2 | 1 | SITE9 | 0.033264 | 1 | 0 | 15.52795 | 0 |
| 3 | 2 | SITE9 | 0.083005 | 1 | 0 | 15.47821 | 0 |
| 4 | 3 | SITE3 | 0.123138 | 0 | 1 | 7.790942 | 1 |
| 5 | 4 | SITE1 | 0.129578 | 0 | 0 | 15.43164 | 0 |
| 6 | 5 | SITE9 | 0.187238 | 0 | 1 | 4.249194 | 1 |
| 7 | 6 | SITE1 | 0.191141 | 1 | 1 | 0.05485 | 1 |
| 8 | 7 | SITE1 | 0.218275 | 1 | 0 | 15.34294 | 0 |
| 9 | 8 | SITE9 | 0.287513 | 1 | 1 | 2.748406 | 1 |
| 10 | 9 | SITE1 | 0.295866 | 1 | 1 | 10.83515 | 1 |

The subject data file can be imported into East using the **Import** button in the Home ribbon:



Once imported, the file will appear in the **Library** pane as nodes in the active workbook, with extension **.cydx**. Choose the menu item **Analysis>Predict>Events and Enrollment-Blinded Data**.



In the ensuing dialog box, select the **Accrual** option, **Complete**. Select data set **RALES_iLook1_SubjectData.cydx**. Map the variables from the data to the

68 Enrollment/Events Prediction - Analysis

Analysis: Predict: Events and Enrollment - Blinded Data (Step 2 of 2)

Target Sample Size:

Target No. of Events:

Response Generation

Accrual / Dropouts

Simulation Controls

| | |
|------------------------------|--|
| Current Sample Size: 1205 | Subjects are followed: <input type="text" value="Until End of Study"/> |
| Current Number Censored: 982 | Current Calendar Time: 15.561 |
| Current No. of Events: 206 | |
| Current No. of Dropouts: 17 | Accrual: Complete (No New Accrual) |

Dropout Information

| | |
|---|---|
| # of Pieces: <input type="text" value="1"/> | Input Method: <input type="text" value="Hazard Rates"/> |
| Dropout Hazard Rate: <input type="text" value="0.00221"/> | |

Subject-level Data: RALES_iLook1_SubjectData.cyx

The **Current Calendar Time** is accrual time of the last subject in the data which is 15.561.

You can edit the values of hazard rate, targeted number of events etc. You can choose a specific follow up period as well by selecting **For Fixed Period** in the **Subjects are followed** textbox. The number of hazard pieces in the Drop out information also can be increased to specify different hazard rates for different time periods. The **Number of pieces** equal to 0 will assume that there aren't going to be any drop outs. For now, let us proceed further with all the default values. Go to **Simulation Controls** tab. Check the Output Options for saving the outputs.

Analysis: Predict: Events and Enrollment - Blinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation Accrual / Dropouts Simulation Controls

Number of Simulations:
 Refresh Frequency:

Random Number Seed
 Clock
 Fixed

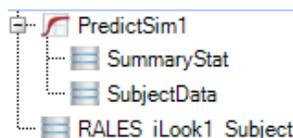
Suppress All Intermediate Output
 Pause after Refresh
 Stop at End

Output Options
 Output Type:
 Save summary statistics for every simulation run
 Save subject level data for simulation runs
 Note: Max. 100,000 records will be saved.

Output for All Trials

| Percentile (%) |
|----------------|
| 5.000 |
| 25.000 |
| 50.000 |
| 75.000 |
| 95.000 |

Click **Simulate**. East simulates the arrival of subjects according to Poisson Arrival process with inter-arrival times following exponential distribution. After a few seconds East will display the message '**Simulations complete. Waiting for User's action**'. Click **Close**. This will save the **Predict Simulation** in the Output Preview window first. Once you save it in the Workbook, it will create a node **PredictSim1** with sub-nodes for **SummaryStat** and **SubjectData** in the **Library**.



To view the detailed summary output of the simulations, double click the node **PredictSim1** in the Library. The following output is displayed.

68 Enrollment/Events Prediction - Analysis

Conditional Simulation: Blinded Prediction of Event Timelines

| Test Parameters | |
|--------------------------------|--------------------|
| Simulation ID | PredictSim1 |
| Accrual | Complete |
| Input | Subject-level Data |
| Target Sample Size | 1205 |
| Target No. of Events | 1205 |
| Response Generation Parameters | |
| Hazard Rate | 0.02683 |
| Accrual / Dropouts Parameters | |
| Subject are followed | Until End of Study |
| Dropout Hazard Rate | 0.00221 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 1000 |

⊖ Actuals from Interim Trial Data: Sample Size and Events

| Sample Size | Events | Dropouts | Pipeline | Average Follow up | Accrual Duration | Current Time |
|-------------|--------|----------|----------|-------------------|------------------|--------------|
| 1205 | 206 | 17 | 982 | 6.372 | 15.549 | 15.561 |

⊖ Conditional Simulation: Average Sample Size and Events

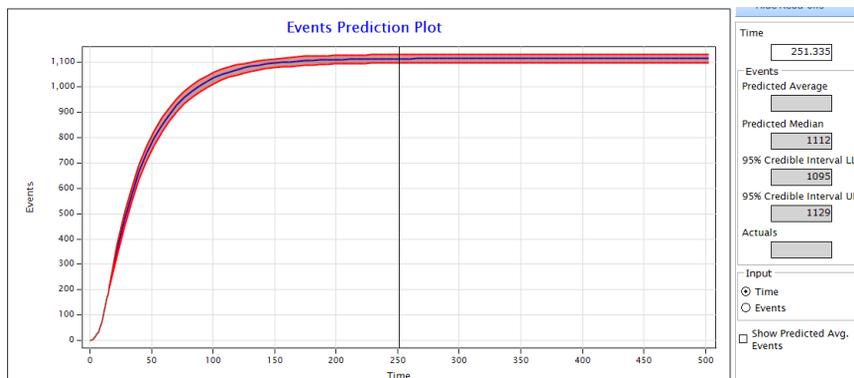
| Average Sample Size | Average Events | Average Dropouts | Average Follow up | Average Study Duration |
|---------------------|----------------|------------------|-------------------|------------------------|
| 1205 | 1113.103 | 91.897 | 34.443 | 273.185 |

⊖ Overall Output

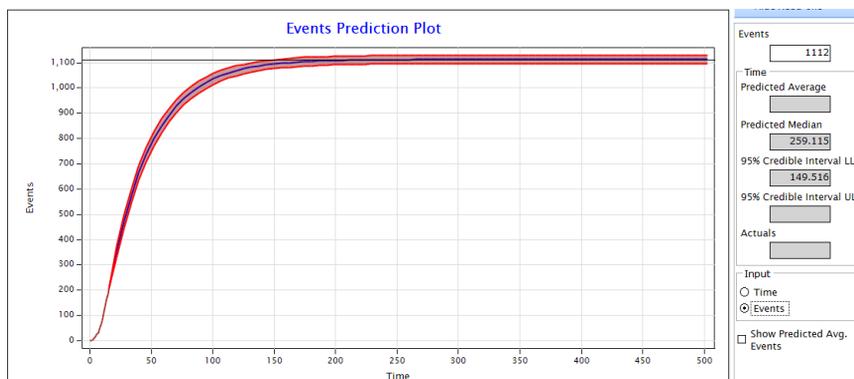
| | Stage | Percentile | No. of Events | No. of Accruals | No. of Dropouts | Study Duration |
|---|--------------|------------|---------------|-----------------|-----------------|----------------|
| | Interim | Actuals | 206 | 1205 | 17 | 15.561 |
| ⊖ | End of Trial | 5% | 1099 | 1205 | 79 | 216.232 |
| | | 25% | 1107 | 1205 | 86 | 242.356 |
| | | 50% | 1113 | 1205 | 92 | 268.125 |
| | | 75% | 1119 | 1205 | 98 | 296.104 |
| | | 95% | 1126 | 1205 | 106 | 351.112 |
| | | Average | 1113.103 | 1205 | 91.897 | 273.185 |

The table at the left describes the Simulation scenario. This summary contains an overview of the actual data we observed. Since the accruals were complete, the **Target Sample Size** and the **Target Number of Events** are the same and equal to 1205. The table **Actuals from Interim Trial Data: Sample Size and Events** presents the detailed information of the subject data such as events, drop outs, average follow up etc. Observe that at the end of the current time, the subjects in pipeline are 982. These are followed till the end of the study. The study is complete when all the subjects in pipeline either experience events or drop out. The table **Average Sample Size and Events** provide information about the average study duration, average number of events, average drop outs, average follow up time etc. From this table it should be noted that it will take on an average 273.185 units of time to complete the study. Average follow up time for an individual is 34.443. The **Overall Output** table describes the details of the distribution of **Average Study Duration** across 1000 simulations. Note that the column **No of Accruals** has all values 1205 since the accruals were complete and only events are being forecasted. Since there are a few drop outs, we expect on an average a lesser, say around 1113 events to occur out of 1205 subjects. It is worth noting that the 5th percentile of the Average Study Duration is going to give 1099 events by the time 216.232. The 95th percentile is 351.112 which is the maximum duration the study can take. You could have changed the percentiles input, if you want to be more specific.

For instance you can input 100% to get the value of maximum study duration. Let us have a look at the individual files stored in the **Library**. The **Events Prediction Plot** (invoked using the  tool in the **Library** pane) shows that the median number of events 1112 are reached in a duration of 251.335 units of time.



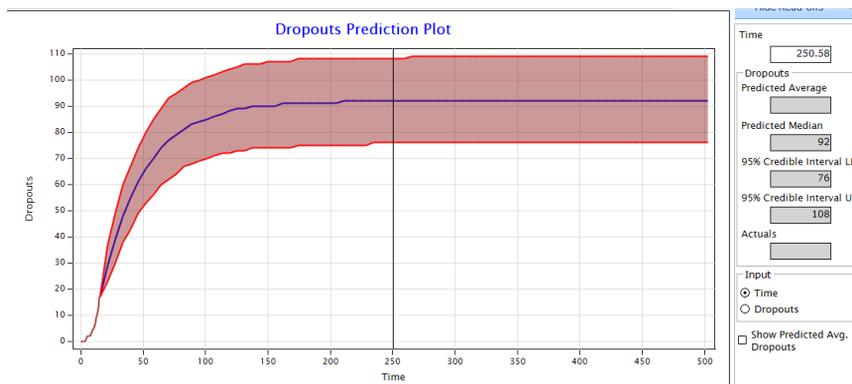
In order to find the average study duration for getting the median number of events, select the Input option **Events** in the plot. Enter 1112 for Events.



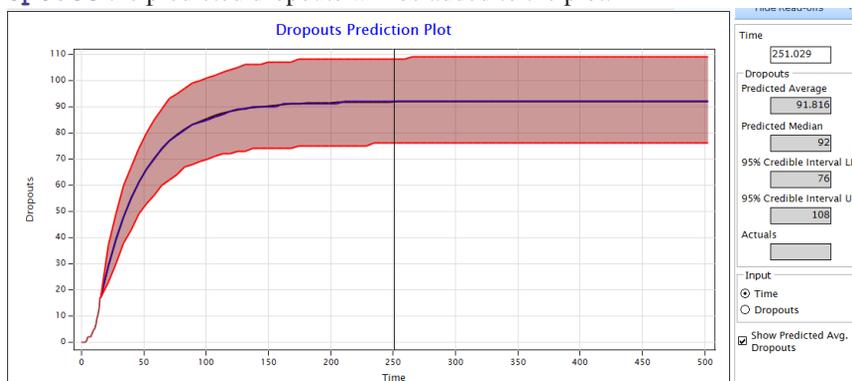
The median study duration is 259.115. Since the 95% upper limit does not exist, one can not forecast the latest time to reach the target.

68 Enrollment/Events Prediction - Analysis

Invoke the (**Dropout Prediction Plot**) using the  tool in the **Library** pane.



In this plot, the simulation results indicate that by the end of 250.58, the median number of dropouts in both the control and treatment arms would be 92 in a 95% confidence interval of 76 to 108. If you select the **Show Predicted Avg. Dropouts** the predicted dropouts will be added to the plot.



Summary Data In the earlier section we saw how to generate the events and follow all subjects till they experience either events or drop out. We estimated the study duration with the help of Predict feature in East. For this to use, we assumed that an interim subject-level data was available which had information on individual arrival time, status etc. However, many a times, the subject-level data may not be available. Instead, the summary of the accruals that have happened till date can be available. For

68 Enrollment/Events Prediction - Analysis

Click **Next**. The next input dialog appears as shown here:

Analysis: Predict: Events and Enrollment - Blinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | Simulation Controls

Survival Information

of Hazard Pieces: Input Method:

Hazard Rate (λ):

The default value for **Hazard Rate** is shown in the dialog. Note that this is not estimated from any data as we don't have the individual subject data as input. Let us use the same hazard rate input as in the previous section, namely 0.02683.

Analysis: Predict: Events and Enrollment - Blinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | Simulation Controls

Survival Information

of Hazard Pieces: Input Method:

Hazard Rate (λ):

Go to the **Accruals/Dropouts** tab. Suppose instead of drop out hazard rates, the information is available on the probabilities of drop out. Suppose the probability of drop out for a subject receiving any of the Control or Treatment is 0.5% which is

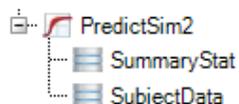
applicable from the current calendar time onwards 15.55. Give all these inputs.

Go to **Simulation Controls** tab. Give a fixed seed 12345. Save the outputs for Summary and Subject data.

Click on **Simulate**. East simulates the events according to Poisson Arrival process with inter-arrival times following exponential distribution. The parameters are derived from the specified hazard rates for Control and Treatment. For details refer to the Appendix M. After a few seconds East will display the message **Simulations complete. Waiting for User's action**. Click **Close**. This will save the **Predict Simulation** in the Output Preview window first. Once you save it in the Workbook, it will create a node **PredictSim2** with sub-nodes for

68 Enrollment/Events Prediction - Analysis

SummaryStat and SubjectData in the Library.



To view the detailed summary output of the simulations, double click the node **PredictSim2** in the Library. The following output is displayed.

Conditional Simulation: Blinded Prediction of Event Timelines

| Test Parameters | |
|--------------------------------|--------------------|
| Simulation ID | PredictSim2 |
| Accrual | Complete |
| Input | Summary Data |
| Target Sample Size | 1205 |
| Target No. of Events | 1205 |
| Response Generation Parameters | |
| Hazard Rate | 0.02683 |
| Accrual / Dropouts Parameters | |
| Subject are followed | Until End of Study |
| Dropout Hazard Rate | 0.005 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 1000 |

Actuals from Interim Trial Data: Sample Size and Events

| Sample Size | Events | Dropouts | Pipeline | Current Time |
|-------------|--------|----------|----------|--------------|
| 1205 | 206 | 17 | 982 | 15.55 |

Conditional Simulation: Average Sample Size and Events

| Average Sample Size | Average Events | Average Dropouts | Average Study Duration |
|---------------------|----------------|------------------|------------------------|
| 1205 | 1033.55 | 171.45 | 249.489 |

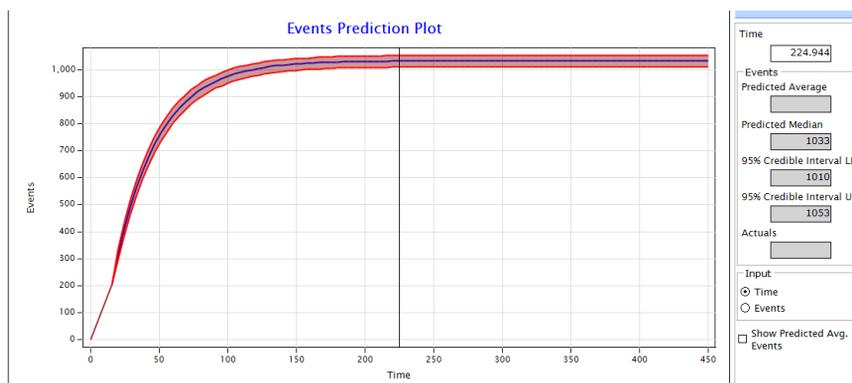
Overall Output

| Stage | Percentile | No. of Events | No. of Accruals | No. of Dropouts | Study Duration |
|--------------|------------|---------------|-----------------|-----------------|----------------|
| Interim | Actuals | 206 | 1205 | 17 | 15.55 |
| | 5% | 1014 | 1205 | 154 | 199.048 |
| End of Trial | 25% | 1026 | 1205 | 164 | 222.563 |
| | 50% | 1034 | 1205 | 171 | 244.173 |
| | 75% | 1041 | 1205 | 179 | 270.544 |
| | 95% | 1051 | 1205 | 191 | 315.872 |
| | Average | 1033.55 | 1205 | 171.45 | 249.489 |

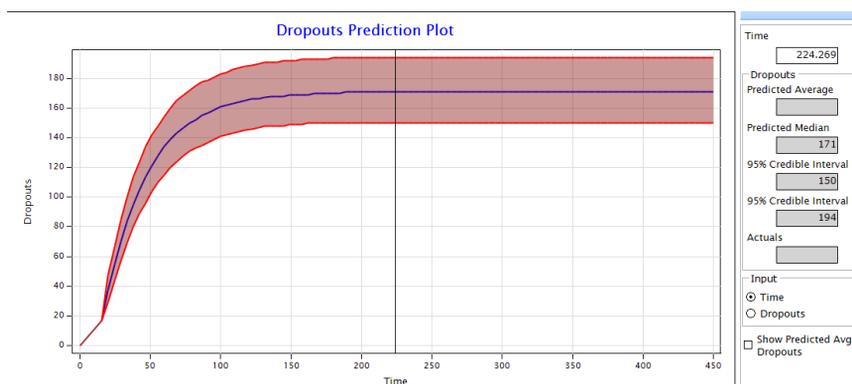
The table at the left describes the Simulation scenario. This summary contains an overview of the actual data we observed. Since the accruals were complete, the **Target Sample Size** and the **Target Number of Events** are 1205. The table **Actuals from Interim Trial Data: Sample Size and Events** presents the detailed information of the subject data such as number of events, drop outs, average follow up etc. Observe that at the end of the current time, the subjects in pipeline are 982. These are followed till the end of the study. The study is complete when in all 1205 events occur. The table **Average Sample Size and Events** provide information about the average study duration, average number of events, average drop outs, average follow up time etc. From this table it should be noted that it will take on an average 249.489 units of time to complete the study. The **Overall Output** table describes the details of the distribution of **Average Study Duration** across 1000 simulations. Note that the column **No of Accruals** has all values 1205 since the accruals were complete and only events are

being forecasted. Although the targeted number of events was 1205, the **No of Events** column shows different values less than 1205 as the accruals were complete and only these subjects were followed till they produce events or drop out. It is worth noting that the 5th percentile of the Average Study Duration is going to give 1014 events, by the time 199.048 and the 95th percentile is 315.872 by which again only 1051 events have been occurred. The investigator has to decide whether to wait for a longer time for getting a few more events.

The **Events Prediction Plot** (invoked using the  tool in the **Library** pane) shows that the median number of events 1033 are reached in a duration of 224.944 units of time.

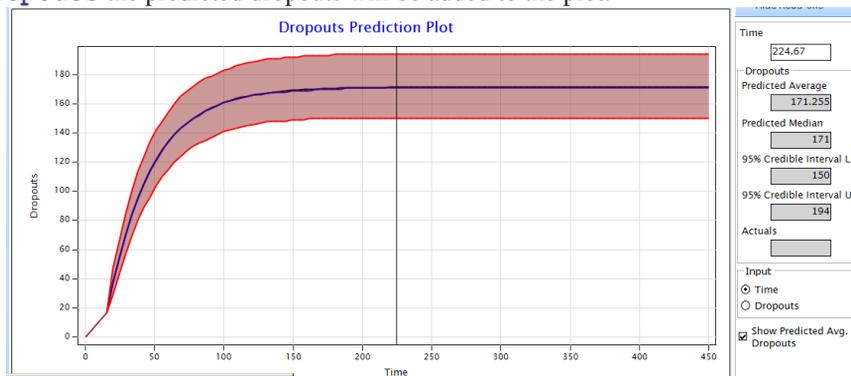


Invoke the **(Dropout Prediction Plot)** using the  tool in the **Library** pane.



68 Enrollment/Events Prediction - Analysis

In this plot, the simulation results indicate that by the end of 224.269, the median number of dropouts in both the control and treatment arms would be 171 in a 95% confidence interval of 150 to 194. If you select the **Show Predicted Avg. Dropouts** the predicted dropouts will be added to the plot.



68.3.2 Events and Enrollment-Blinded Data: Accrual Ongoing

Subject-level Data The simulation of the RALES trial in Chapter 66 indicated a required sample size of 1638 subjects with an expected accrual period of around 20 months. The total duration of the study was around 72 months. Assume that an interim look has been taken and subject data are available at this time point. The trial is still accruing subjects and we are interested in forecasting Accrual Duration as well as Study Duration. The trial has accrued in all 1205 subjects so far. The Subject data are available in the file **RALES_iLook1_SubjectData.csv** in the Samples folder of East installation directory. The file RALES_iLook1_SubjectData.csv contains a list of subjects accrued so far and the following data for each subject:

- SiteID: the site at which the subject arrived.
- ArrivalTime: the time at which the subject arrived.
- TreatmentID: a variable indicating which group the subject was randomized to ('1' for treatment, '0' for placebo).
- TimeOnStudy: the length of time the subject has been in the study, corresponding to survival time.
- CensorIndicator: a variable indicating whether the subject is a completer ('1'), a dropout ('-1') or in the pipeline ('0').
- CensorInd: a variable indicating whether the subject is a completer ('1') or a non-completer ('0'). A non-completer can be either a dropout or in the pipeline.

A portion of the RALES_iLook1_SubjectData file is shown below:

| | A | B | C | D | E | F | G |
|----|-----------|--------|------------|----------|-----------------|----------|-----------|
| 1 | PatientID | SiteID | ArrivalTim | Treatmen | CensorIndicator | TimeOnSt | CensorInd |
| 2 | 1 | SITE9 | 0.033264 | 1 | 0 | 15.52795 | 0 |
| 3 | 2 | SITE9 | 0.083005 | 1 | 0 | 15.47821 | 0 |
| 4 | 3 | SITE3 | 0.123138 | 0 | 1 | 7.790942 | 1 |
| 5 | 4 | SITE1 | 0.129578 | 0 | 0 | 15.43164 | 0 |
| 6 | 5 | SITE9 | 0.187238 | 0 | 1 | 4.249194 | 1 |
| 7 | 6 | SITE1 | 0.191141 | 1 | 1 | 0.05485 | 1 |
| 8 | 7 | SITE1 | 0.218275 | 1 | 0 | 15.34294 | 0 |
| 9 | 8 | SITE9 | 0.287513 | 1 | 1 | 2.748406 | 1 |
| 10 | 9 | SITE1 | 0.295866 | 1 | 1 | 10.83515 | 1 |

The trial is assumed to enroll subjects from several sites, the information about which is provided in the file **RALES_iLook1_SiteData .csv** also available in **Samples** folder. The file contains the following data for each site:

- SiteID: the identification number of the site.
- SiteReadyTime: the time at which the site was initiated.
- SiteAccrRate: the site accrual rate specified in the enrollment plan.
- SubjectsAccrued: the number of subjects accrued at the site.
- LastSubjectRand: the randomization time of the last subject arriving at the site.
- ObsrvdAccrualRate: the observed accrual rate at the site.
- PosteriorAccrualRate: the updated site accrual rate.
- SIP_Start: the start of the initiation period of the site.
- SIP_End: the end of the initiation period of the site.
- Ecap: the enrollment cap, representing the maximum number of subjects that can be enrolled at the site.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|-----|--------------|--------|---------------|--------------|-----------------|-----------------|-----------------|----------------|-------------------|----------|--------|-------|------|
| 1 | Sno | SiteOpenFlag | SiteID | SiteReadyTime | SiteAccrRate | SubjectsAccrued | LastSubjectRand | AccrualDuration | ObsrvdAccrRate | PosteriorAccrRate | SIPstart | SIPend | Arate | Ecap |
| 2 | 1 | 1 | SITE1 | 0 | 8 | 110 | 15.49742096 | 15.5612138 | 7.068857313 | 7.076277385 | 0 | 0 | 8 | 1000 |
| 3 | 2 | 1 | SITE2 | 0.667776346 | 8 | 125 | 15.54130249 | 14.65942746 | 4.507199242 | 4.50290795 | 0 | 1 | 8 | 1000 |
| 4 | 3 | 1 | SITE3 | 0.096637821 | 4 | 47 | 15.30974738 | 15.46437598 | 3.039243229 | 3.054527908 | 0 | 1 | 4 | 1000 |
| 5 | 4 | 1 | SITE4 | 0.270185643 | 4 | 82 | 15.28105966 | 15.29102816 | 5.362621737 | 5.340701989 | 0 | 1 | 4 | 1000 |
| 6 | 5 | 1 | SITE5 | 0.816564942 | 4 | 64 | 15.52819167 | 14.74464886 | 4.340557758 | 4.33487977 | 0 | 1 | 4 | 1000 |
| 7 | 6 | 1 | SITE6 | 0.335109399 | 4 | 59 | 15.47260272 | 15.2261044 | 3.874924172 | 3.876944639 | 0 | 1 | 4 | 1000 |
| 8 | 7 | 1 | SITE7 | 0.988711395 | 4 | 61 | 15.38173883 | 14.57250241 | 4.183966026 | 4.182829477 | 0 | 1 | 4 | 1000 |
| 9 | 8 | 1 | SITE8 | 0.643672688 | 4 | 62 | 15.40828988 | 14.91754112 | 4.155180936 | 4.153606674 | 0 | 1 | 4 | 1000 |
| 10 | 9 | 1 | SITE9 | 0.029872714 | 4 | 67 | 15.54004467 | 15.53134109 | 4.313857999 | 4.30888602 | 0 | 1 | 4 | 1000 |
| 11 | 10 | 1 | SITE10 | 0.574258361 | 4 | 57 | 15.4884442 | 14.9895544 | 3.803307498 | 3.806534725 | 0 | 1 | 4 | 1000 |
| 12 | 11 | 1 | SITE11 | 0.865589555 | 4 | 67 | 15.54905438 | 14.69562425 | 4.559180261 | 4.549826683 | 0 | 1 | 4 | 1000 |
| 13 | 12 | 1 | SITE12 | 0.467937141 | 4 | 65 | 15.39676354 | 15.09327666 | 4.306553272 | 4.30155836 | 0 | 1 | 4 | 1000 |
| 14 | 13 | 1 | SITE13 | 0.533081979 | 4 | 60 | 15.3618035 | 15.02813182 | 3.992512223 | 3.992634748 | 0 | 1 | 4 | 1000 |
| 15 | 14 | 1 | SITE14 | 0.46896191 | 4 | 65 | 15.49317175 | 15.09735189 | 4.305396804 | 4.300416154 | 0 | 1 | 4 | 1000 |
| 16 | 15 | 1 | SITE15 | 0.230835496 | 4 | 55 | 14.70403722 | 15.33035831 | 3.54765261 | 3.59426972 | 0 | 1 | 4 | 1000 |
| 17 | 16 | 1 | SITE16 | 0.460707765 | 4 | 48 | 15.28051965 | 15.10050604 | 3.17870421 | 3.152077178 | 0 | 1 | 4 | 1000 |
| 18 | 17 | 1 | SITE17 | 0.572273339 | 2 | 28 | 14.33945704 | 14.9894046 | 1.86804398 | 1.872303665 | 0 | 1 | 2 | 1000 |
| 19 | 18 | 1 | SITE18 | 0.891082007 | 2 | 31 | 15.3240577 | 14.6701318 | 2.113137116 | 2.109408173 | 0 | 1 | 2 | 1000 |
| 20 | 19 | 1 | SITE19 | 0.939732219 | 2 | 26 | 15.53443802 | 14.62148158 | 1.77820557 | 1.785539324 | 0 | 1 | 2 | 1000 |
| 21 | 20 | 1 | SITE20 | 0.771935426 | 2 | 26 | 15.54880295 | 14.78927838 | 1.758030334 | 1.765943384 | 0 | 1 | 2 | 1000 |

Site-level Data

Select Workbook:

Select Dataset:

Choose Variables from Dataset

Site ID: **Site Initiation: Unopened Sites**

Accrual Rate/Site: Start Time:

Enrollment Cap: End Time:

Site Initiation: Opened Sites

Site Initiation Time:

Click **Next**. The next input dialog appears as shown here:

Analysis: Predict: Events and Enrollment - Blinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation **Accrual / Dropouts** Simulation Controls

Survival Information

of Hazard Pieces: Input Method:

Hazard Rate (λ):

Subject-level Data: RALES_iLook1_SubjectData.cydx
 Site-level Data: RALES_iLook1_SiteData.cydx

The default value for **Hazard Rate** is estimated from the subject data. Note that this is estimation of common hazard rate ignoring whether the event is occurring on the treatment arm or control arm. Go to the **Accrual/ Dropouts** tab. In the ensuing dialog, you will see almost all values filled in. Change the **Accrual Model** to **Poisson**.

68 Enrollment/Events Prediction - Analysis

Analysis: Predict: Events and Enrollment - Blinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation
 Accrual / Dropouts
 Simulation Controls

Current Sample Size: 1205 Subjects are followed:

Current Number Censored: 982 Current Calendar Time: 15.561

Current No. of Events: 206

Current No. of Dropouts: 17 Accrual Model:

Total Number of Sites: Number of Sites Opened: 20

| Site ID | Site Initiation | Period | Accrual | Enrollment | Planned Accrual | Site Initiation | No. of |
|---------|-----------------|--------|-----------|------------|-----------------|-----------------|----------|
| | Start | End | Rate/Site | Cap | Rate/Site | Time | Accruals |
| SITE1 | NA | NA | 7.074 | 1000 | 15.497 | 0 | 110 |
| SITE2 | NA | NA | 8.514 | 1000 | 15.541 | 0.868 | 125 |
| SITE3 | NA | NA | 3.042 | 1000 | 15.31 | 0.097 | 47 |
| SITE4 | NA | NA | 5.367 | 1000 | 15.281 | 0.27 | 82 |
| SITE5 | NA | NA | 4.344 | 1000 | 15.528 | 0.817 | 64 |
| SITE6 | NA | NA | 3.878 | 1000 | 15.473 | 0.335 | 59 |
| SITE7 | NA | NA | 4.189 | 1000 | 15.382 | 0.989 | 61 |
| SITE8 | NA | NA | 4.16 | 1000 | 15.408 | 0.644 | 62 |
| SITE9 | NA | NA | 4.317 | 1000 | 15.54 | 0.03 | 67 |

Dropout Information

of Pieces: Input Method:

Dropout Hazard Rate:

Subject-level Data: RALES_iLook1_SubjectData.cydx

Site-level Data: RALES_iLook1_SiteData.cydx

The **Current Calendar Time** is accrual time of the last subject in the data which is 15.561. Again, the drop out hazard rate is estimated from the subject data assuming that the data are blinded on treatments. The **Target Sample Size** is 1807 which is $1.5 * SampleSize$. You are free to change the **Target Sample Size** as we are assuming that the study is still accepting enrollment. The **Target Number of Events** default value is 1205, the same as in the data. This value can be edited. You can edit the values of hazard rate, target number of events etc. You can choose a specific follow up period as well by selecting **For Fixed Period** in the **Subjects are followed** textbox. The number of hazard pieces in the Drop out information also can be increased to specify different hazard rates for different time periods. The **Number of pieces** equal to 0 will assume that there aren't going to be any drop outs. For now, let us proceed further with all the default values. Go to **Simulation Controls** tab. Give a fixed seed 12345. Check the Output Options

for saving the outputs.

Analysis: Predict: Events and Enrollment - Blinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | **Simulation Controls**

Number of Simulations:
 Refresh Frequency:

Random Number Seed
 Clock
 Fixed

Suppress All Intermediate Output
 Pause after Refresh
 Stop at End

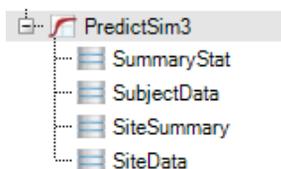
Output Options
 Output Type:
 Save summary statistics for every simulation run
 Save subject level data for simulation runs
 Save site-wise summary for every simulation run
 Save site parameters data for simulation runs
 Note: Max. 100,000 records will be saved.

Output for All Trials

| Percentile (%) |
|----------------|
| 5.000 |
| 25.000 |
| 50.000 |
| 75.000 |

Subject-level Data: RALES_iLook1_SubjectData.cydx
 Site-level Data: RALES_iLook1_SiteData.cydx

Click **Simulate**. East simulates the arrival of subjects according to Poisson Arrival process with inter-arrival times following exponential distribution. After a few seconds East will display the message **Simulations complete. Waiting for User's action**. Click **Close**. This will save the **Predict Simulation** in the Output Preview window first. Once you save it in the Workbook, it will create a node **PredictSim1** with sub-nodes for **SummaryStat**, **SubjectData**, **SiteSummary** and **SitePara** in the **Library**.



To view the detailed summary output of the simulations, double click the node

68 Enrollment/Events Prediction - Analysis

PredictSim1 in the Library. The following output is displayed.

Conditional Simulation: Blinded Prediction of Enrollment and Event Timelines

| Test Parameters | |
|--------------------------------|--------------------|
| Simulation ID | PredictSim3 |
| Accrual | Ongoing |
| Input | Subject-level Data |
| Include Site Info | Yes |
| Target Sample Size | 1807 |
| Target No. of Events | 1205 |
| Response Generation Parameters | |
| Hazard Rate | 0.02683 |
| Accrual / Dropouts Parameters | |
| Subject are followed | Until End of Study |
| Accrual Model | Poisson |
| Total No. of Sites | 20 |
| Dropout Hazard Rate | 0.00221 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 1000 |

Actuals from Interim Trial Data: Sample Size and Events

| Sample Size | Events | Dropouts | Average Follow up | Accrual Duration | Current Time |
|-------------|--------|----------|-------------------|------------------|--------------|
| 1205 | 206 | 17 | 6.372 | 15.549 | 15.561 |

Conditional Simulation: Average Sample Size and Events

| Average Sample Size | Average Events | Average Dropouts | Average Follow up | Average Accrual Duration | Average Study Duration |
|---------------------|----------------|------------------|-------------------|--------------------------|------------------------|
| 1807 | 1205 | 99.064 | 24.855 | 23.058 | 56.573 |

Overall Output

| Stage | Percentile | No. of Events | No. of Sites Opened | No. of Accruals | No. of Dropouts | Accrual Duration | Study Duration |
|--------------|------------|---------------|---------------------|-----------------|-----------------|------------------|----------------|
| Interim | Actuals | 206 | 0 | 1205 | 17 | 15.549 | 15.561 |
| | 5% | 1205 | 20 | 1807 | 84 | 22.542 | 54.29 |
| End of Trial | 25% | 1205 | 20 | 1807 | 93 | 22.86 | 55.647 |
| | 50% | 1205 | 20 | 1807 | 99 | 23.059 | 56.503 |
| | 75% | 1205 | 20 | 1807 | 105 | 23.245 | 57.453 |
| | 95% | 1205 | 20 | 1807 | 115 | 23.556 | 58.866 |
| | Average | 1205 | 20 | 1807 | 99.064 | 23.058 | 56.573 |

The table at the left describes the Simulation scenario. This summary contains an overview of the actual data we observed. The table **Actuals from Interim Trial Data: Sample Size and Events** presents the detailed information of the subject data such as events, drop outs, average follow up etc. The table **Average Sample Size and Events** provide information about the average study duration, average number of events considering both the arms together, average drop outs, average follow up time etc. From this table it should be noted that it will take on an average 56.573 units of time to complete the study getting all the 1205 events. Average follow up time for an individual is 24.855. The **Average Accrual Duration** is 23.058. The **Overall Output** table describes the details of the distribution of **Accrual Duration** and **Study Duration** across 1000 simulations. It is worth noting that the 5th percentile of the Average Study Duration is going to give 1205 events pretty early, by the time 54.29. The 95th percentile is 58.866 which is the maximum duration the study can take. You could have changed the percentiles input, if you want to be more specific. For instance, you can input 100% to get the value of maximum study duration. Open the **SummaryStat** data by double clicking the sub node. You will see the following display of data (shown in parts).

| ScenarioID | SimIndex | Stage | Status | TotAccruals | TotDropOuts | TotPendings |
|------------|----------|---------|---------|-------------|-------------|-------------|
| 1 | 1 | Interim | SUCCESS | 1205 | 17 | 982 |
| 1 | 1 | Final | SUCCESS | 1807 | 107 | 495 |
| 1 | 2 | Interim | SUCCESS | 1205 | 17 | 982 |
| 1 | 2 | Final | SUCCESS | 1807 | 94 | 508 |
| 1 | 3 | Interim | SUCCESS | 1205 | 17 | 982 |
| 1 | 3 | Final | SUCCESS | 1807 | 96 | 506 |
| 1 | 4 | Interim | SUCCESS | 1205 | 17 | 982 |
| 1 | 4 | Final | SUCCESS | 1807 | 108 | 494 |
| 1 | 5 | Interim | SUCCESS | 1205 | 17 | 982 |
| 1 | 5 | Final | SUCCESS | 1807 | 94 | 508 |
| 1 | 6 | Interim | SUCCESS | 1205 | 17 | 982 |
| 1 | 6 | Final | SUCCESS | 1807 | 102 | 500 |
| 1 | 7 | Interim | SUCCESS | 1205 | 17 | 982 |
| 1 | 7 | Final | SUCCESS | 1807 | 112 | 490 |
| 1 | 8 | Interim | SUCCESS | 1205 | 17 | 982 |
| 1 | 8 | Final | SUCCESS | 1807 | 92 | 510 |

| TotEvents | LookTime | AvgFollowUp | AccrDurtn |
|-----------|------------|-------------|------------|
| 206 | 15.5612138 | 6.37188162 | . |
| 1205 | 58.7788267 | 25.9763865 | 23.6794267 |
| 206 | 15.5612138 | 6.37188162 | . |
| 1205 | 57.3955021 | 25.4300969 | 22.8068019 |
| 206 | 15.5612138 | 6.37188162 | . |
| 1205 | 55.9089273 | 24.8431129 | 22.8376137 |
| 206 | 15.5612138 | 6.37188162 | . |
| 1205 | 57.2235217 | 24.858187 | 22.9155648 |
| 206 | 15.5612138 | 6.37188162 | . |
| 1205 | 56.0206866 | 24.8088062 | 23.2308946 |
| 206 | 15.5612138 | 6.37188162 | . |
| 1205 | 56.336879 | 25.0240714 | 22.8906787 |
| 206 | 15.5612138 | 6.37188162 | . |
| 1205 | 56.6849826 | 24.8888107 | 23.004226 |
| 206 | 15.5612138 | 6.37188162 | . |
| 1205 | 58.2624171 | 25.5864867 | 23.0333272 |

68 Enrollment/Events Prediction - Analysis

Note that the file contains overall information and not on individual treatment since the data are blinded. Observe that the **TotAccruals** for every simulation is 1807, and the **TotEvents** is 1205. The

tt TotPending values for **Final Look** are the subjects which have neither experienced events nor have dropped out till the end of the study. This is because the study is concluded after getting 1205 events and does not proceed till all the subjects experience the event as was the case in the previous section. It is worth noting that the **LookTime** corresponding to the Final stage is essentially the study duration observed in that particular simulation.

Open the **Subject Data** file which stores detailed information about one simulation.

| ScenarioID | SimulationID | SubjectID | SiteID | ArrivalTime | SurvivalTime | DropOutTime | CensorInd_1 |
|------------|--------------|-----------|--------|-------------|--------------|-------------|-------------|
| 1 | 4 | 1 | SITE9 | 0.033264221 | 15.6442149 | 72.4325353 | 1 |
| 1 | 4 | 2 | SITE9 | 0.083004659 | 89.2386969 | 1401.13179 | 0 |
| 1 | 4 | 3 | SITE3 | 0.123138121 | 7.79094196 | . | 1 |
| 1 | 4 | 4 | SITE1 | 0.129577928 | 39.9958859 | 153.508787 | 1 |
| 1 | 4 | 5 | SITE9 | 0.18723815 | 4.24919401 | . | 1 |
| 1 | 4 | 6 | SITE1 | 0.191141046 | 0.054849627 | . | 1 |
| 1 | 4 | 7 | SITE1 | 0.218274781 | 101.950753 | 502.870168 | 0 |
| 1 | 4 | 8 | SITE9 | 0.287513081 | 2.74840645 | . | 1 |
| 1 | 4 | 9 | SITE1 | 0.295866118 | 10.8351481 | . | 1 |
| 1 | 4 | 10 | SITE15 | 0.333387424 | 0.929750744 | . | 1 |
| 1 | 4 | 11 | SITE15 | 0.380771385 | 26.9309624 | 703.997202 | 1 |
| 1 | 4 | 12 | SITE1 | 0.388893193 | 4.005856 | . | 1 |
| 1 | 4 | 13 | SITE9 | 0.460566744 | 3.01041245 | . | 1 |
| 1 | 4 | 14 | SITE16 | 0.496907792 | 16.6690848 | 524.9554 | 1 |
| 1 | 4 | 15 | SITE14 | 0.516395459 | 116.787629 | 93.8083191 | 0 |
| 1 | 4 | 16 | SITE15 | 0.532471377 | 9.19540687 | . | 1 |

For all the subjects, the survival times and drop out times are generated. **Survival Time** is the duration for which the subject was alive in the study. **DropoutTime** is the duration of time the subject was present in the study before dropping out. For the existing interim data, the **Drop out times** are generated. These are generated using the specified drop out probability. For the new arrivals, accrual times, survival times as well as drop out times are generated. Open the **SiteSummary** file which stores detailed information about one simulation.

| SiteID | AvgInitiation | AvgLastSubj | AvgNumOfSu | AvgAccrualD | AvgAccrualR | SiteOpenedS |
|--------|---------------|-------------|------------|-------------|-------------|-------------|
| SITE1 | 0 | 22.9222579 | 163 | 23.0575281 | 7.07025361 | 1000 |
| SITE2 | 0.867776346 | 22.9518059 | 189.168 | 22.1897517 | 8.52683923 | 1000 |
| SITE3 | 0.096837821 | 22.752676 | 69.795 | 22.9606902 | 3.04035644 | 1000 |
| SITE4 | 0.270185643 | 22.8877262 | 122.632 | 22.7873424 | 5.38261544 | 1000 |
| SITE5 | 0.816564942 | 22.8373052 | 96.628 | 22.2409631 | 4.34542623 | 1000 |
| SITE6 | 0.335109399 | 22.8086504 | 88.173 | 22.7224187 | 3.88111013 | 1000 |
| SITE7 | 0.988711395 | 22.8280228 | 92.36 | 22.0688167 | 4.18576332 | 1000 |
| SITE8 | 0.643672688 | 22.8269208 | 93.259 | 22.4138554 | 4.16134149 | 1000 |
| SITE9 | 0.029872714 | 22.8376078 | 99.001 | 23.0276553 | 4.29992848 | 1000 |
| SITE10 | 0.574258361 | 22.8159865 | 85.634 | 22.4832697 | 3.80949086 | 1000 |
| SITE11 | 0.865589555 | 22.8556951 | 101.235 | 22.1919385 | 4.56259999 | 1000 |
| SITE12 | 0.467937141 | 22.8588598 | 97.559 | 22.5895909 | 4.31953834 | 1000 |
| SITE13 | 0.533081979 | 22.811424 | 89.93 | 22.5244461 | 3.9931811 | 1000 |
| SITE14 | 0.46386191 | 22.8355528 | 97.34 | 22.5936661 | 4.30919861 | 1000 |
| SITE15 | 0.230855496 | 22.7959215 | 81.76 | 22.8266726 | 3.58232337 | 1000 |
| SITE16 | 0.460707765 | 22.765897 | 71.921 | 22.5968203 | 3.18347215 | 1000 |
| SITE17 | 0.572273339 | 22.5347011 | 41.904 | 22.4852547 | 1.86407625 | 1000 |
| SITE18 | 0.891082007 | 22.5886827 | 46.96 | 22.1664461 | 2.11892058 | 1000 |
| SITE19 | 0.939732219 | 22.5110776 | 39.456 | 22.1177958 | 1.78419397 | 1000 |
| SITE20 | 0.771935426 | 22.4844368 | 39.285 | 22.2855926 | 1.76307358 | 1000 |

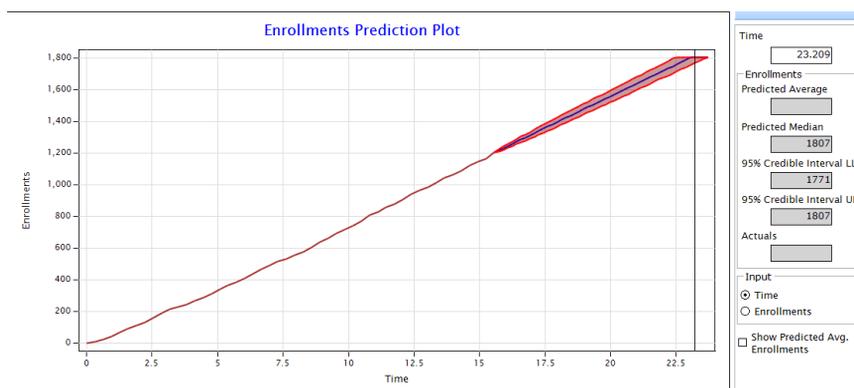
The file contains averages across all simulations for each site of the quantities such as initiation times, accrual duration, number of subjects enrolled, accrual rate, number of sites opened etc.

Open the [SiteData](#) file which stores detailed information about one simulation.

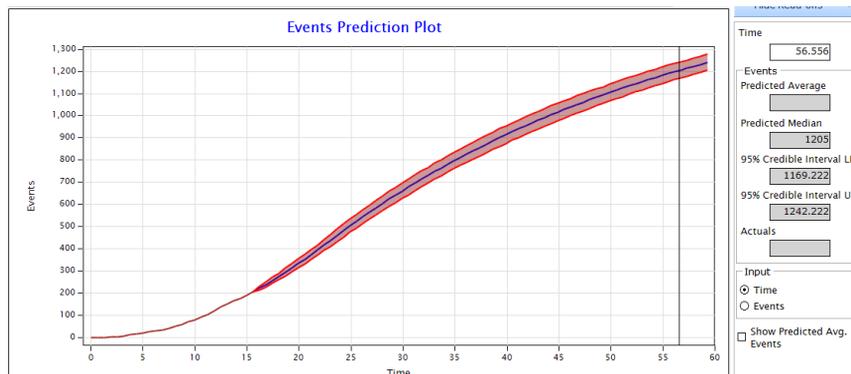
68 Enrollment/Events Prediction - Analysis

| SimulationID | SiteOp | SiteAlre | SiteID | SiteInitia | SiteAccrRate | Subjects | LastSubjectR | AccrualDurati | ObsrvdAccru |
|--------------|--------|----------|--------|------------|--------------|----------|--------------|---------------|-------------|
| 4 | 1 | 1 | SITE1 | 0 | 7.07438519 | 165 | 22.9155648 | 22.9155648 | 7.2003462 |
| 4 | 1 | 1 | SITE2 | 67776346 | 8.51424513 | 184 | 22.8790416 | 22.0477885 | 8.34550823 |
| 4 | 1 | 1 | SITE3 | 96837821 | 3.04163482 | 66 | 22.743859 | 22.818727 | 2.89236118 |
| 4 | 1 | 1 | SITE4 | 70185643 | 5.36688949 | 132 | 22.6282023 | 22.6453792 | 5.82900375 |
| 4 | 1 | 1 | SITE5 | 16564942 | 4.34414023 | 100 | 22.9121889 | 22.0989999 | 4.52509166 |
| 4 | 1 | 1 | SITE6 | 35109399 | 3.87802112 | 96 | 22.7716584 | 22.5804554 | 4.2514643 |
| 4 | 1 | 1 | SITE7 | 88711395 | 4.18946175 | 98 | 22.8111897 | 21.9268534 | 4.46940554 |
| 4 | 1 | 1 | SITE8 | 43672688 | 4.15957144 | 94 | 22.7702357 | 22.2718921 | 4.22056642 |
| 4 | 1 | 1 | SITE9 | 29872714 | 4.31723795 | 89 | 22.6057527 | 22.8856921 | 3.88889266 |
| 4 | 1 | 1 | SITE10 | 74258361 | 3.80639576 | 93 | 22.875309 | 22.3413064 | 4.162693 |
| 4 | 1 | 1 | SITE11 | 65589555 | 4.56295573 | 104 | 22.775434 | 22.0499752 | 4.71655858 |
| 4 | 1 | 1 | SITE12 | 67937141 | 4.31002551 | 93 | 22.7052255 | 22.4476277 | 4.14297677 |
| 4 | 1 | 1 | SITE13 | 33081979 | 3.99574522 | 91 | 22.816217 | 22.3824828 | 4.06567943 |
| 4 | 1 | 1 | SITE14 | 46386191 | 4.30886116 | 88 | 22.7570596 | 22.4517029 | 3.91952452 |
| 4 | 1 | 1 | SITE15 | 30855496 | 3.59050045 | 84 | 22.6555191 | 22.6847093 | 3.70293482 |
| 4 | 1 | 1 | SITE16 | 60707765 | 3.18126308 | 70 | 21.2342624 | 22.454857 | 3.11736565 |
| 4 | 1 | 1 | SITE17 | 72273339 | 1.86956062 | 41 | 21.4813529 | 22.3432915 | 1.83500269 |
| 4 | 1 | 1 | SITE18 | 91082007 | 2.11489006 | 43 | 22.2902027 | 22.0244828 | 1.95237275 |
| 4 | 1 | 1 | SITE19 | 39732219 | 1.77968558 | 39 | 22.8373258 | 21.9758326 | 1.77467679 |
| 4 | 1 | 1 | SITE20 | 71935426 | 1.75947694 | 37 | 22.7354377 | 22.1436294 | 1.67090947 |

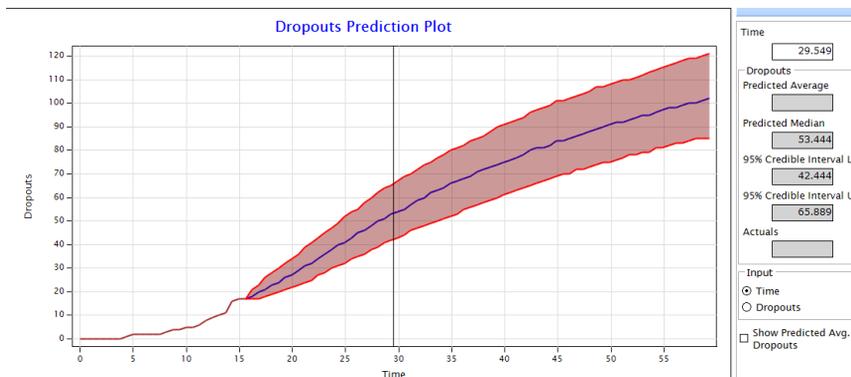
Click on the **PredictSim** node. The **Enrollments Prediction Plot** (invoked using the  tool in the **Library** pane) shows that the median number of enrollments 1807 are reached in a duration of 23.209 units of time.



Invoke the **(Events Prediction Plot)** using the  tool in the **Library** pane.



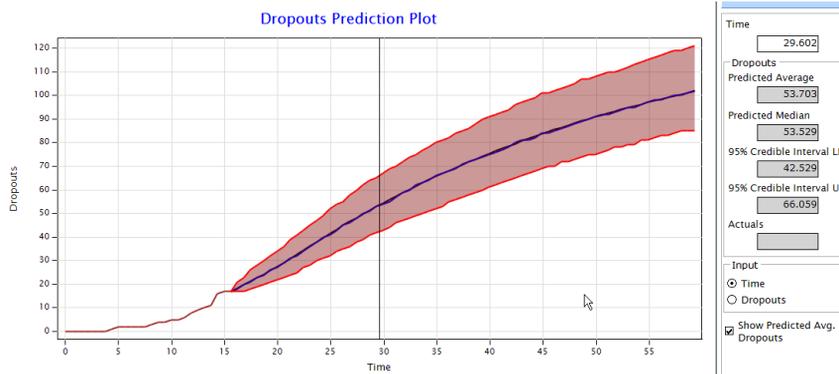
In this plot, the simulation results indicate that by the end of 56.556, the median number of events in both the control and treatment arms would be 1205 in a 95% confidence interval of 1169 to 1242. Invoke the **(Dropouts Prediction Plot)** using the  tool in the **Library** pane.



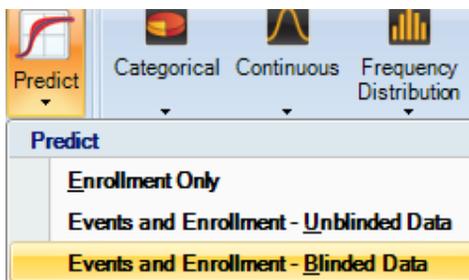
In this plot, the simulation results indicate that by the end of 29.549, the median number of drop outs in both the control and treatment arms would be 53 in a 95% confidence interval of 42 to 65.

If you select the **Show Predicted Avg. Dropouts** the predicted dropouts will be added to the plot.

68 Enrollment/Events Prediction - Analysis



Summary Data In the earlier section we saw how to generate the events and follow them till the required number of events occur or the target sample size is reached. We estimated the study duration with the help of Predict feature in East. For this to use, we assumed that an interim subject-level data was available which had information on individual arrival time, status etc. However, many a times, the subject-level data may not be available. What can be available is the summary of the accruals that have happened till date. For example, in the case of Rales trial considered above, the DMC statistician may have the information that there have been 1205 subjects accrued so far. The number of events occurred so far 206 considering both Control and Treatment. In all 17 subjects have dropped out. The last subject arrived at time 15.55. The DMC statistician is interested in knowing the total study duration when all the accrued 1205 subjects are followed till end. East through its Predict feature makes it possible to still come up with an estimate of average study duration based on simulating events from Poisson process based on the specified or default hazard rate. To see this, choose the menu item **Analysis>Predict>Events and Enrollment-Blinded Data**.



68 Enrollment/Events Prediction - Analysis

values, the input dialog looks as shown below:

Analysis: Predict: Events and Enrollment - Blinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | Simulation Controls

Survival Information

of Hazard Pieces: Input Method:

Hazard Rate (λ):

Go to the **Accruals/Dropout** tab. Suppose instead of drop out hazard rates, the information is available on the probability of drop out. Suppose the probability of drop out for a subject is 0.5% and is applicable from the current calendar time onwards which is 15.55. Give all these inputs.

Analysis: Predict: Events and Enrollment - Blinded Data (Step 2 of 2)

Target Sample Size: Target No. of Events:

Response Generation | **Accrual / Dropouts** | Simulation Controls

Current Sample Size: 1205 Subjects are followed:

Current Number Censored: 982 Current Calendar Time: 15.55

Current No. of Events: 206

Current No. of Dropouts: 17 Accrual Model: Poisson

Accrual Information

of Accrual Periods: Input Method:

| Period # | Starting at Time | Accrual Rate |
|----------|------------------|--------------|
| 1 | 0.000 | 77.492 |
| 2 | 15.550 | 38.714 |

Dropout Information

of Pieces: Input Method:

By Time:

Probability of Dropout:

Go to **Simulation Controls** tab. Give a fixed seed 12345. Save the outputs for

Summary and Subject data.

Target Sample Size: Target No. of Events:

Response Generation Accrual / Dropouts Simulation Controls

Number of Simulations:
 Refresh Frequency:

Random Number Seed
 Clock
 Fixed

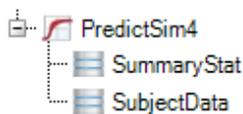
Suppress All Intermediate Output
 Pause after Refresh
 Stop at End

Output Options
 Output Type:
 Save summary statistics for every simulation run
 Save subject level data for simulation runs
 Note: Max. 100,000 records will be saved.

Output for All Trials

| Percentile (%) |
|----------------|
| 5.000 |
| 25.000 |
| 50.000 |
| 75.000 |
| 95.000 |

Click on **Simulate**. East simulates the events according to Poisson Arrival process with inter-arrival times following exponential distribution. The parameter is derived from the specified hazard rate. For details refer to the Appendix M. After a few seconds East will display the message **Simulations complete. Waiting for User's action**. Click **Close**. This will save the **Predict Simulation** in the Output Preview window first. Once you save it in the Workbook, it will create a node **PredictSim4** with sub-nodes for **SummaryStat** and **SubjectData** in the **Library**.



To view the detailed summary output of the simulations, double click the node

68 Enrollment/Events Prediction - Analysis

PredictSim4 in the Library. The following output is displayed.

Conditional Simulation: Blinded Prediction of Enrollment and Event Timelines

| Test Parameters | |
|--------------------------------|--------------------|
| Simulation ID | PredictSim4 |
| Accrual | Ongoing |
| Input | Summary Data |
| Target Sample Size | 1807 |
| Target No. of Events | 1205 |
| Response Generation Parameters | |
| Hazard Rate | 0.02683 |
| Accrual / Dropouts Parameters | |
| Subject are followed | Until End of Study |
| Prob. of Dropout by Time = | 15.55 |
| Prob. of Dropout | 0.005 |
| Simulation Control Parameters | |
| Starting Seed | Fixed |
| Number of Simulations | 1000 |

Actuals from Interim Trial Data: Sample Size and Events

| Sample Size | Events | Dropouts | Current Time |
|-------------|--------|----------|--------------|
| 1205 | 206 | 17 | 15.55 |

Conditional Simulation: Average Sample Size and Events

| Average Sample Size | Average Events | Average Dropouts | Average Accrual Duration | Average Study Duration |
|---------------------|----------------|------------------|--------------------------|------------------------|
| 1807 | 1205 | 29.134 | 31.104 | 56.187 |

Accrual / Dropouts Parameters

Accrual Input Method: Accrual Rates

| Period # | Starting at Time | Accrual Rate |
|----------|------------------|--------------|
| 1 | 0 | 77.49196 |
| 2 | 15.55 | 38.71383 |

Overall Output

| Stage | Percentile | No. of Events | No. of Accruals | No. of Dropouts | Accrual Duration | Study Duration |
|--------------|------------|---------------|-----------------|-----------------|------------------|----------------|
| Interim | Actuals | 206 | 1205 | 17 | 15.55 | 15.55 |
| End of Trial | 5% | 1205 | 1807 | 24 | 30.062 | 54.231 |
| | 25% | 1205 | 1807 | 27 | 30.683 | 55.362 |
| | 50% | 1205 | 1807 | 29 | 31.093 | 56.121 |
| | 75% | 1205 | 1807 | 31 | 31.527 | 56.965 |
| | 95% | 1205 | 1807 | 35 | 32.152 | 58.321 |
| | Average | 1205 | 1807 | 29.134 | 31.104 | 56.187 |

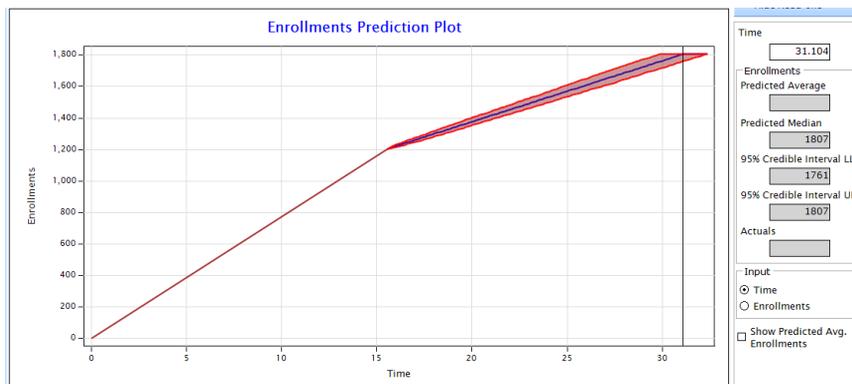
Simulation Seed and Elapsed Time

Starting Seed: 12345
 Total Number of Simulations: 1000
 Elapsed Time: 00:00:33

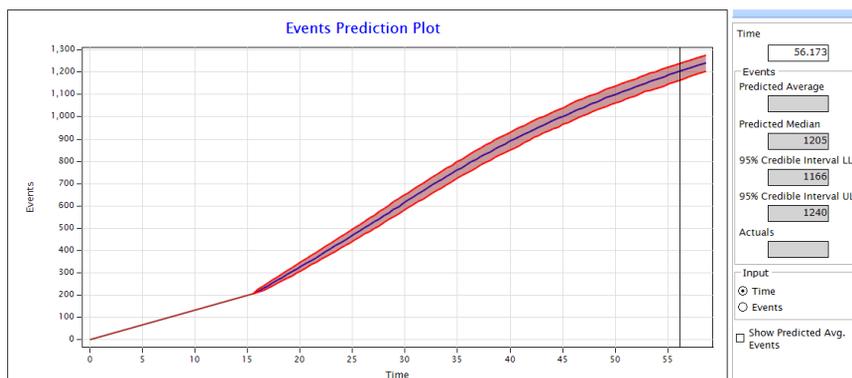
The table at the left describes the Simulation scenario. This summary contains an overview of the actual data we observed. The table **Actuals from Interim Trial Data: Sample Size and Events** presents the summary data input. The study is complete when in all 1205 events occur. The table **Average Sample Size and Events** provide information about the average study duration, average number of events, average drop outs, average follow up time etc. From this table it should be noted that it will take on an average 56.187 units of time to complete the study. The **Overall Output** table describes the details of the distribution of

Average Study Duration across 1000 simulations. Note that the column **No of Events** has all values 1205(except for actuals) meaning thereby the target sample size of 1807 was adequate in giving the required number of events. It is worth noting that the 5th percentile of the Average Study Duration is going to give 1205 events pretty early, by the time 54.231. The 95th percentile is 58.321 by which almost in all cases the target would be achieved. Click on the **PredictSim4** node. The

Enrollments Prediction Plot (invoked using the  tool in the **Library** pane) shows that the median number of enrollments 1807 are reached in a duration of 31.104 units of time.



Invoke the **(Events Prediction Plot)** using the  tool in the **Library** pane.

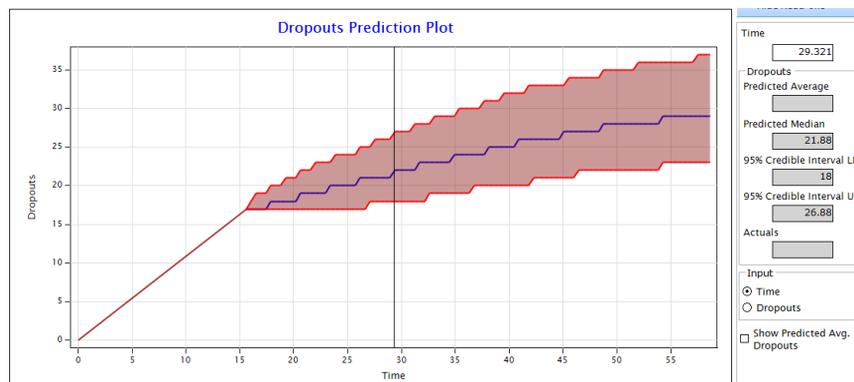


In this plot, the simulation results indicate that by the end of 56.173, the median

68 Enrollment/Events Prediction - Analysis

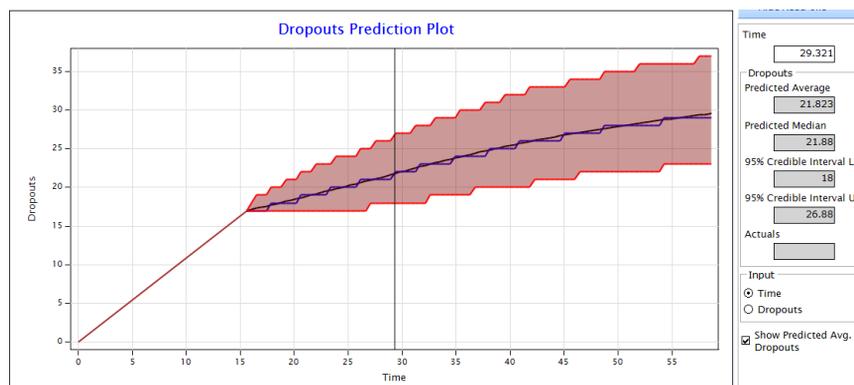
number of events in both the control and treatment arms would be 1205 in a 95% confidence interval of 1166 to 1240. Invoke the (**Dropouts Prediction Plot**) using the

 tool in the **Library** pane.



In this plot, the simulation results indicate that by the end of 29.321, the median number of drop outs in both the control and treatment arms would be 22 in a 95% confidence interval of 18 to 27.

If you select the **Show Predicted Avg. Dropouts** the predicted dropouts will be added to the plot.



69 *Interfacing with East PROCs*

69.1 *What is East PROCs*

East PROCs is a special version of **East 6.3** developed specially for *SAS*^(R) users. **East PROCs** contains an external SAS procedure for Interim Monitoring of a design created in **East 6.3**. While it has all the capabilities of **East** Interim Monitoring, it requires *SAS*^(R) system on your machine.

Proc EASTMONITOR from **East PROCs**, through its various options can perform monitoring of the following group sequential designs in **East**.

1. Continuous Endpoints
 - One Sample: Single mean
 - One Sample: Mean of paired differences
 - Two Samples: Difference of means from independent populations
2. Discrete Endpoints
 - One Sample: Single binomial proportion
 - One Sample: McNemar's for matched pairs of binomial responses
 - Two Samples: Difference of binomial proportions from independent populations
 - Two Samples: Ratio of binomial proportions from independent populations
 - Two Samples: Odds ratio of proportions from independent populations
 - Two Samples: Common odds ratio for stratified 2x2 tables
3. Survival Endpoints
 - Two Samples: Logrank test given accrual duration and accrual rates
 - Two Samples: Logrank test given accrual duration and study duration
4. General
 - Information based
 - Sample Size based

The trials can be Superiority or Noninferiority with either or both efficacy and futility boundaries. For details of combinations of efficacy and futility boundaries, boundary families etc allowed per test, the user is referred to **East 6** user manual.

Apart from Interim Monitoring of the above mentioned group sequential designs, **Proc EastMonitor** can also monitor Adaptive Trials created in **East 6** based on the following tests:

- Continuous: Difference of means from two independent populations

69 Interfacing with East PROCs

- Binary: Difference of proportions from two independent populations
- Binary: Ratio of proportions from two independent populations
- Survival: Logrank test given accrual duration and accrual rates
- Survival: Logrank test given accrual duration and study duration

The syntax for each of the above mentioned interim monitoring is described in the **East PROCs** user manual.

69.2 Why Proc EastMonitor

Clinical trial data are generally analyzed using SAS. **Proc EastMonitor** has been developed to enable the interim analysis of clinical trial data using SAS considering that the **East** has been used for designing the study. In other words, you don't need **East** to be available for interim monitoring. What you need is a design created in **East**. This design and interim look data are inputs to **Proc EastMonitor**. **Proc EastMonitor** then performs the interim analysis exactly the same way as East Interim Monitoring module would have done it. This is possible because **Proc EastMonitor** calls the East interim monitoring programs internally. The resulting output is available in SAS data sets as well as in the list files which include the decisions of the interim analysis regarding the continuation of the trial or otherwise. The generated output data sets can be subjected to SAS' graphical and reporting tools for creating reports as per requirement.

East being the pioneering software in designing of phase 3 clinical trials, encompasses numerous combinations of efficacy and futility boundaries and other features such as accrual, drop out etc. The boundaries that are available in **East** run the gamut between extreme conservatism and extreme liberality for early stopping. It can also handle the designs with missing efficacy or futility boundaries at some looks. All these designs can be monitored using **Proc EastMonitor**. Besides the non-adaptive designs, **East** can formulate adaptive designs following Cui, Hung and Wang (1999). This adaptive design allows modification of sample size and effect size at an interim look. In effect, the adaptive designs are also amenable for interim monitoring in SAS through Proc EastMonitor.

As a result, with **Proc EastMonitor** as an add on to SAS, the whole interim monitoring capability of **East** becomes available in SAS and will continue to be so for further new designs in **East**.

69.3 Continuous Endpoint: Orlistat Trial

Consider the Orlistat trial described in Section 10.1.1 where we would like to test the null hypothesis that treatment does not lead to weight loss, $H_0: \delta = 0$, against the alternative hypothesis that the treatment does result in a loss of weight, $H_1: \delta = 3$. Suppose we have designed this trial in East 6 and following is the detailed output of this design.

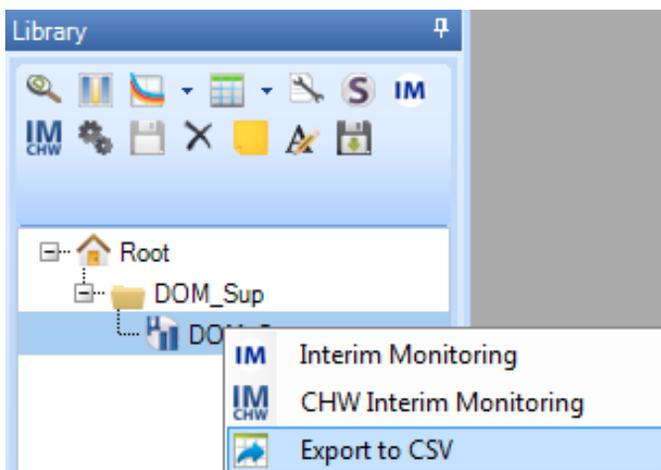
Design: Continuous Endpoint: Two-Sample Test - Parallel Design - Difference of Means

| Test Parameters | | | |
|--------------------------------|------------------|-------------|--|
| Design ID | DOM_Sup | Superiority | |
| Design Type | Superiority | | |
| Number of Looks | 3 | | |
| Test Type | 1-Sided | | |
| Specified α | 0.05 | | |
| Power | 0.9 | | |
| Model Parameters | | | |
| Test Statistic | Z | | |
| Input Method | Individual Means | | |
| Mean Control (μ_c) | 6 | | |
| Mean Treatment (μ_t) | 9 | | |
| $\delta = \mu_t - \mu_c$ | | | |
| Under H_0 | 0 | | |
| Under H_1 | 3 | | |
| Std. Deviation (σ) | 8 | | |
| Allocation Ratio (n_t/n_c) | 3 | | |
| Boundary Parameters | | | |
| Spacing of Looks | Equal | | |
| Efficacy Boundary | LD (OF) | | |

| Sample Size Information | | | |
|------------------------------|-------------|---------------|---------|
| | Control Arm | Treatment Arm | Total |
| Sample Size (n) | | | |
| Maximum | 83 | 248 | 331 |
| Expected H_1 | 64.143 | 192.448 | 256.591 |
| Expected H_0 | 82.521 | 246.594 | 329.115 |
| Maximum Information (I) 0.97 | | | |

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries Efficacy Z | Boundary Crossing Probability (Incremental) | |
|--------|--------------------------|-----------------|---------------------------|-----------------------|---|-------------|
| | | | | | Under H_0 | Under H_1 |
| 1 | 0.332 | 110 | 6.741E-4 | 3.205 | 6.741E-4 | 0.066 |
| 2 | 0.668 | 221 | 0.016 | 2.139 | 0.016 | 0.543 |
| 3 | 1 | 331 | 0.05 | 1.695 | 0.034 | 0.291 |

Let us monitor this trial using PROC EASTMONITOR. Save the design details into the CSV format. Right click on the design node in the Library and select Export to CSV.



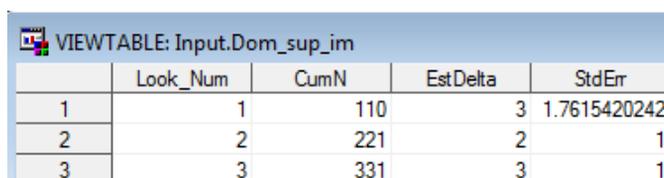
This CSV file will serve as an input to the PROC EASTMONITOR.

69 Interfacing with East PROCs

Launch East PROCs and import the above CSV file in SAS. The code in SAS to import the file may look like as shown below:

```
PROC IMPORT OUT= INPUT.orlistat_des
            DATAFILE= "D:\Work\EAST6.3\ProcIM\Orlistat\Orlistat.csv"
            DBMS=CSV REPLACE;
            GETNAMES=YES;
            DATAROW=2;
RUN;
```

Now suppose you have the data to be used for interim monitoring in a SAS file.



| | Look_Num | CumN | EstDelta | StdErr |
|---|----------|------|----------|--------------|
| 1 | 1 | 110 | 3 | 1.7615420242 |
| 2 | 2 | 221 | 2 | 1 |
| 3 | 3 | 331 | 3 | 1 |

The following code reads the design information from the dataset **orlistat_des**; IM data from the dataset **Orlistat_im**; monitors the trial; computes the look-by-look output quantities and saves the output in the form of a SAS datasets.

```
libname input "D:\Work\EAST6.3\ProcIM\Orlistat";run;
libname out "D:\Work\EAST6.3\ProcIM\Orlistat\out";run;

options nodate nonumber;
PROC EASTMONITOR DESIGN=input.Orlistat_Des DATA=input.Orlistat_im;
CONDPower OUT=out.cp_Orlistat ;
PHP OUT=out.php_Orlistat ;
ERRSPEND OUT=out.errspd_Orlistat ;
CI OUT=out.ci_Orlistat;
BOUNDARY OUT=out.bdd_Orlistat ;
OUTPUT OUT=out.IM_INFO_Orlistat ;
run;
```

The output from PROC EASTMONITOR can be seen in the Output window of SAS. It is divided into two parts - Design Output and IM Output. The Design Output part contains all the information exported from East 6. The IM output is actually the output we are interested in.

The SAS System

Output from East (r) PROCs (v1.0) under _SAS9_2 or latter
 Copyright (c) 1987-2014 Cytel Inc., Cambridge, MA, USA.

INTERIM MONITORING: DIFFERENCE OF MEANS

Design Input Parameters

```
-----
Design ID       : DOM_Sup
Design DataSet  : <INPUT.ORLISTAT_DES>
IM DataSet      : <INPUT.ORLISTAT_IM>
-----
```

Test Parameters

```
-----
Design Type     : Superiority
No. of Looks    : 3
Test Type       : 1-Sided
Specified Alpha : 0.0500
Power           : 0.9001
-----
```

Model Parameters

```
-----
Input Method           : Individual Means
Diff. in Mean          : 3.0000
Mean Control           : 6.0000
Mean Treatment         : 9.0000
Std. Deviation         : 8.0000
Test Statistic         : Z
Allocation Ratio(nt/nc): 3.0000
-----
```

Boundary Parameters

Efficacy Boundary : LD(OF)

The SAS System

Output from East (r) PROCs (v1.0) under _SAS9_2 or latter
 Copyright (c) 1987-2014 Cytel Inc., Cambridge, MA, USA.

Detailed Design: Two-Sample Test- Parallel Design - Difference of Means

Sample Size Information

| | Control Arm | Treatment Arm | Total |
|-----------------|----------------|------------------|----------|
| Sample Size (n) | | | |
| Maximum: | 83 | 248 | 331 |
| Expected H1: | 64.1428 | 192.4477 | 256.5906 |
| Expected H0: | 82.5210 | 246.5940 | 329.1150 |

Maximum Information for this design is 0.9697

The SAS System

69 Interfacing with East PROCs

Output from East (r) PROCs (v1.0) under _SAS9_2 or latter
 Copyright (c) 1987-2014 Cytel Inc., Cambridge, MA, USA.

Stopping Boundaries: Look by Look

| Look No. | Info Fract (n/n_max) | Sample Size (n) | Cumulative Alpha Spent | Boundaries Efficacy (Z) | Boundary Crossing Probability (Incremental) | Under H0 Efficacy | Under H1 Efficacy |
|----------|----------------------|-----------------|------------------------|-------------------------|---|-------------------|-------------------|
| 1 | 0.3323 | 110 | 0.0007 | 3.2055 | 0.0007 | | 0.0665 |
| 2 | 0.6677 | 220 | 0.0165 | 2.1387 | 0.0158 | | 0.5429 |
| 3 | 1.0000 | 330 | 0.0500 | 1.6950 | 0.0335 | | 0.2907 |

The SAS System

Output from East (r) PROCs (v1.0) under _SAS9_2 or latter
 Copyright (c) 1987-2014 Cytel Inc., Cambridge, MA, USA.

Interim Monitoring Output

| Look No | Information Fraction | Cumulative Sample Size | Test Statistic | Efficacy |
|---------|----------------------|------------------------|----------------|----------|
| 1 | 0.3323 | 110 | 1.7031 | 3.2055 |
| 2 | 0.6677 | 221 | 2.0000 | 2.1387 |
| 3*& | 1.0000 | 331 | 3.0000 | 1.6950 |

| Look No. | Cumulative Sample Size | Effect Size | Standard Error | Repeated Lower | 95.00% CI Upper | Repeated p-value |
|----------|------------------------|-------------|----------------|----------------|-----------------|------------------|
| 1 | 110 | 3.0000 | 1.7615 | -2.6466 | Infinity | 0.2462 |
| 2 | 221 | 2.0000 | 1.0000 | -0.1387 | Infinity | 0.0636 |
| 3*& | 331 | 3.0000 | 1.0000 | 1.3050 | Infinity | 0.0014 |

| Look No. | Cumulative Sample Size | INLP | CP | Predictive Power |
|----------|------------------------|------|--------|------------------|
| 1 | 110 | 326 | 0.9438 | 0.8229 |
| 2 | 221 | 331 | 0.9041 | 0.8570 |
| 3*& | 331 | NA | NA | NA |

*: At Look 3 the value of Test Statistic is >= the critical point for efficacy, H0 is rejected.

&: At Look 3 with the current cumulative sample size, the desired power is achieved or exceeded. In order to preserve the operating characteristics of the study, East has forced this to be the last look.

The SAS System

Output from East (r) PROCs (v1.0) under _SAS9_2 or latter
 Copyright (c) 1987-2014 Cytel Inc., Cambridge, MA, USA.

Final Inference

```
Final Outputs at Look          : 3
Adj. p-value                   : 0.0167
Adj. Pt. Est. for Effect Size  : 2.5047

Adj. 90.00% CI for Effect Size

Upper confidence bound         : 4.3144
Lower confidence bound         : 0.5825
Post-Hoc Power                 : 0.9001
```

Notice that the **PROC EASTMONITOR** prints the decision which was taken at the end of the trial. In the end, it prints the final inference as well.

One can also see this output back in the East6. To do that, export the output from SAS to a CSV file. The code to export the output dataset may look like as shown below:

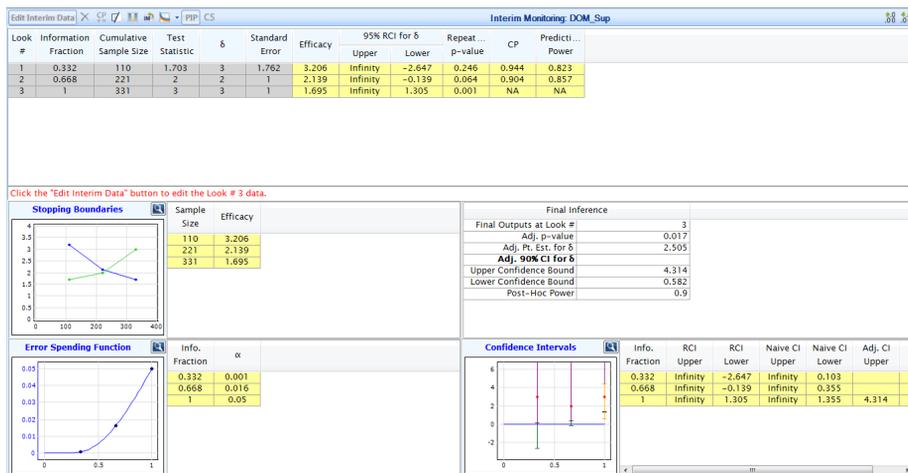
```
PROC EXPORT DATA= OUT.Im_info_orlistat
             OUTFILE= "D:\Work\EAST6.3\ProcIM\Orlistat\out\imout_orlistat.csv"
             DBMS=CSV REPLACE;
             PUTNAMES=YES;
RUN;
```

Activate East 6 and go back to the design node in **Library**. Insert a new IM dashboard by clicking the **IM** icon. Right click on the Interim Monitoring node and select **Import PROC EM Output**. Import the CSV file **imout_orlistat.csv**.

The Interim Monitoring dashboard gets updated with the output from **PROC**

69 Interfacing with East PROCs

EASTMONITOR.



Volume 9 *Analysis*

| | |
|--|-------------|
| <i>70 Introduction to Volume 9</i> | <i>1798</i> |
| <i>71 Tutorial: Analysis</i> | <i>1806</i> |
| <i>72 Analysis-Descriptive Statistics</i> | <i>1827</i> |
| <i>73 Analysis-Analytics</i> | <i>1837</i> |
| <i>74 Analysis-Plots</i> | <i>1854</i> |
| <i>75 Analysis-Normal Superiority One-Sample</i> | <i>1890</i> |
| <i>76 Analysis-Normal Noninferiority Paired-Sample</i> | <i>1901</i> |
| <i>77 Analysis-Normal Equivalence Paired-Sample</i> | <i>1907</i> |
| <i>78 Analysis-Normal Superiority Two-Sample</i> | <i>1913</i> |
| <i>79 Analysis-Normal Noninferiority Two-Sample</i> | <i>1926</i> |
| <i>80 Analysis-Normal Equivalence Two-Sample</i> | <i>1941</i> |
| <i>81 Analysis-Nonparametric Two-Sample</i> | <i>1956</i> |

| | | |
|----|--|------|
| 82 | <i>Analysis-ANOVA</i> | 1976 |
| 83 | <i>Analysis-Regression Procedures</i> | 1987 |
| 84 | <i>Analysis-Multiple Comparison Procedures for Continuous Data</i> | 2024 |
| 85 | <i>Analysis-Multiple Endpoints for Continuous Data</i> | 2055 |
| 86 | <i>Analysis-Binomial Superiority One-Sample</i> | 2060 |
| 87 | <i>Analysis-Binomial Superiority Two-Sample</i> | 2069 |
| 88 | <i>Analysis-Binomial Noninferiority Two-Sample</i> | 2088 |
| 89 | <i>Analysis-Binomial Equivalence Two-Samples</i> | 2106 |
| 90 | <i>Analysis-Discrete: Many Proportions</i> | 2111 |
| 91 | <i>Analysis-Binary Regression Analysis</i> | 2131 |
| 92 | <i>Analysis- Multiple Comparison Procedures for Binary Data</i> | 2180 |
| 93 | <i>Analysis-Comparison of Multiple Comparison Procedures for Continuous Data- Analysis</i> | 2207 |
| 94 | <i>Analysis-Multiple Endpoints for Binary Data</i> | 2211 |

| | |
|---|-------------|
| <i>95 Analysis-Agreement</i> | <i>2216</i> |
| <i>96 Analysis-Survival Data</i> | <i>2219</i> |
| <i>97 Analysis-Multiple Comparison Procedures for Survival Data</i> | <i>2240</i> |

70 *Introduction to Volume 9*

This volume describes the procedures for analyzing data for continuous, binary, discrete and survival endpoints. Analysis of data arising from clinical trials with one arm, two arm as well as multiple arms is possible with the help of the **Analysis** module of **East**. The procedures include **Basic Statistics** and **Plots** used for exploratory analysis of data and at higher level, Logistic and Probit regression as well as tests for handling analysis of crossover data. Exact Inference tests for two by two categorical data and multiple comparison tests for continuous and discrete data also belong to the Analysis menu. For a few tests, a link to SAS[®] is provided which enable user to perform analysis in SAS and display the output in **East**.

Chapter 4 introduces the data editor features such as creating a new data, manipulating existing data, sorting, filtering, transforming variables, generating random numbers from distributions etc. **East** caters to Case Data and Crossover Data.

Chapter 71 explains the workflow in Analysis used for analyzing any data. This chapter describes how you can use the data editor capabilities effectively and perform the statistical test you want.

Chapter 72 deals with preliminary exploration of data using elementary tools such as computation of summary measures, classification, cross tabulation of the data. Descriptive Statistics helps statisticians to choose statistical analysis techniques to arrive at meaningful inference.

Chapter 73 describes some of the commonly used univariate procedures: t-test (paired and independent), one-way and two-way (without interaction) analysis of variance (*ANOVA*) and multiple linear regression. The topics of correlations and Multivariate Analysis of Variance (*MANOVA*) are also included in this chapter.

Chapter 74 deals with data exploration plots for case data and crossover data.

Chapter 75 demonstrates how **East** can be used to perform inferences on data collected from a single-sample superiority study with continuous endpoint. This may consist of a random sample of observations from either a single treatment or paired observations from two treatments.

Chapter 76 explores how we can use **East** to perform inference on continuous data collected from a paired-sample noninferiority study.

Chapter 77 demonstrates how inference on continuous data collected from a paired-sample equivalence study can be performed.

Chapter 78 deals with analysis of continuous data coming from two independent samples and crossover superiority studies.

Chapter 79 deals with analysis of continuous data coming from two independent samples and crossover noninferiority studies.

Chapter 80 explains how we can use **East** to perform analysis of continuous data that comes from two independent samples and crossover equivalence studies.

Chapter 81 describes analysis using Wilcoxon-Mann-Whitney nonparametric test for parallel as well as crossover designs. Analysis of data from both superiority and noninferiority studies is possible.

Chapter 82 focuses on Analysis of Variance (*ANOVA*). The technique is useful in clinical trial data analysis whenever there are multiple responses or multiple doses of an experimental drug being compared with placebo. The chapter deals with **One way**, **Two way** and **One way repeated measures ANOVA** with SAS connection.

Chapter 83 demonstrates how to run regression analysis in **East**. East can perform multiple linear regression, repeated measure regression and fit linear mixed effect (LME) model on data obtained from 2×2 crossover design. Link to SAS is also available for the repeated measures regression and linear mixed effects model.

Chapter 84 deals with multiple comparison procedures in which multiple treatments are compared against a placebo or active control. The response is of continuous type. The procedures included are parametric and p-value based. For multiple comparison procedures in **East** we can either provide the dataset containing the observations under each arm or the raw p-values to obtain the adjusted p-values.

Chapter 86 demonstrates how to perform inferences on data collected from a single-sample superiority study when the observations on a binary variable have an unknown probability of success. You need to either test a null hypothesis about the probability, or compute an exact confidence interval for the probability of success. The section also discusses the analysis of paired data on a binary random variable. The chapter also discusses Exact test for paired samples.

Chapter 87 explores how to analyze data from two independent binomial samples

70 *Introduction to Volume 9*

generated while conducting a superiority trial. This comparison is based on difference of response probabilities, ratio of proportions or odds ratio of the two populations. Exact inference in case of difference of proportions and ratio of proportions is described.

Chapter 88 deals with noninferiority trials involving data from two independent binomial samples. This comparison is based on difference of proportions, ratio of proportions or odds ratio of the two populations. For difference of proportions and ratio of proportions exact inference is supported which is described in this chapter.

Chapter 89 explains how we can use **East** to perform analysis of data that comes from two independent binomial samples equivalence studies. Both asymptotic and Exact options are described.

Chapter 90 deals with situations for discrete data, where the data are either coming from many binomial populations or the responses are from multinomial distribution. In case of multiple binomial populations, the interest lies in testing whether the success probability differs across several binomial populations, in particular does it increase or decrease with reference to an index variable. For data coming from multinomial distributions, one is interested in testing if the cell probabilities are according to some theoretical law. **East** can be used to analyze both these types of data. Chi-square tests, Wilcoxon rank sum test for ordered categorical data, trend in R ordered populations are some of the tests described in this chapter.

Chapter 91 focuses on how to run binary regression analysis. **East** provides logistic, probit, and complementary log-log regression models for data with a binary response variable. Along with regular maximum likelihood inference for logistic model, **East** provides Firth bias-correction for asymptotic estimates for unstratified logistic regression. Profile likelihood based confidence intervals for estimates are available for unstratified data.

Chapter 92 explains how to analyze data arising out of multiple comparison studies where more than one treatment are compared against a placebo or active control. The procedures included are parametric and p-value based. For multiple comparison procedures in **East** we can either provide the dataset containing the observations under each arm or the raw p-values to obtain the adjusted p-values.

Chapter 93 deals with comparison of different multiple testing procedures for continuous end point through an illustrative example.

Chapter 95 discusses Cohen's Kappa and the Weighted Kappa measures. These two measures are used to assess the level of agreement between two observers classifying a sample of objects on the same categorical scale.

Chapter 96 deals with comparison of two survival curves using Logrank Test in superiority and noninferiority studies. The chapter also demonstrates how one can obtain a plot of multi-arm Kaplan Meier Estimator in **East**.

Chapter 97 explains how to analyze data arising out of multiple comparison studies with survival endpoint where more than one treatment are compared against a placebo or active control.

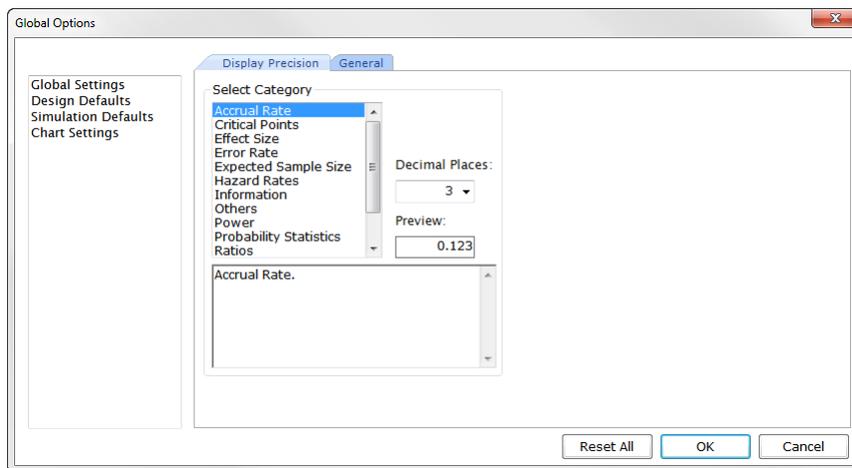
The following section discusses the Global Options in East 6. Most of them are not applicable to Analysis menu but some options like Data Path or Display Precision settings for analysis can be set.

70 Introduction to Volume 9

70.1 Settings



Click the **Global Options** icon in the **Home** menu to adjust default values in East 6.

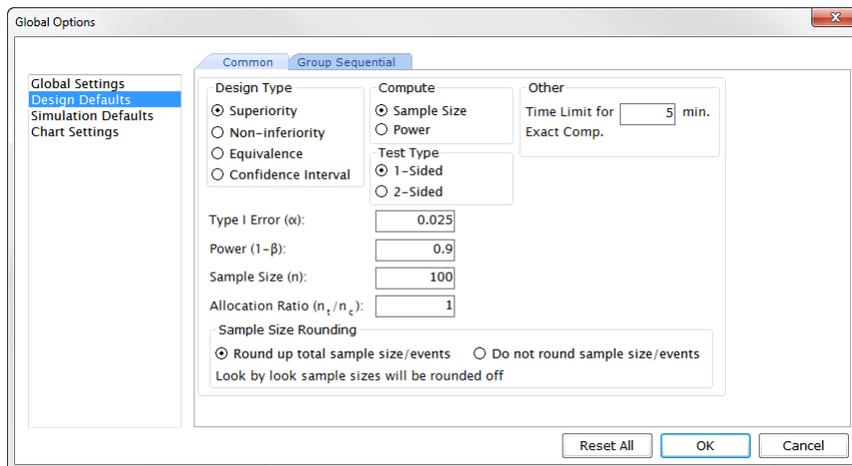


The options provided in the **Display Precision** tab are used to set the decimal places of numerical quantities. The settings indicated here will be applicable to all tests in East 6 under the **Design** and **Analysis** menus.

All these numerical quantities are grouped in different categories depending upon their usage. For example, all the average and expected sample sizes computed at simulation or design stage are grouped together under the category "Expected Sample Size". So to view any of these quantities with greater or lesser precision, select the corresponding category and change the decimal places to any value between 0 to 9.

The **General** tab has the provision of adjusting the paths for storing workbooks, files, and temporary files. These paths will remain throughout the current and future sessions even after East is closed. This is the place where we need to specify the installation directory of the R software in order to use the feature of R Integration in East 6.

The **Design Defaults** is where the user can change the settings for trial design:



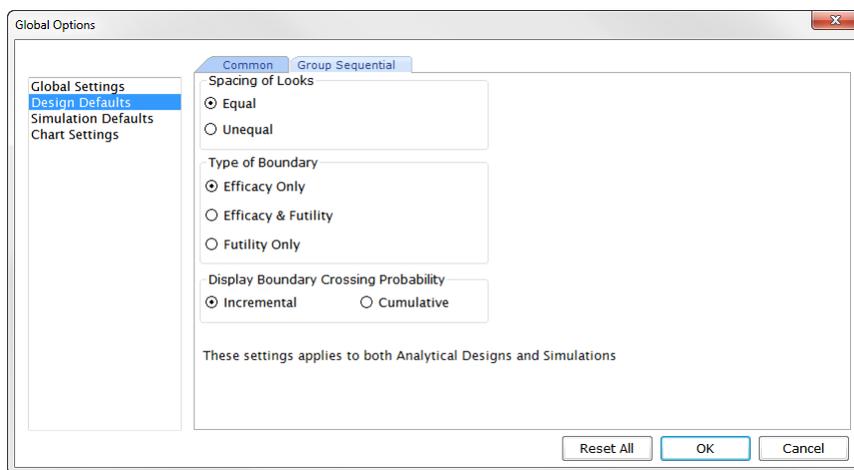
Under the **Common** tab, default values can be set for input design parameters.

You can set up the default choices for the design type, computation type, test type and the default values for type-I error, power, sample size and allocation ratio. When a new design is invoked, the input window will show these default choices.

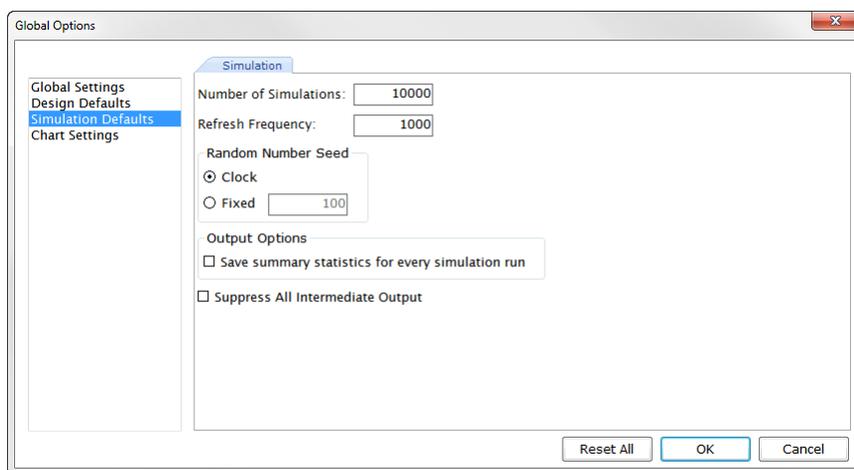
- **Time Limit for Exact Computation**
 This time limit is applicable only to exact designs and charts. Exact methods are computationally intensive and can easily consume several hours of computation time if the likely sample sizes are very large. You can set the maximum time available for any exact test in terms of minutes. If the time limit is reached, the test is terminated and no exact results are provided. Minimum and default value is 5 minutes.
- **Type I Error for MCP**
 If user has selected 2-sided test as default in global settings, then any MCP will use half of the alpha from settings as default since MCP is always a 1-sided test.
- **Sample Size Rounding**
 Notice that by default, East displays the integer sample size (events) by rounding up the actual number computed by the East algorithm. In this case, the look-by-look sample size is rounded off to the nearest integer. One can also see the original floating point sample size by selecting the option "Do not round sample size/events".

70 Introduction to Volume 9

Under the **Group Sequential** tab, defaults are set for boundary information. When a new design is invoked, input fields will contain these specified defaults. We can also set the option to view the Boundary Crossing Probabilities in the detailed output. It can be either Incremental or Cumulative.



Simulation Defaults is where we can change the settings for simulation:

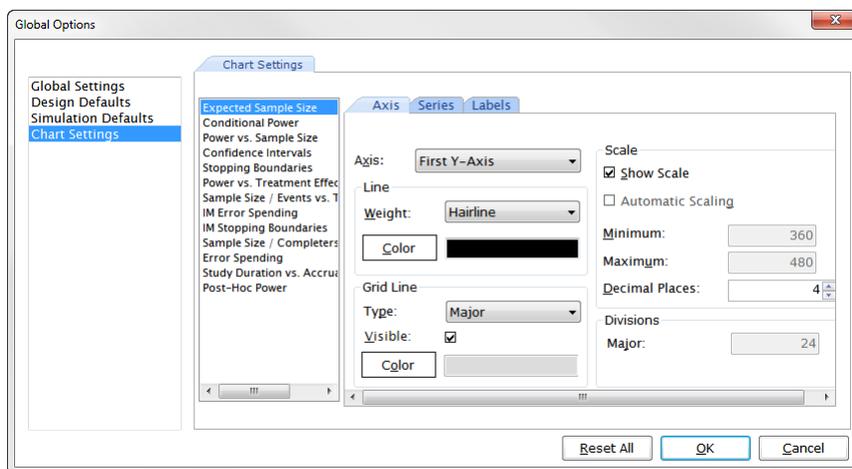


If the checkbox for "Save summary statistics for every simulation" is checked, then East simulations will by default save the per simulation summary data for all the

simulations in the form of a case data.

If the checkbox for "Suppress All Intermediate Output" is checked, the intermediate simulation output window will be always suppressed and you will be directed to the **Output Preview** area.

The **Chart Settings** allows defaults to be set for the following quantities on East6 charts:



We suggest that you do not alter the defaults until you are quite familiar with the software.

71 *Tutorial: Analysis*

The Analysis menu of **East 6.3** contains various procedures for analyzing data. The procedures include **Basic Statistics** and **Plots** used for exploratory analysis of data and at higher level, Logistic and Probit regression as well as tests for handling analysis of crossover data. Exact Inference tests for two by two categorical data and multiple comparison tests for continuous and discrete data belong to the Analysis menu.

For a few tests, we provide a link to SAS[®] for invoking SAS[®] procedure or user's SAS program which will do the analysis in SAS[®] and display the results on **East** screen.

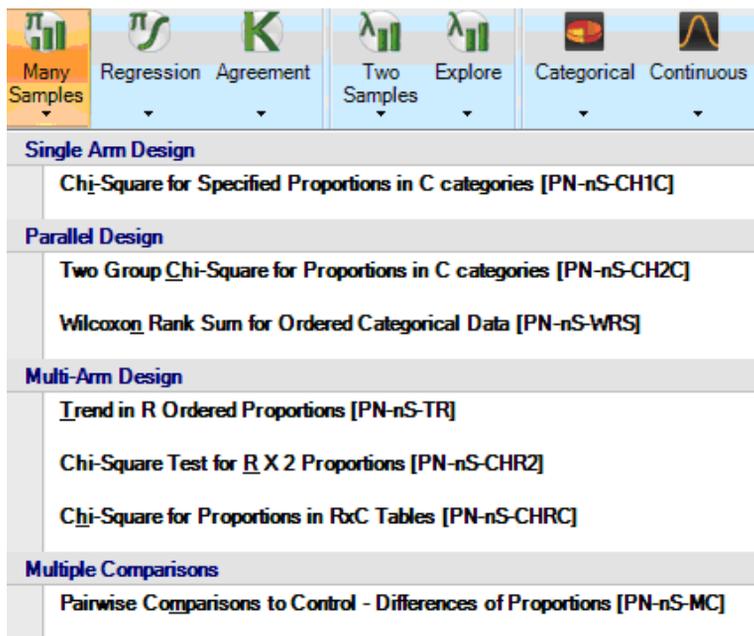
All the procedures in the Analysis menu are broadly grouped under the following categories.

- Basic Statistics
- Continuous
- Discrete
- Events
- Basic Plots
- Crossover Plots

Each of these categories is further divided into several sub menus consisting of the procedures related to that particular category. For example, if you traverse

Analysis > (Discrete) Many Samples

You will see the following list of available procedures.



Note that the procedures are grouped under **Single Arm Design**, **Parallel Design** etc. In this tutorial, we will take you on a tour of Analysis procedures available in **East**.

71.1 DataTypes

East can do analysis on data in **Case Data** or **Crossover Data** formats. Except for the procedures specifically marked for crossover analysis, all other procedures can be carried out on **Case Data**. **Case data** can be viewed and modified using the **case data editor**. The case data editor displays case data in the form of a sheet where Rows represent records and Columns represent variables. The variables can be of binary, string, categorical or continuous types. You can create a new Case Data sheet by clicking **New Data** on **File menu**. For more details about **Case Data Editor** and **Cross over Data Editor** refer to the Chapter 4.

For illustrative purposes we have included Data files in the **Samples** folder available in the **Installation Directory** of **East**. A typical **Case data file** in a case data editor looks as shown below:

71 Tutorial: Analysis

| | Dose | Animal | Week | Weight | var |
|----|------|--------|------|--------|-----|
| 1 | No | 1 | 1 | 455 | |
| 2 | No | 1 | 3 | 460 | |
| 3 | No | 1 | 4 | 510 | |
| 4 | No | 1 | 5 | 504 | |
| 5 | No | 1 | 6 | 436 | |
| 6 | No | 1 | 7 | 466 | |
| 7 | No | 2 | 1 | 467 | |
| 8 | No | 2 | 3 | 565 | |
| 9 | No | 2 | 4 | 610 | |
| 10 | No | 2 | 5 | 596 | |
| 11 | No | 2 | 6 | 542 | |
| 12 | No | 2 | 7 | 587 | |
| 13 | No | 3 | 1 | 445 | |
| 14 | No | 3 | 3 | 530 | |
| 15 | No | 3 | 4 | 580 | |
| 16 | No | 3 | 5 | 597 | |
| 17 | No | 3 | 6 | 582 | |
| 18 | No | 3 | 7 | 619 | |
| 19 | No | 4 | 1 | 485 | |
| 20 | No | 4 | 3 | 542 | |
| 21 | No | 4 | 4 | 594 | |
| 22 | No | 4 | 5 | 583 | |

This is a view of **Body_weight.cyd** data opened in **East**. These data contain 22 records on four variables: Dose, Animal, Week and Weight. While **Dose** is string variables, all others are of numeric type.

A typical **Crossover Data file** in crossover data editor looks as follows:

No. of Groups:

No. of Periods:

Treatment Assignment Table

| | P1 | P2 |
|----|----|----|
| G1 | T | R |
| G2 | R | T |

| | PatientId | GroupId | P1_Resp | P2_Resp | var |
|----|-----------|---------|---------|---------|-----|
| 1 | 1 | G1 | 228.04 | 288.79 | |
| 2 | 2 | G2 | 339.03 | 329.76 | |
| 3 | 3 | G1 | 288.21 | 343.37 | |
| 4 | 4 | G2 | 242.64 | 258.19 | |
| 5 | 5 | G2 | 249.94 | 201.56 | |
| 6 | 6 | G1 | 217.97 | 225.77 | |
| 7 | 7 | G1 | 133.13 | 235.89 | |
| 8 | 8 | G2 | 184.32 | 249.64 | |
| 9 | 9 | G1 | 213.78 | 215.14 | |
| 10 | 10 | G1 | 248.98 | 245.48 | |
| 11 | 11 | G1 | 163.93 | 134.89 | |
| 12 | 12 | G2 | 209.3 | 231.98 | |
| 13 | 13 | G2 | 207.4 | 234.19 | |
| 14 | 14 | G1 | 245.92 | 223.39 | |
| 15 | 15 | G2 | 239.84 | 241.25 | |
| 16 | 16 | G2 | 211.24 | 255.6 | |

The above is a view of the **Euphylong.cyd file**, that contains 2x2 crossover data on two drugs T and R administered in two sequences G1=T,R and G2=R,T. The variables P1_Resp and P2_Resp are the variables representing responses of patients in the first and second periods respectively.

71 Tutorial: Analysis

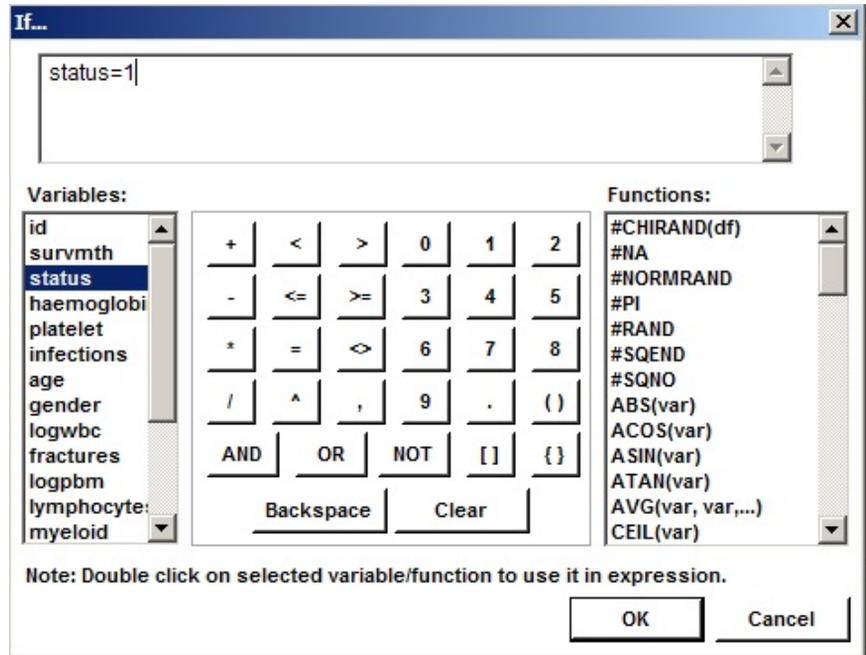
71.2 Using Case Data Editor Features

There are a whole lot of capabilities available with the **Case Data editor**. You can sort, filter transform variables, etc., and perform the analysis on the modified data. For instance, suppose you open the data set **Leukemia.cyd** from **Samples** folder. The data consist of three variables **Drug**, **Time** and **Status**, a part of which is shown below.

| | Drug | Time | Status |
|----|---------|------|--------|
| 1 | Placebo | 1 | 1 |
| 2 | 6-MP | 10 | 1 |
| 3 | Placebo | 22 | 1 |
| 4 | 6-MP | 7 | 1 |
| 5 | Placebo | 3 | 1 |
| 6 | 6-MP | 32 | 0 |
| 7 | Placebo | 12 | 1 |
| 8 | 6-MP | 23 | 1 |
| 9 | Placebo | 8 | 1 |
| 10 | 6-MP | 22 | 1 |
| 11 | Placebo | 17 | 1 |
| 12 | 6-MP | 6 | 1 |
| 13 | Placebo | 2 | 1 |
| 14 | 6-MP | 16 | 1 |

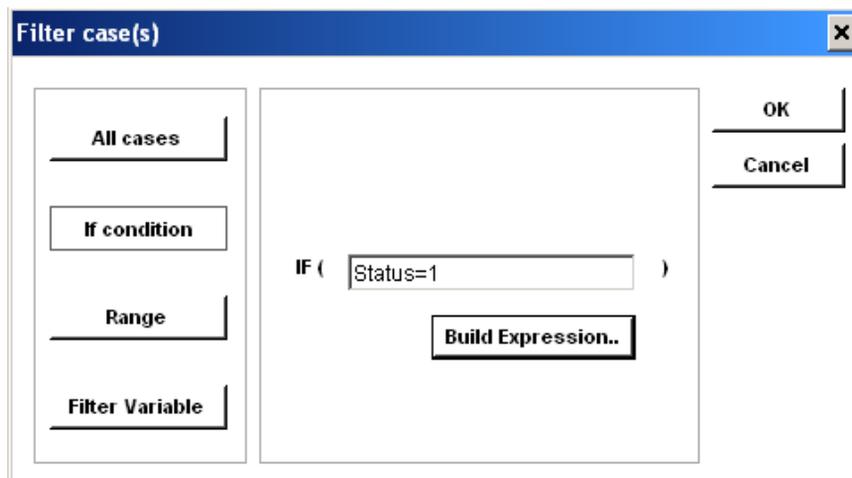
Suppose you want to consider data only for **Status =1**. This is possible by filtering the data using filter command as described below.

Click on the  filter icon on the **Data Editor** menu. In the ensuing dialog box, press **If Condition** and then **Build expression** button. You will see the following dialog box.



71 Tutorial: Analysis

Use the selections available to build the expression as shown below.



Press **OK**. This will select the data on **Status =1** as active. All other inactive records in

the data will be displayed in blue color. A partial data is shown in the following figure.

| Drug | Time | Status |
|---------|------|--------|
| Placebo | 1 | 1 |
| 6-MP | 10 | 1 |
| Placebo | 22 | 1 |
| 6-MP | 7 | 1 |
| Placebo | 3 | 1 |
| 6-MP | 32 | 0 |
| Placebo | 12 | 1 |
| 6-MP | 23 | 1 |
| Placebo | 8 | 1 |
| 6-MP | 22 | 1 |
| Placebo | 17 | 1 |
| 6-MP | 6 | 1 |
| Placebo | 2 | 1 |
| 6-MP | 16 | 1 |
| Placebo | 11 | 1 |
| 6-MP | 34 | 0 |

Let us perform the Difference of Means Test on the filtered data. Select Difference of Means Test as:

Analysis > (Continuous) Two Samples > (Parallel Design) Difference of Means

In the ensuing dialog box, select **Drug** as **Population Id**, **Placebo** as **Control** and **Time** as **Response variable**. No need to select any variable as **Frequency Variable**.

71 Tutorial: Analysis

The Input dialog box will look as follows:

Analysis: Continuous Endpoint Two-Sample Test - Parallel Design - Difference of Means

Data Set: leukemia.cyd

Main Advanced

Trial Type: Superiority Test Type: t-test Frequency Variable:

Variance Type: Equal Response Variable: Time

Population Id: Drug

Control: Placebo

Treatment: 6-MP

Click **OK**. You will see the following output.

Analysis: Continuous Response: Difference of Means for Independent Data

Hypothesis

$H_0 : \mu_t - \mu_c = 0$ Vs. $H_1 : \mu_t - \mu_c \neq 0$ for 2-Sided test

Either $H_1 : \mu_t - \mu_c < 0$

Or $H_1 : \mu_t - \mu_c > 0$ for 1-Sided test

Input Parameters

Data File: leukemia.cyd
 Trial Type: Superiority
 Drug(Treatment=6-MP,
 Population Id: MP,
 Control=Placebo)
 Response Variable: Time
 Variance Type: Equal
 Test Type: t-test
 Confidence Level: 0.95

Response Variable: Time
 Total Number of Records: 42
 Number of Records Rejected: 12

Summary of the Observed Data:

| Treatment Group | Min | Max | Median | Mean | Std.Dev | n |
|-----------------|-----|-----|--------|--------|---------|----|
| Placebo | 1 | 23 | 8 | 8.667 | 6.468 | 21 |
| 6-MP | 6 | 23 | 10 | 12.111 | 6.846 | 9 |

Test of Hypothesis:

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | 1-Sided | | 2-Sided | 95% Confidence Interval(2-Sided) | |
|----|---------------------|------------------|------------|-------|----|---------|------|---------|----------------------------------|-------------|
| | | | | | | p-value | Tail | p-value | Lower Limit | Upper Limit |
| 30 | 3.444 | 0.524 | 2.621 | 1.314 | 28 | 0.1 | G.E. | 0.199 | -1.924 | 8.813 |

The output is displayed in three parts. The **first part** specifies the **Null** as well as **Alternative hypotheses** to be tested. **East** considers testing of both one sided and two sided alternative hypotheses and computes the corresponding p- values.

In the **second part**, Input details such as **data file name**, **Population Id** and the **test to be performed** are mentioned.

In the **last part**, the output contains number of records in the data, number of record rejected, summary of data and the Inference. From the above output, it is clear that there are in all 42 records, out of which 12 are rejected by filtering and the t test is applied on the remaining 30 records. Both one sided and two sided p-values indicate that the null hypothesis Ho cannot be rejected at 5% level of significance.

71.3 Sub-Group Analysis using By variable

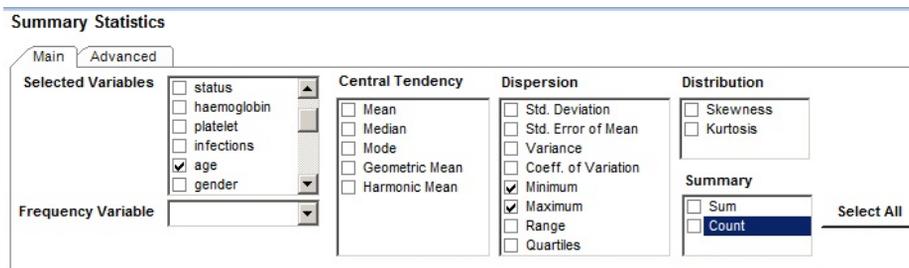
In almost every data analysis feature, **East** provides facility for doing sub-group analysis on maximum of two variables. For instance, consider running 'One way analysis of variance (ANOVA)' procedure on the **Myeloma.cyd** data set available in the **Samples** folder of the installation directory of the product. The authors Krall, Uthoff and Harley (1975) provide data of a survival study that include the survival times, in months, of 65 multiple myeloma patients with data on 15 concomitant variables. In this example, we like to perform one way ANOVA on the variable 'survmth' in subgroups of Males and Females separately.

Suppose we want to see if average survival is different across age groups, we would first like to categorize variable age into a factor. This we can do using the **RCODE** function available in **Transform Variable** in **Case Data Editor**. Open **Myeloma.cyd** from the **Samples** folder of **East**. In order to know the Minimum and Maximum value of age, choose from the menu:

Analysis > (Basic Statistics) Descriptive Statistics > Summary Statistics

71 Tutorial: Analysis

Select the variable **age** and **Minimum** and **Maximum** as shown in the following dialog box.



Click **OK**. You will see the following output.

Analysis: Summary Statistics

Input Parameters

Data File: Myeloma.cyd
Selected Variable (s): age
Statistics: Minimum, Maximum

Output

| | Min | Max |
|-----|-----|-----|
| age | 27 | 82 |

Accordingly let us form the age groups as follows:

1. AgeCode=1 for Age from 21 to 40
2. AgeCode=2 for Age from 41 to 60
3. AgeCode=3 for Age from 61 to 80
4. AgeCode=4 for Age not less than 81

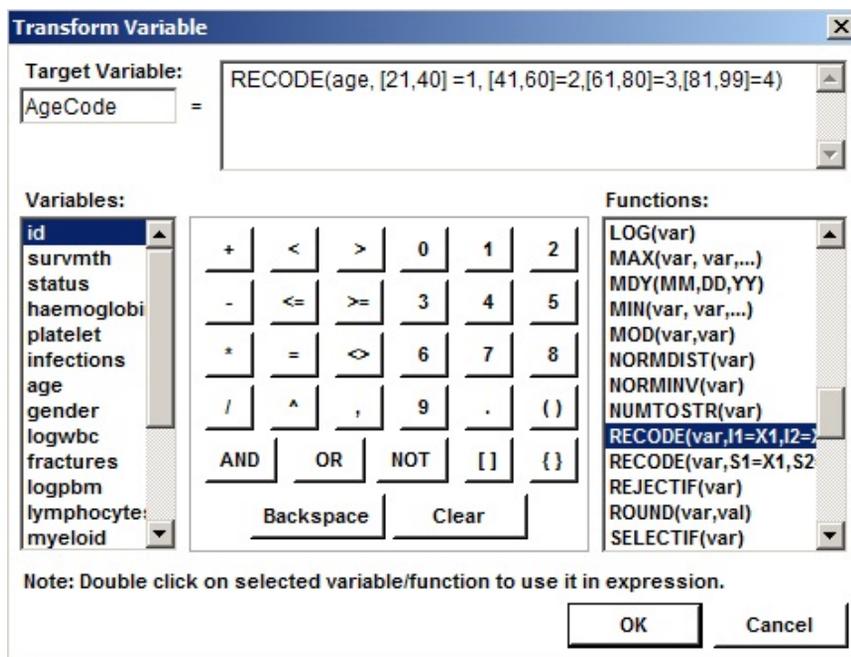
This grouping can be done using **RCODE** function. Here we show how to accomplish this. For demonstrating the grouping capability, we would use **Myeloma** dataset. This dataset when opened looks like this:

| | id | survmth | status | haemoglobin | platelet | infections |
|----|----|---------|--------|-------------|----------|------------|
| 2 | 2 | 1.25 | 1 | 12 | 1 | 1 |
| 3 | 3 | 2 | 1 | 9.8 | 1 | 1 |
| 4 | 4 | 2 | 1 | 11.3 | 0 | 0 |
| 5 | 5 | 2 | 1 | 5.1 | 0 | 0 |
| 6 | 6 | 3 | 1 | 6.7 | 1 | 1 |
| 7 | 7 | 5 | 1 | 10.1 | 1 | 0 |
| 8 | 8 | 5 | 1 | 6.5 | 1 | 0 |
| 9 | 9 | 6 | 1 | 9 | 1 | 1 |
| 10 | 10 | 6 | 1 | 10.2 | 0 | 0 |
| 11 | 11 | 6 | 1 | 9.7 | 1 | 0 |
| 12 | 12 | 6 | 1 | 10.4 | 1 | 0 |
| 13 | 13 | 7 | 1 | 9.5 | 1 | 0 |
| 14 | 14 | 7 | 1 | 5.1 | 0 | 0 |
| 15 | 15 | 7 | 1 | 11.4 | 1 | 0 |
| 16 | 16 | 9 | 1 | 8.2 | 1 | 0 |
| 17 | 17 | 11 | 1 | 14 | 1 | 0 |
| 18 | 18 | 11 | 1 | 12 | 1 | 0 |
| 19 | 19 | 11 | 1 | 13.2 | 1 | 0 |
| 20 | 20 | 11 | 1 | 7.5 | 1 | 0 |
| 21 | 21 | 11 | 1 | 9.6 | 1 | 0 |

In the next column after sercalcium (which is the last variable in the dataset) construct a new variable by clicking the Transform icon  on the **Data Editor menu**. In the

71 Tutorial: Analysis

ensuing dialog box, type the **Transform command** as shown below.



Click **OK**. This will generate a new variable **AgeCode** with values from 1 to 4. To run **One way ANOVA** procedure, follow the steps:

Analysis > (Continuous) Many Samples > One way ANOVA

You will see the following input dialog box. In it select **AgeCode** as **Factor** and **survmth** as **Response**.

Data Set: Myeloma.cyd

The screenshot shows a software interface with two tabs: 'Main' and 'Advanced'. The 'Main' tab is selected. Below the tabs, there are two dropdown menus. The first is labeled 'Factor:' and contains the text 'age'. The second is labeled 'Response:' and contains the text 'survmth'. To the right of these dropdowns is a checkbox labeled 'Contrast' which is currently unchecked.

Click **Advanced Tab**. Select **gender** in **By variable 1** drop down box.

The screenshot shows the same software interface but with the 'Advanced' tab selected. The 'By Variable 1:' dropdown menu now contains the text 'gender'. The 'By Variable 2:' dropdown menu is currently empty.

Click **OK**. The output obtained is as shown below.

71 Tutorial: Analysis

Analysis: ANOVA: One Way

Hypothesis

$H_0 : \mu_1 = \mu_2 = \dots = \mu_r$

Vs.

$H_1 : \text{At least one } \mu_i \text{ differs, for } i=1,2,3 \dots r \text{ where } r \text{ is the number of treatment groups}$

Input Parameters

Data File: Myeloma.cyd

Factor: AgeCode

Response: survmth

By Variable (s): gender

By: gender =
1

Total Number of Records: 38

Number of Records Rejected: 0

ANOVA Table:

| Source | DF | Sum of Squares | Mean Square | F Statistic | p-value |
|-----------|----|----------------|-------------|-------------|---------|
| AgeCode | 3 | 1515.209 | 505.07 | 0.878 | 0.462 |
| Residuals | 34 | 19568.593 | 575.547 | | |
| Total | 37 | 21083.803 | | | |

Summary Table:

| Overall Mean | Root MSE | R-Square | Adjusted R-Square | Coeff. of Variation |
|--------------|----------|----------|-------------------|---------------------|
| 24.908 | 23.991 | 0.072 | -0.01 | 96.317 |

By: gender =
2

Total Number of Records: 27

Number of Records Rejected: 0

ANOVA Table:

| Source | DF | Sum of Squares | Mean Square | F Statistic | p-value |
|-----------|----|----------------|-------------|-------------|---------|
| AgeCode | 3 | 1101.185 | 367.062 | 0.617 | 0.611 |
| Residuals | 23 | 13672 | 594.435 | | |
| Total | 26 | 14773.185 | | | |

Summary Table:

| Overall Mean | Root MSE | R-Square | Adjusted R-Square | Coeff. of Variation |
|--------------|----------|----------|-------------------|---------------------|
| 22.741 | 24.381 | 0.075 | -0.046 | 107.213 |

Notice that **East** has performed one way ANOVA procedure on the two subgroups formed by **gender=1** and **gender=2** respectively. There is no significant effect of **Age** on the survival. This is true in case of both male and female patients.

71.4 Workflow for Analysis

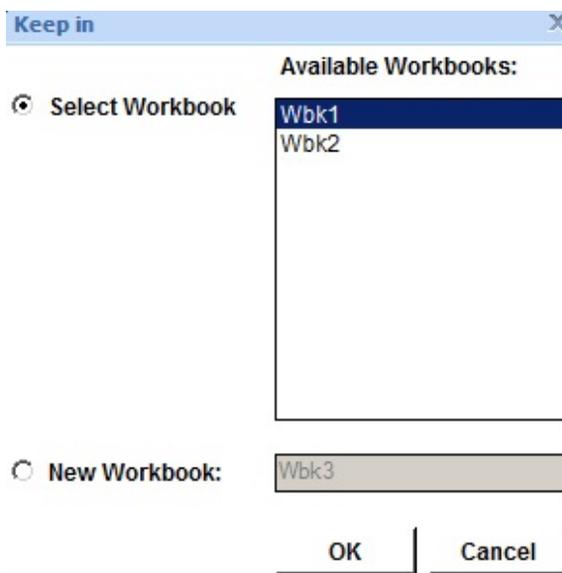
In this section we will walk you through the steps that will be generally followed while performing any analysis in **East**.

71.4.1 Getting Data into East

Data may be entered into **East** as case data or as crossover data, read in as a previously saved East file (.cydx) through the **Open** command in the **File** menu, or read in from another software package through the **Import command**. In this tutorial, you will read in a previously saved data file using the **Open** command.

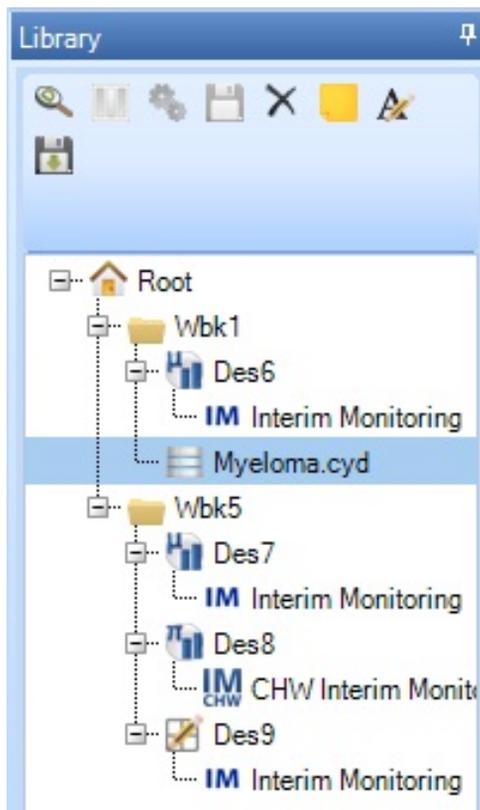
For illustrative purposes, let us consider performing **Difference of Means Test** on the data **Myeloma.cyd** available in the **Samples** folder.

Open the data set **Myeloma.cyd** from **Samples** folder. If there are several workbooks in Library, **East** will ask for the workbook you would like to store the data as shown in the following dialog:



71 Tutorial: Analysis

Suppose you choose **Wbk1**. A node named **Myeloma.cydx** will be created in **Wbk1** as shown below:



You may rename the data node by right clicking on it.

71.4.2 Choose the Test

Choose the test from the appropriate submenu of Analysis. In this case, select;

Analysis > (Continuous) Two Samples > (Parallel Designs) Difference of Means.

In the ensuing dialog, select the variables as shown below.

Data Set: Myeloma.cyd

Main Advanced

Trial Type: Superiority Test Type: t-test Frequency Variable:

Variance Type: Equal

Population Id: status Response Variable: platelet

Control: 0

Treatment: 1

Click **OK** to execute the test. You will see the following output.

71.4.3 Output

Analysis: Continuous Response: Difference of Means for Independent Data

Hypothesis

$H_0 : \mu_t - \mu_c = 0$ Vs. $H_1 : \mu_t - \mu_c \neq 0$ for 2-Sided test

Either $H_1 : \mu_t - \mu_c < 0$

Or $H_1 : \mu_t - \mu_c > 0$ for 1-Sided test

Input Parameters

Data File: Myeloma.cyd
 Trial Type: Superiority
 Population Id: status(Treatment=1, Control=0)
 Response Variable: platelet
 Variance Type: Equal
 Test Type: t-test
 Confidence Level: 0.95

Response Variable: platelet

Total Number of Records: 65

Number of Records Rejected: 0

Summary of the Observed Data:

| Treatment Group | Min | Max | Median | Mean | Std.Dev | n |
|-----------------|-----|-----|--------|-------|---------|----|
| 0 | 0 | 1 | 1 | 0.941 | 0.243 | 17 |
| 1 | 0 | 1 | 1 | 0.833 | 0.377 | 48 |

Test of Hypothesis:

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | 1-Sided | | 2-Sided | | 95% Confidence Interval(2-Sided) | |
|----|---------------------|------------------|------------|------|----|---------|------|---------|-------------|----------------------------------|--|
| | | | | | | p-value | Tail | p-value | Lower Limit | Upper Limit | |
| 65 | -0.108 | -0.31 | 0.098 | -1.1 | 63 | 0.138 | L.E. | 0.276 | -0.304 | 0.088 | |

The output is divided into three sections. The first one is the **Hypothesis** where the null and alternative hypothesis for 2 sided and 1-sided tests are stated.

The next section is the **Input Parameters** section. This section tells us the name of

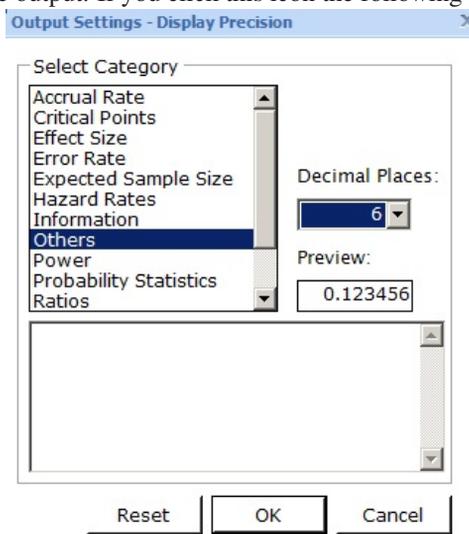
71 Tutorial: Analysis

data file and response variable used in the analysis, type of test performed, confidence level set for the analysis and other parameter(s) used in the analysis. This section is very important to review to make sure that we specified all the input correctly.

The last section is **Output**. First part of the output is the **Summary of the observed Data**. It contains descriptive statistics such as minimum, maximum, mean, median and standard deviation of the response variable within the two treatments groups. The remaining part of the output contains **inference for t test**. The standardized effect size is **-0.31** with t statistic value as **-1.1** which with **63 d.f.** is non-significant. Accordingly, the test fails to reject the null hypothesis at 5% level of significance. This is substantiated by observing that the 95 % confidence interval includes value 0.

You will see three icons at the top of the Analysis output. Using  icon you can print the output. The  icon is used to save the output as HTML.

With  icon you can readily change the display settings, in particular the number of decimal points on the output. If you click this icon the following dialog comes up.



Change the display precision for '**Others**' category to 6 decimals as shown in the above dialog box and click **OK**. You will see the following output with, other than

Beta and p- values, all values are displayed up to 6 decimals.

Analysis: Continuous Response: Difference of Means for Independent Data

Hypothesis

$H_0 : \mu_t - \mu_c = 0$ Vs. $H_1 : \mu_t - \mu_c \neq 0$ for 2-Sided test
 Either $H_1 : \mu_t - \mu_c < 0$
 Or $H_1 : \mu_t - \mu_c > 0$ for 1-Sided test

Input Parameters

Data File: Myeloma.cyd
 Trial Type: Superiority
 Population Id: status(Treatment=1, Control=0)
 Response Variable: platelet
 Variance Type: Equal
 Test Type: t-test
 Confidence Level: 0.95

Response Variable: platelet
 Total Number of Records: 65
 Number of Records Rejected: 0
 Summary of the Observed Data:

| Treatment Group | Min | Max | Median | Mean | Std.Dev | n |
|-----------------|-----|-----|--------|-------|---------|----|
| 0 | 0 | 1 | 1 | 0.941 | 0.243 | 17 |
| 1 | 0 | 1 | 1 | 0.833 | 0.377 | 48 |

Test of Hypothesis:

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | 1-Sided | | 2-Sided | 95% Confidence Interval(2-Sided) | |
|----|---------------------|------------------|------------|-----------|----|---------|------|---------|----------------------------------|-------------|
| | | | | | | p-value | Tail | p-value | Lower Limit | Upper Limit |
| 65 | -0.108 | -0.31 | 0.098 | -1.099564 | 63 | 0.138 | L.E. | 0.276 | -0.303837 | 0.08815 |

71.4.4 Links to SAS

The Analysis module of **East** facilitates analysis using SAS Procedures on the data in two ways.

1. Invoking **SAS** through SAS link '**Run Using SAS**' provided on the **Advanced tab** for the tests **Linear Mixed Effects Model: Difference of Means and Ratio of Means**. These tests are part of the **Regression** menu from **Analysis: Continuous**. By doing this, **East** will invoke Mixed procedure of **SAS**. You can also choose not to use **SAS**. If you use **SAS**, you will have the option of including covariates in your model. Without **SAS**, your model will not include covariates.
2. Using **SAS** command option available in the tests for 2x2 crossover tests in the **Regression** menu **Analysis: Continuous**. With this option, you can use your own data set and **SAS** commands. This will utilize the data in **East** and run the **PROC** specified in your **SAS** code. The output will be displayed in the **East's** main window. The flexibility offered allows you to write any **SAS** code. The only exception is that your code should not contain **SAS** graphics.

71 *Tutorial: Analysis*

For more details of these aspects you are referred to the respective chapters.

72 *Analysis-Descriptive Statistics*

Descriptive Statistics, under the **Basic Statistics** menu, deals with preliminary exploration of data using elementary tools such as computation of summary measures, classification, cross tabulation of the data. Descriptive Statistics helps statisticians to choose statistical analysis techniques to arrive at meaningful inference.

In this Chapter, Section (72.1) describes the descriptive statistical measures available in **East**. Section (72.2) describes the procedure for obtaining frequency distribution for one or more variables in a data set. Section (72.3) details the procedure for obtaining a cross-tabulation of any two variables in a case data file. All these procedures are available only for case data.

Note: All measures are computed after dropping observations with missing values.

72.1 *Summary Statistics*

72.1.1 *Example: Summary Statistics*

East provides results for a set of 16 predefined univariate summary measures for numeric variables in a data set. These measures help you to select the type of analysis to carry out later.

The following Descriptive Statistics or Univariate Summary Measures are available.

| Central Tendency | Dispersion | Distribution | Summary |
|---------------------------|--------------------------|--------------|---------|
| Mean (Std. Error of Mean) | Standard Deviation | Skewness | Count |
| Median | Variance | Kurtosis | Sum |
| Mode | Coefficient of variation | | |
| Geometric Mean | Maximum | | |
| Harmonic Mean | Minimum | | |
| | Range | | |

72.1.1 *Example: Summary Statistics*

Dataset: Myeloma.cydx

Data Description:

The authors Krall, Uthoff and Harley (1975) have provided data of a survival study. It included the survival times in months of 65 multiple myeloma patients with data on 15 concomitant variables.

72 Analysis-Descriptive Statistics

Purpose of the Analysis:

To compute summary measures for the variables **survmth**, **haemoglobin** and **bjprotein** grouped based on the survival status **status** and gender **gender**.

Analysis Steps:

1. Open the dataset from Samples folder.
2. Choose the menu item:

Analysis > (Basic Statistics) Descriptive Statistics > Summary Statistics

3. In the **Main** tab, select the variables of interest. In this example select the following variables: **survmth**, **haemoglobin** and **bjprotein**. Click the **Select All** button to get the results for all the summary measures.

4. Thereafter, under the **Advanced** tab choose the variables as shown below:

5. Click **OK**. You will see the Analysis results as shown below.

| | | Mean | Std.Error | Median | Mode | GeoMean |
|--------------------|--------|--------|-----------|--------|------|---------|
| survmth | | | | | | |
| status | gender | | | | | |
| 0 | 1 | 33.25 | 9.697 | 28.5 | N.A. | 21.578 |
| 0 | 2 | 13.556 | 2.467 | 12 | 12 | 11.741 |
| 1 | 1 | 22.683 | 4.168 | 14.5 | 11 | 12.829 |
| 1 | 2 | 27.333 | 6.567 | 17 | 6 | 16.921 |
| haemoglobin | | | | | | |
| status | gender | | | | | |
| 0 | 1 | 12.338 | 0.525 | 12.65 | 10.2 | 12.256 |
| 0 | 2 | 9.856 | 0.854 | 10 | N.A. | 9.498 |
| 1 | 1 | 10.277 | 0.458 | 10.1 | 9 | 9.961 |
| 1 | 2 | 9.3 | 0.616 | 10.1 | 7.5 | 8.917 |
| bjprotein | | | | | | |
| status | gender | | | | | |
| 0 | 1 | 1.625 | 0.183 | 2 | 2 | 1.542 |
| 0 | 2 | 2 | 0 | 2 | 2 | 2 |
| 1 | 1 | 1.6 | 0.091 | 2 | 2 | 1.516 |
| 1 | 2 | 1.556 | 0.121 | 2 | 2 | 1.47 |

72.2 Example: Frequency Distribution

The procedure **Frequency Distribution** displays a separate frequency distribution table for each of the variables specified in a list. The default display includes the values of the variable in sorted order in the first column and the frequencies in the second column. Additional display can be obtained by choosing one or more of the options Percentage, Cumulative<=, Cumulative >= or Compute Percentiles.

Dataset: *Myeloma.cydx* as described in Section 72.1.1.

Purpose of the Analysis:

To obtain Frequency Distribution table for the variables **haemoglobin**, **age**, **fractures**, and **bjprotein**.

Analysis Steps:

1. Open the dataset from Samples folder.
2. Choose the menu item:
Analysis > (Basic Statistics) Descriptive Statistics > Frequency Distribution

72 Analysis-Descriptive Statistics

- In the ensuing dialog box (under the **Main** tab), select the variables of interest in the **Selected Variables** box. For this example, select the variables: **haemoglobin**, **age**, **fractures**, and **bjprotein**. In the **Frequency Output** select all the three checkboxes and select the **Compute Percentile** check box.

Frequency Distribution

Main | Advanced

Selected Variables

- logpbm
- lymphocytes
- myeloid
- uria
- bjprotein
- totserprotein

Frequency Variable

Frequency Output

- Percentage
- Cumulative <=
- Cumulative >=

Compute Percentile

- Thereafter, under the **Advanced** tab choose the By Variables as shown below:

Frequency Distribution Data :

Main | Advanced

By Variable

Percentile Values

- Quartiles
- Percentile(s)
- Cut Points for Equal groups

| |
|----|
| 33 |
| 67 |

5. Click **OK**. A partial output of the Analysis results is as shown below.

Analysis: Frequency Distribution

Input Parameters

Data File: Myeloma.cyd
 Selected Variable (s): haemoglobin, age, fractures, bprotein
 By Variable (s): gender

Output

By Variable (s): gender(=1)
 Selected Variable (s): haemoglobin
 No. of obs. Analysed: 65
 No. of obs. Rejected: 0

| Value | Freq | % | Cum.<= Freq | Cum.<= % | Cum.>= Freq | Cum.>=% |
|-------|------|-------|-------------|----------|-------------|---------|
| 5.1 | 1 | 2.632 | 1 | 2.632 | 38 | 100 |
| 6.5 | 1 | 2.632 | 2 | 5.263 | 37 | 97.368 |
| 6.6 | 1 | 2.632 | 3 | 7.895 | 36 | 94.737 |
| 7 | 1 | 2.632 | 4 | 10.526 | 35 | 92.105 |
| 7.5 | 1 | 2.632 | 5 | 13.158 | 34 | 89.474 |
| 7.7 | 1 | 2.632 | 6 | 15.789 | 33 | 86.842 |
| 8.2 | 1 | 2.632 | 7 | 18.421 | 32 | 84.211 |
| 9 | 3 | 7.895 | 10 | 26.316 | 31 | 81.579 |
| 9.4 | 1 | 2.632 | 11 | 28.947 | 28 | 73.684 |
| 9.5 | 1 | 2.632 | 12 | 31.579 | 27 | 71.053 |
| 9.7 | 1 | 2.632 | 13 | 34.211 | 26 | 68.421 |
| 9.8 | 1 | 2.632 | 14 | 36.842 | 25 | 65.789 |
| 10 | 1 | 2.632 | 15 | 39.474 | 24 | 63.158 |
| 10.2 | 3 | 7.895 | 18 | 47.368 | 23 | 60.526 |
| 10.6 | 2 | 5.263 | 20 | 52.632 | 20 | 52.632 |
| 11 | 1 | 2.632 | 21 | 55.263 | 18 | 47.368 |

72 Analysis-Descriptive Statistics

| | | | | | | |
|--------------|----|--------|----|--------|----|--------|
| 11.2 | 1 | 2.632 | 22 | 57.895 | 17 | 44.737 |
| 11.3 | 1 | 2.632 | 23 | 60.526 | 16 | 42.105 |
| 12 | 3 | 7.895 | 26 | 68.421 | 15 | 39.474 |
| 12.4 | 1 | 2.632 | 27 | 71.053 | 12 | 31.579 |
| 12.5 | 1 | 2.632 | 28 | 73.684 | 11 | 28.947 |
| 12.8 | 2 | 5.263 | 30 | 78.947 | 10 | 26.316 |
| 13 | 1 | 2.632 | 31 | 81.579 | 8 | 21.053 |
| 13.2 | 1 | 2.632 | 32 | 84.211 | 7 | 18.421 |
| 14 | 4 | 10.526 | 36 | 94.737 | 6 | 15.789 |
| 14.4 | 1 | 2.632 | 37 | 97.368 | 2 | 5.263 |
| 14.6 | 1 | 2.632 | 38 | 100 | 1 | 2.632 |
| Total | 38 | 100 | | | | |

72.3 Example: Tabulate

The **Tabulate** procedure allows cross-tabulation of any two specified variables of a data set. It also allows specification of a **Frequency variable** and a **By Variable** for subgroup analysis. The optional output from this procedure includes row, column, and overall percentages, expected values and chi-square statistics with p-values.

Dataset: Job-case.cyx

Data Description:

This example refers to the data obtained in a general social survey conducted by National Opinion Research Center (1991) among black American women and men. The data is in case data form. For exploratory analysis and presentation, it is useful to summarize these data into a tabular form.

The data consist of the following annual income levels:

- <\$5,000
- 5,000-15,000
- 15,000-25,000
- > 25,000

and job satisfaction levels:

- Very Dissatisfied
- A little satisfied
- Moderately satisfied
- Very satisfied

for 64 women and 40 men.

Purpose of the Analysis:

To cross tabulate the data for the variables **Incomegrp** and **Jobsatis** grouped by gender.

Analysis Steps:

1. Open the dataset from Samples folder.
2. Choose the menu item:

Analysis > (Basic Statistics) Descriptive Statistics > Tabulate

3. In the ensuing dialog box, choose the variables as shown below.

Tabulate

Main | Advanced

Row Variable: Incomegrp Column Variable: Jobsatis Display: Chi-Square Test

Frequency Variable: Persons

Row Percent
 Column Percent
 Overall Percent
 Expected Value

4. Thereafter, under the **Advanced** tab choose the variables as shown below.

Tabulate

Main | Advanced

By Variable: Gender

5. Click **OK**. You will see the Analysis results as shown below.

72 Analysis-Descriptive Statistics

Analysis: Tabulate

Input Parameters

Data File: Job-case.cyd
 Row Variable: Incomegrp
 Column Variable: Jobsatis
 Frequency Variable: Persons
 By Variable (s): Gender

Output

Gender 1

| Frequency | 1 | 2 | 3 | 4 | Total |
|-----------------|--------|--------|--------|--------|-------|
| Expected Values | | | | | |
| 1 | 1 | 3 | 11 | 2 | |
| | 1.563 | 4.688 | 17.188 | 3.125 | |
| | 5.882 | 17.647 | 64.706 | 11.765 | |
| | 33.333 | 33.333 | 27.5 | 16.667 | |
| | 0.797 | 2.391 | 10.625 | 3.188 | |
| 2 | 2 | 3 | 17 | 3 | |
| | 3.125 | 4.688 | 26.563 | 4.688 | |
| | 8 | 12 | 68 | 12 | |
| | 66.667 | 33.333 | 42.5 | 25 | |
| | 1.172 | 3.516 | 15.625 | 4.688 | |
| 3 | 0 | 1 | 8 | 5 | |
| | 0 | 1.563 | 12.5 | 7.813 | |
| | 0 | 7.143 | 57.143 | 35.714 | |
| | 0 | 11.111 | 20 | 41.667 | |
| | 0.656 | 1.969 | 8.75 | 2.625 | |

| | | | | | |
|--------------|--------------|---------------|-------------|--------------|------------|
| | 0 | 1 | 8 | 5 | |
| | 0 | 1.563 | 12.5 | 7.813 | |
| 3 | 0 | 7.143 | 57.143 | 35.714 | |
| | 0 | 11.111 | 20 | 41.667 | |
| | 0.656 | 1.969 | 8.75 | 2.625 | |
| | 0 | 2 | 4 | 2 | |
| | 0 | 3.125 | 6.25 | 3.125 | |
| 4 | 0 | 25 | 50 | 25 | |
| | 0 | 22.222 | 10 | 16.667 | |
| | 0.375 | 1.125 | 5 | 1.5 | |
| Total | 4.688 | 14.063 | 62.5 | 18.75 | 100 |

| Statistic | Value | DF | P-Value |
|------------|-------|----|---------|
| Chi_square | 6.82 | 9 | 0.656 |

72 Analysis-Descriptive Statistics

Gender 2

| Frequency | 1 | 2 | 3 | 4 | Total |
|-----------------|--------------------------------|-----------------------------------|--|--------------------------------------|-------|
| Expected Values | | | | | |
| 1 | 1 2.5 20 100 0.125 | 1 2.5 20 20 0.625 | 2 5 40 8.696 2.875 | 1 2.5 20 9.091 1.375 | |
| 2 | 0 0 0 0 0.225 | 3 7.5 33.333 60 1.125 | 5 12.5 55.556 21.739 5.175 | 1 2.5 11.111 9.091 2.475 | |
| 3 | 0 0 0 0 0.25 | 0 0 0 0 1.25 | 7 17.5 70 30.435 5.75 | 3 7.5 30 27.273 2.75 | |
| 4 | 0 0 0 0 0.4 | 1 2.5 6.25 20 2 | 9 22.5 56.25 39.13 9.2 | 6 15 37.5 54.545 4.4 | |
| Total | 2.5 | 12.5 | 57.5 | 27.5 | 100 |

| Statistic | Value | DF | P-Value |
|------------|--------|----|---------|
| Chi_square | 14.234 | 9 | 0.114 |

73 *Analysis-Analytcs*

This chapter describes some of the commonly used univariate procedures: t-test (paired and independent), one-way and two-way (without interaction) analysis of variance (ANOVA) and multiple linear regression. The topic of correlations is also included in this chapter. References for the procedures covered in this chapter are provided in the table shown below:

| Test | References |
|--|---|
| t-tests | Snedecor & Cochran (1989) |
| ANOVA | Kreyszig (1970) |
| Pearson’s Product-Moment Correlation Spearman’s Product-Moment Correlation Kendall’s Tau | Siegel & Castellan (1988) |
| Regression procedures Collinearity diagnostics Residuals and Influence | Maindonald J (1984) Belsley, Kuh, & Welsh (1980) Cook & Weisberg (1982) |

Note: Any observation with a missing value for a variable that is included in the model is excluded from the analysis.

73.1 Example: t-test
 73.1.1 Independent t-test
 73.1.2 Paired t-test

This section describes t-test procedures for analyzing data of **independent** and **paired** samples.

73.1.1 Independent t-test

Dataset: *Myeloma.cyx* as described in Section 72.1.1.

Purpose of the Analysis:

To compare mean **Uria** level between two groups indicated by the variable **status** (0-alive, 1-dead).

Analysis Steps:

1. Open the dataset from Samples folder.
2. Choose the menu item:

Analysis > (Basic Statistics) Analytcs > (t-test) Independent t-test

73 Analysis-Analytics

3. In the ensuing dialog box, choose the variables as shown below.

The screenshot shows the 'Independent t-Test' dialog box with the following settings:

- Tab: Main
- Population: status
- Response: uria
- Equal Variance:
- Unequal Variance:

4. Click **OK**. The results will appear as shown below.

Analysis: Two Sample t-test (Equal Variance)

Input Parameters

Data File: Myeloma.cyd

Population: status

Response: uria

| N | Difference | Standard Error | 95% C.I.(Lower Limit) | 95% C.I.(Upper Limit) | t | DF | P-Value |
|----|------------|----------------|-------------------------|-------------------------|--------|----|---------|
| 65 | -3.521 | 1.654 | -6.826 | -0.215 | -2.129 | 63 | 0.037 |

You can try running the t-test with unequal variance by selecting the **Unequal Variance** option on the main tab.

73.1.2 Paired t-test

Dataset: Azt1.cyd

Data Description:

The data from Makuch and Parks (1988) documents the response of serum antigen level to AZT in 20 AIDS patients. Two sets of antigen levels are provided for each patient:

- Pre-treatment
- Post-treatment

Purpose of the Analysis:

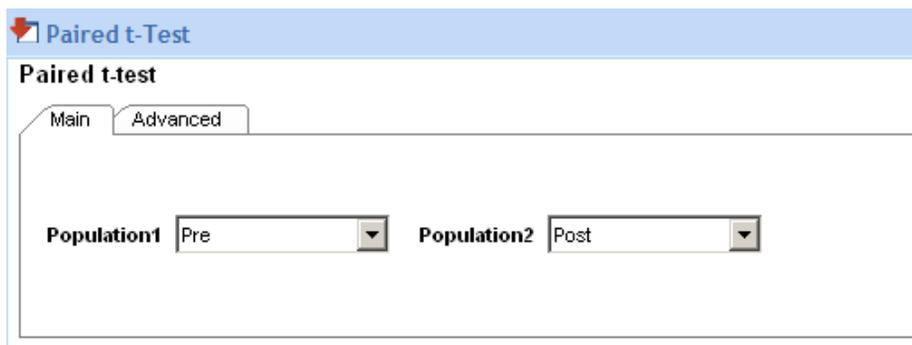
To compare the mean antigen level among patients after administering the treatment with the mean antigen level before administering the treatment.

Analysis Steps:

1. Open the dataset from Samples folder.
2. Choose the menu item:

Analysis > (Basic Statistics) Analytics >(t-test) Paired t-test

3. In the ensuing dialog box, choose the variables as shown below.



4. Click **OK**. You will see the analysis results as shown below.

Analysis: Paired t test

Input Parameters

Data File: Azt1.cyd
 Population1: Pre
 Population2: Post

| N | Difference | Standard Error | 95% C.I.(Lower Limit) | 95% C.I.(Upper Limit) | t | DF | P-Value |
|----|------------|----------------|-------------------------|-------------------------|-----|----|---------|
| 20 | 232.7 | 129.256 | -37.835 | 503.235 | 1.8 | 19 | 0.088 |

73 Analysis-Analytcs

73.2 Analysis of Variance

73.2.1 One-way Analysis of Variance

73.2.2 Two-way Analysis of Variance

The ANOVA procedure available under **Analytcs** menu can perform simple **one-way** and **two-way** analysis of variance, described in this section.

73.2.1 One-way Analysis of Variance

Dataset: Leukocyte.cyd

Data Description:

This data comes from a study done by Kontula et al. (1980) (1982) in which the Glucocorticoid Receptor (GR) Sites per Leukocyte Cell in normal subjects (Group 1) were compared to those in patients with hairy-cell leukemia (Group 2), chronic lymphatic leukemia (Group 3), chronic myelocytic leukemia (Group 4) or acute leukemia (Group 5). One of the aims of the study was to find whether there were any significant differences in the mean number of GR sites per leukocyte cells between these five groups.

Purpose of the Analysis:

To test whether the mean GR sites per Leukocyte Cell is the same across all groups.

Analysis Steps:

1. Open the dataset from Samples folder.
2. Choose the menu item:

Analysis > (Basic Statistics) Analytcs > ANOVA

3. In the ensuing dialog box, choose the variables as shown below.

ANOVA

ANOVA

Value GR Factor1 Group Factor2

4. Click **OK**. You will see the analysis results as shown below.

Analysis: One way ANOVA

Input Parameters

Data File: leukocyte.cyd

Value: GR

Factor1: Group

Output

| Source | Sum of Squares | DF | Mean Square | F | P-Value |
|-----------|----------------|----|--------------|-------|---------|
| Group | 353735187.391 | 4 | 88433796.848 | 4.308 | 0.007 |
| Residuals | 656834693.69 | 32 | 20526084.178 | | |

| | |
|--------------|----------|
| Overall Mean | 7107.568 |
| Root MSE | 4530.572 |
| R-Square | 0.35 |
| Coeff Var | 63.743 |

73.2.2 Two-way Analysis of Variance

Dataset: swine.cyd

Data Description:

The data for this example is a subset of the data that comes from a study reported by Snedecor and Cochran (1989). This data relates to dressing percentages of 20 swine that have been classified by breed (5 categories) and sex (2 categories) with 2 swine under each combination of breed and sex categories.

Purpose of the Analysis:

The aim of this study is to test for the effect of breed and sex on the study measure taken on the animals.

73 Analysis-Analytics

Analysis Steps:

1. Open the dataset from Samples folder.
2. Choose the menu item:

Analysis > (Basic Statistics) Analytics > ANOVA

3. In the ensuing dialog box, choose the variables as shown below.



The screenshot shows a dialog box titled "ANOVA". Inside the dialog, there are three dropdown menus. The first is labeled "Value" and is set to "Measure". The second is labeled "Factor1" and is set to "Breed". The third is labeled "Factor2" and is set to "Sex".

4. Click **OK**. You will see the Analysis results as shown below.

Analysis: Two way ANOVA

Input Parameters

Data File: swine.cyd

Value: Measure

Factor1: Breed

Factor2: Sex

Output

| Source | Sum of Squares | DF | Mean Square | F | P-Value |
|-----------|----------------|----|-------------|-------|---------|
| Breed | 39.932 | 4 | 9.983 | 1.346 | 0.302 |
| Sex | 27.613 | 1 | 27.613 | 3.724 | 0.074 |
| Residuals | 103.805 | 14 | 7.415 | | |

| | |
|-----------------------|--------|
| Max no. of Replicates | 2 |
| Overall Mean | 12.405 |
| Root MSE | 2.723 |
| R-Square | 0.394 |
| Coeff Var | 21.951 |

73.3 Correlations

73.3.1 When to Use Each Measure

73.3.2 Example: Correlations

The **Correlations** procedure under **Analytcs** can be used to compute the following correlation measures for pairs of variables in a data set:

- Pearson's Correlation
- Spearman's Rho
- Kendall's Tau

All these measures of correlation range between -1 and +1 with:

- 0 signifying no association
- -1 signifying perfect negative association
- +1 signifying perfect positive association.

73.3.1 When to Use Each Measure

All the measures of correlation or association in this section capture in a single number

73 Analysis-Analytics

the relationship between two ordered data series. But one measure might be more appropriate than the others under different assumptions about the data. Here are some guidelines on when to use each measure.

Pearson: Use the Pearson product-moment correlation coefficient when you can assume that two correlated data series follow a bivariate normal distribution.

Spearman: Use the Spearman rank-order correlation coefficient when you cannot make a normality assumption about the two data series.

Kendall's Tau: Use Kendall's Tau to capture the association between two data series that are ordered implicitly but not numerically.

73.3.2 Example: Correlations

Dataset: **Myeloma.cyd** as described in Section 72.1.1.

Purpose of the Analysis:

To compute correlations among all pairs of variables from the variables **age**, **bjprotein**, **haemoglobin**, and **survmth**.

Analysis Steps:

1. Open the dataset from Samples folder.
2. Choose the menu item:

Analysis > (Basic Statistics) Analytics > Correlation

3. In the ensuing dialog box, select **age**, **bjprotein**, **haemoglobin**, and **survmth** as the **Selected Variables** and select all the three checkboxes inside the

Correlation box.

Correlations

Correlations

Selected Variables

id
 survmth
 status
 haemoglobin

Frequency Variable

[Empty]

Correlation

Pearson's Product Moment
 Spearman's Rho
 Kendall's Tau

Analysis: Correlation

Input Parameters
 Data File: Mydata.csf
 Selected Variable(s): survmth haemoglobin age bprotein
 Total Observations: 65
 Observations Rejected: 0

Output

Sample Size

| | survmth | haemoglobin | age | bprotein |
|-------------|---------|-------------|-----|----------|
| survmth | 65 | 65 | 65 | 65 |
| haemoglobin | 65 | 65 | 65 | 65 |
| age | 65 | 65 | 65 | 65 |
| bprotein | 65 | 65 | 65 | 65 |

Pearson's Product Moment

| | survmth | haemoglobin | age | bprotein |
|-------------|---------|-------------|--------|----------|
| survmth | 1 | 0.195 | 0.041 | -0.243 |
| haemoglobin | 0.195 | 1 | -0.103 | -0.147 |
| age | -0.041 | -0.103 | 1 | 0.175 |
| bprotein | -0.243 | -0.147 | 0.175 | 1 |

Spearman's Rho Correlation

| | survmth | haemoglobin | age | bprotein |
|-------------|---------|-------------|--------|----------|
| survmth | 1 | 0.202 | -0.033 | -0.23 |
| haemoglobin | 0.202 | 1 | -0.141 | -0.16 |
| age | -0.033 | -0.141 | 1 | 0.188 |
| bprotein | -0.23 | -0.16 | 0.188 | 1 |

73 Analysis-Analytics

4. Click **OK**. You will see the Analysis results as shown below.

Analysis: Correlation

Input Parameters

Data File: Myeloma.cyd
 Selected Variable (s): survmth,haemoglobin,age,bjprotein
 Total Observations: 65
 Observations Rejected: 0

Output

Sample Size

| | survmth | haemoglobin | age | bjprotein |
|-------------|---------|-------------|-----|-----------|
| survmth | - | 65 | 65 | 65 |
| haemoglobin | 65 | - | 65 | 65 |
| age | 65 | 65 | - | 65 |
| bjprotein | 65 | 65 | 65 | - |

Pearson's Product Moment

| | survmth | haemoglobin | age | bjprotein |
|-------------|---------|-------------|--------|-----------|
| survmth | 1 | 0.195 | -0.041 | -0.349 |
| haemoglobin | 0.195 | 1 | -0.159 | -0.147 |
| age | -0.041 | -0.159 | 1 | 0.175 |
| bjprotein | -0.349 | -0.147 | 0.175 | 1 |

Spearman's Rho Correlation

| | survmth | haemoglobin | age | bjprotein |
|-------------|---------|-------------|--------|-----------|
| survmth | 1 | 0.232 | -0.033 | -0.23 |
| haemoglobin | 0.232 | 1 | -0.149 | -0.16 |
| age | -0.033 | -0.149 | 1 | 0.198 |
| bjprotein | -0.23 | -0.16 | 0.198 | 1 |

73.4 Multiple Linear Regression

73.4.1 Available procedures

73.4.2 Example: Multiple Linear Regression

This section describes the method of fitting a multiple linear regression model for a selected data set. The regression procedures are performed using a variance-covariance updating procedure described in Maindonald, J (1984). The least squared solution is facilitated by using Cholesky decomposition.

73.4.1 Available procedures

The procedure available in this section fits a linear model of the form $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$ where Y is the dependent variable (response) and X_1, \dots, X_k are the independent variables (predictors) and ε is a random error with a normal distribution having mean=0 and variance= σ^2 . The multiple linear regression algorithm computes the estimates $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$, of the regression coefficients $\beta_0, \beta_1, \dots, \beta_k$, so as to minimize the sum of squares of residuals.

The regression procedure

- Calculates the estimates of the regression coefficients, their standard errors, p-values, R^2 , and the contribution of each variable to reducing the total sum of squares.
- Performs the Wald test on groups of specified variables.
- Allows control of multicollinearity criterion (default 0.05) and number of components for collinearity diagnostics to be displayed (default 8).
- Computes the fitted values, ANOVA table and covariance matrix of the coefficients estimates.
- Computes various types of residuals-unstandardized, standardized, studentized and deleted.
- Computes influence statistics-Cook's distance, DFFIT's, covariance ratios and hat matrix diagonals.

73.4.2 Example: Multiple Linear Regression

Dataset: Werner.cyx

Data Description:

In this example, consider the data from a blood chemistry study described by Werner, et al (1985). Eight variables were recorded for n=188 women. The data includes the information on age, weight, birth pill (1=user, 2=nonuser), cholesterol, albumin and calcium. One of the aims of this study is to find the relationship between the variable **Cholesterol** and other variables.

Purpose of the Analysis:

To fit multiple linear regression model to the data with **Cholesterol** as dependent variable and **Age, Height, Weight, Birthpill, Albumin, Calcium, and Uric Acid** as

73 Analysis-Analytcs

independent variables. Also to obtain collinearity diagnostics and perform Wald test for **Albumin**, **Calcium**, and **Uric_Acid**.

Analysis Steps:

1. Open the dataset from Samples folder.
2. Choose the menu item:

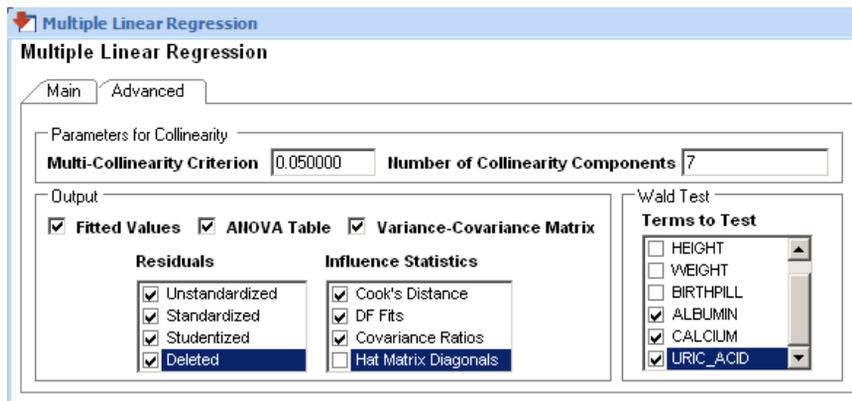
Analysis > (Basic Statistics) Analytics > Multiple Linear Regression

3. In the **Main** tab, select **Cholesterol** as the **Dependent Variable** and select the checkboxes against the remaining 7 variables, **Age**, **Height**, **Weight**, **Birthpill**, **Albumin**, **Calcium**, and **Uric_Acid** as independent variables. Click the **Wald Test** and **Collinearity Diagnostics** checkboxes.

The screenshot shows the 'Multiple Linear Regression' dialog box in SPSS, with the 'Main' tab selected. The 'Dependent Variable' is set to 'CHOLESTEROL'. The 'Independent Variables' list includes 'AGE', 'HEIGHT', 'WEIGHT', 'BIRTHPILL', and 'ALBUMIN', all of which are checked. The 'Wald Test' and 'Collinearity Diagnostics' checkboxes are also checked. The 'No Intercept Term' checkbox is unchecked. The 'Weightage Variable' field is empty.

4. In the **Advanced** tab, enter 7 as the **Number of Collinearity Components**. In the **Wald Test** box select the variables **albumin**, **calcium**, and **uric acid** for carrying out the Wald Test. In the **Output** box select all the checkboxes except

Hat Matrix Diagonals.



5. Click **OK**. You will see the Analysis results as shown below.

Analysis: Multiple Linear Regression

Input Parameters

Data File: Wemer.cyd
 Model: CHOLESTEROL=%CONST+AGE+HEIGHT+WEIGHT+BIRTHPILL+ALBUMIN+CALCIUM+URIC_ACID
 Number of Observations: 181

Output

Terms dropped due to

| | |
|-------------------|------|
| No variation | None |
| Multicollinearity | None |

Summary Statistics

| Item | Value |
|--------------------|-----------|
| Residual df | 173 |
| Multiple R-squared | 0.252 |
| Std. Dev. Estimate | 39.486 |
| Residual SS | 269731.34 |

Parameter Estimates

| Model term | Coefficient | Std. Error | 95% Confidence Interval | | t-statistic | p-value | SS |
|------------|-------------|------------|-------------------------|-------------|-------------|---------|-------------|
| | | | Lower Limit | Upper Limit | | | |
| %Const | 55.59 | 95.358 | -132.624 | 243.804 | 0.583 | 0.561 | 9951593.818 |
| AGE | 1.545 | 0.311 | 0.931 | 2.16 | 4.964 | 0 | 51320.975 |
| HEIGHT | -2.331 | 1.384 | -5.063 | 0.401 | -1.684 | 0.094 | 1123.665 |
| WEIGHT | 0.051 | 0.182 | -0.308 | 0.41 | 0.281 | 0.779 | 1196.139 |
| BIRTHPILL | 6.913 | 6.157 | -5.239 | 19.066 | 1.123 | 0.263 | 851.006 |
| ALBUMIN | 0.121 | 0.997 | -1.847 | 2.089 | 0.121 | 0.903 | 7263.968 |
| CALCIUM | 2.296 | 0.722 | 0.871 | 3.72 | 3.18 | 0.002 | 19442.367 |
| URIC_ACID | 0.706 | 0.281 | 0.15 | 1.261 | 2.509 | 0.013 | 9811.723 |

73 Analysis-Analytics

Collinearity Diagnostics

| Components | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-------------------|---------|--------|-------|-------|--------|--------|-------|
| Eigenvalues | 0.001 | 0.001 | 0.004 | 0.02 | 0.042 | 0.066 | 0.457 |
| Condition numbers | 114.324 | 74.361 | 45.85 | 19.49 | 13.356 | 10.569 | 4.026 |
| %CONST | 0.924 | 0.023 | 0.051 | 0 | 0.001 | 0 | 0 |
| AGE | 0.007 | 0.001 | 0.001 | 0.038 | 0.034 | 0.917 | 0.002 |
| HEIGHT | 0.761 | 0.19 | 0.048 | 0 | 0.001 | 0.001 | 0 |
| WEIGHT | 0.125 | 0.015 | 0.261 | 0.598 | 0 | 0.001 | 0 |
| BIRTHPILL | 0.013 | 0.002 | 0.065 | 0.014 | 0.001 | 0 | 0.9 |
| ALBUMIN | 0.007 | 0.134 | 0.81 | 0.038 | 0.006 | 0.004 | 0 |
| CALCIUM | 0.122 | 0.861 | 0.013 | 0.002 | 0.001 | 0.001 | 0 |
| URIC_ACID | 0 | 0.016 | 0.006 | 0.06 | 0.914 | 0.001 | 0.001 |

ANOVA

| Source | df | SS | MS | F-statistic | p-value |
|------------|-----|------------|-----------|-------------|---------|
| Regression | 7 | 91009.843 | 13001.406 | 8.339 | 0 |
| Error | 173 | 269731.34 | 1559.141 | | |
| Total | 180 | 360741.182 | | | |

Hypothesis Testing: [ALBUMIN=CALCIUM=URIC_ACID=0]

| Type of test | Statistic | DF | P-value |
|--------------|-----------|----|---------|
| Wald | 23.422 | 3 | 0 |

Multiple Linear Regression – Residuals

| #Obs | AGE | HEIGHT | WEIGHT | BIRTHPILL | ALBUMIN | CALCIUM | URIC_ACID | CHOLESTEROL |
|------|-----|--------|--------|-----------|---------|---------|-----------|-------------|
| 1 | 22 | 67 | 144 | 0 | 43 | 98 | 54 | 200 |
| 3 | 25 | 62 | 128 | 0 | 41 | 104 | 33 | 243 |
| 4 | 25 | 68 | 150 | 1 | 38 | 96 | 30 | 50 |
| 5 | 19 | 64 | 125 | 0 | 41 | 99 | 47 | 158 |
| 6 | 19 | 67 | 130 | 1 | 45 | 105 | 83 | 225 |
| 7 | 20 | 64 | 118 | 0 | 39 | 95 | 40 | 210 |
| 8 | 20 | 65 | 119 | 1 | 38 | 93 | 50 | 192 |
| 9 | 21 | 60 | 107 | 0 | 42 | 101 | 52 | 246 |
| 10 | 21 | 65 | 135 | 1 | 34 | 106 | 48 | 245 |
| 11 | 21 | 63 | 100 | 0 | 38 | 98 | 54 | 208 |
| 12 | 21 | 64 | 120 | 1 | 47 | 106 | 38 | 260 |
| 13 | 21 | 67 | 134 | 0 | 40 | 108 | 34 | 204 |
| 14 | 21 | 67 | 145 | 1 | 39 | 95 | 49 | 192 |
| 15 | 21 | 63 | 138 | 0 | 41 | 102 | 41 | 280 |
| 16 | 21 | 64 | 113 | 1 | 39 | 99 | 38 | 230 |
| 17 | 21 | 63 | 160 | 0 | 39 | 96 | 39 | 215 |
| 18 | 21 | 64 | 115 | 1 | 44 | 105 | 44 | 225 |

The **Terms dropped due to** table refers to some essential pre-processing of the data. If a particular independent variable assumes the same value throughout the data set, it is not really a ‘variable’ and has to be dropped. Its presence creates ‘singularity’ in the so called X matrix. In the present data set we do not have the problem and hence the entry is ‘none’. Multicollinearity is another similar feature of the data which makes the

problem unstable. Again, in the present dataset, no such difficulty is encountered and hence the entry 'None'. The **Summary Statistics** table gives an overview of the results. If the residual degrees of freedom are not adequate, we have too many independent variables. In the present case, the residual df value is 173. The data size is large, relative to the number of independent variables. Multiple R-squared indicates the fraction of the total variation explained by the selected set of independent variables. In this case, the value is 0.2523. If the data under study have high multicollinearity, estimates of regression coefficients become volatile and less dependable. To check this, examine the 'condition numbers'. (They indicate extent of spread in eigen-values of $X'X$). A very large number is a warning. Montgomery et al recommend use of 100 as indicative of moderate concern while a value of 1000 is an alarm trigger (Montgomery, Peck, and Vining, 2003, page 339). For this model, the number is 114.32. Thus, it may be pertinent to take corrective step such as centering the data. The next table is ANOVA. Our data has 181 observations and total degrees of freedom are 180 (one degree being spent on fitting an intercept in the model). There are 7 independent variables giving rise to 7 df for regression and remaining 173 df are assigned to error. The very low p-value shows that the model fitted is 'significant'. We have to reject the null hypothesis that all regression coefficients are zero. Lastly, we can test if a subset of the regression coefficients is zero. A test is carried out to check if coefficients of three independent variables (Albumin, calcium and uric acid) are zero. Here also, the p-value is very small and the hypothesis stands discredited.

73.5 *Multivariate Analysis of Variance*

This section describes various methods for analyzing a data set in which each observation consists of multiple measurements on the same experimental unit. As an example, if our study concerns the size of babies, we may measure length, chest girth, head girth and weight. In that case, there will be four measurements on each baby. We can of course study every measurement separately. However, the fact that they are correlated makes it necessary that they are studied together. All ideas applicable to analysis of univariate data are relevant here too. However, some aspects absent in univariate data arise in multivariate data. Procedures available in **East** for multivariate analysis include Multiple Linear Regression and Multivariate Analysis of Variance (MANOVA). References for the procedures covered in this section are provided in Johnson & Wichern (1998).

Multivariate Analysis of Variance (MANOVA) procedure is a generalization of univariate Analysis of Variance (ANOVA) procedure. When we have samples of observations from different multivariate normal populations having a common variance covariance matrix, we can use the MANOVA procedure to check for the equality of mean vectors.

73 Analysis-Analytcs

73.5.1 Available procedures

The available procedures under MANOVA are:

- One-way MANOVA
- Profile analysis

73.5.2 Example: Multivariate Analysis

Dataset: Root.cyx

Data Description:

The following Example is taken from Rencher (1995). In a classical experiment carried out from 1918 to 1934, apple trees of different rootstocks were compared (Andrews and Herzberg, 1985). The data for eight trees from each of six rootstocks are available. The variables in the data are:

y1 = trunk girth at 4 years (mm x 100)

y2 = extension growth at 4 years (m)

y3 = trunk girth at 15 years (mm x 100)

y4 = weight of tree above ground at 15 years (lb x 1000)

The table of mean vectors of the six rootstocks is shown below:

| Rootstock | Y1 | Y2 | Y3 | Y4 |
|-----------|--------|--------|--------|--------|
| 1 | 1.1375 | 2.9771 | 3.7388 | 0.8711 |
| 2 | 1.1575 | 3.1091 | 4.515 | 1.2805 |
| 3 | 1.1075 | 2.8152 | 4.455 | 1.3914 |
| 4 | 1.0975 | 2.8798 | 3.9063 | 1.039 |
| 5 | 1.08 | 2.5572 | 4.3125 | 1.181 |
| 6 | 1.0362 | 2.2146 | 3.5962 | 0.735 |

Purpose of the Analysis:

To perform One Way Multivariate Analysis of Variance (MANOVA) on the data with ROOTBOX as the group variable and Y1, Y2, Y3, Y4 as the response vector.

Analysis Steps:

1. Open the dataset from Samples folder.
2. Choose the menu item:

Analysis > (Basic Statistics) Analytics > Multiple Analysis of Variance

3. Select **ROOTBOX** in the **Group Variable**. Select all the checkboxes in the **Dependant Variable** box.

Multivariate Analysis of Variance

Group Variable ROOTSTOCK **Dependent Variable**

- Y1
- Y2
- Y3
- Y4

4. Click **OK**. You will see the Analysis results as shown below.

Analysis: Multivariate Analysis of Variance (MANOVA)

Input Parameters

Data File: Root.cyd
 Selected Variable (s): Y1,Y2,Y3,Y4
 Group Variable: ROOTSTOCK
 # Groups: 6
 Number of Obs. in Data File: 48
 No. of obs. Analysed: 48
 Number of Observations Rejected: 0

Output

| | Wilks' Lambda | F-Statistic | Num DF | Den DF | P-value |
|----------------------------|---------------|-------------|--------|--------|-----------|
| Test for equality of means | 0.154 | 4.937 | 20 | 131 | 0 |
| Test for parallel profiles | 0.17 | 6.662 | 15 | 111 | 4.779E-10 |

The p-values are very small and hence we reject the null hypothesis of equality of means as well as the hypothesis of parallel profile.

74 *Analysis-Plots*

The plotting capabilities available in the **Analysis** menu of **East** are of two types.

- Basic Plots
- Crossover Plots

These are essentially data exploration charts for the two types of data, case data and crossover data respectively.

This chapter discusses in detail the various types of basic plots and crossover plots . .

The following types of plots provide data exploration capabilities in **East**:

- Area
- Box
- Bubble
- Cumulative: (Left or Right)
- Density
- Histogram
- Simple Scatter
- Stem and Leaf
- Step Function
- Bar: (simple Bar, Stacked Bar, Horizontal Bar, or Stacked Horizontal Bar)
- Pie
- P-P Normal
- Q-Q Normal

74.1 *Data Exploration Plots*

The plots are further classified into Categorical, Continuous and Frequency Distribution. To generate a data exploration plot, open a data file and then choose from the menu:

Analysis> Basic Plots

Then you can select:

- **Categorical,**
- **Continuous,**
- **Frequency Distribution**

74.2 Categorical

74.2.1 Bar Chart

74.2.2 Pie chart

74.2.1 Bar Chart

Bar chart provides a choice of following graphical displays of the frequencies of the categories of a variable:

- Simple Bar
- Stacked Bar
- Horizontal Bar
- Horizontal Stacked Bar.

The display is in the form of vertical or horizontal bars, the height or length of the bars are proportional to the frequency of the categories shown in the X-axis.

Simple Bar

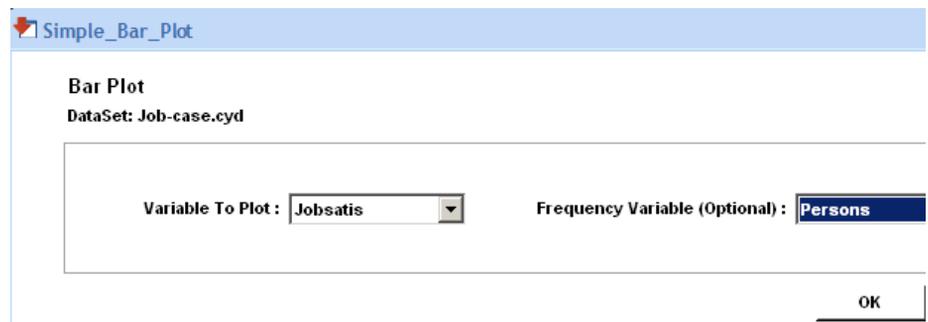
Dataset: Job-Case.cyx as described in Section 72.3.

Purpose of Plot :

For exploratory analysis and presentation, it is useful to summarize these data into a tabular form. The purpose is to generate and display a **Simple Bar** chart for the variable **Jobsatis**.

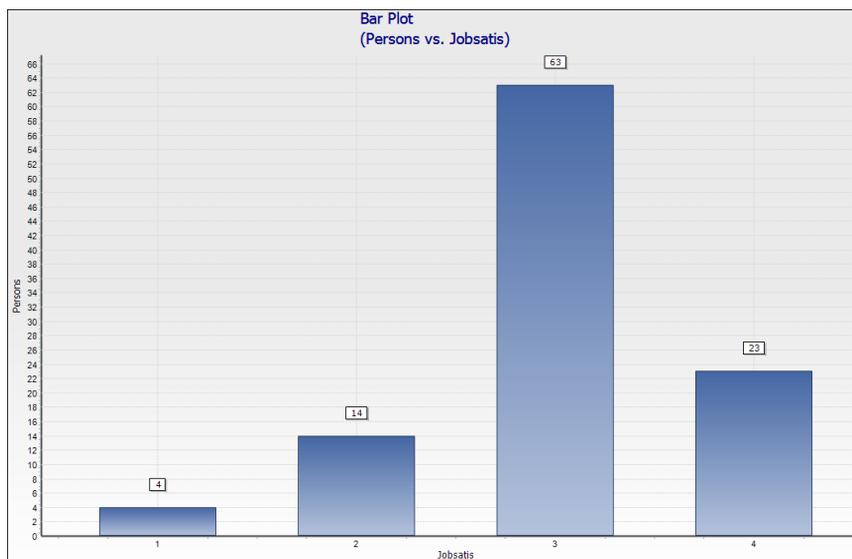
Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Categorical > Bar Chart > Simple Bar
3. In the ensuing dialog box, select **Jobsatis** as the **Variable to Plot** and **Persons** as the **Frequency Variable (Optional)**.



74 Analysis-Plots

4. Click **OK**. The following **Simple Bar** chart is displayed in the main window.



Stacked Bar

Dataset: Job-Case.cydx as described in Section 72.3

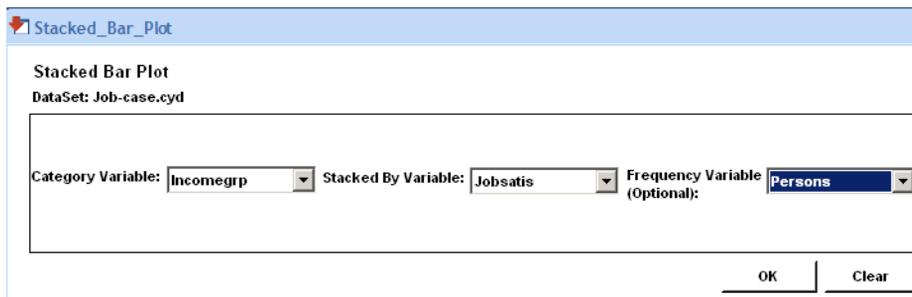
Purpose of Plot :

To generate and display a **Stacked Bar** chart for the variable **Incomegrp** stacked by **Jobsatis** based on the selected dataset.

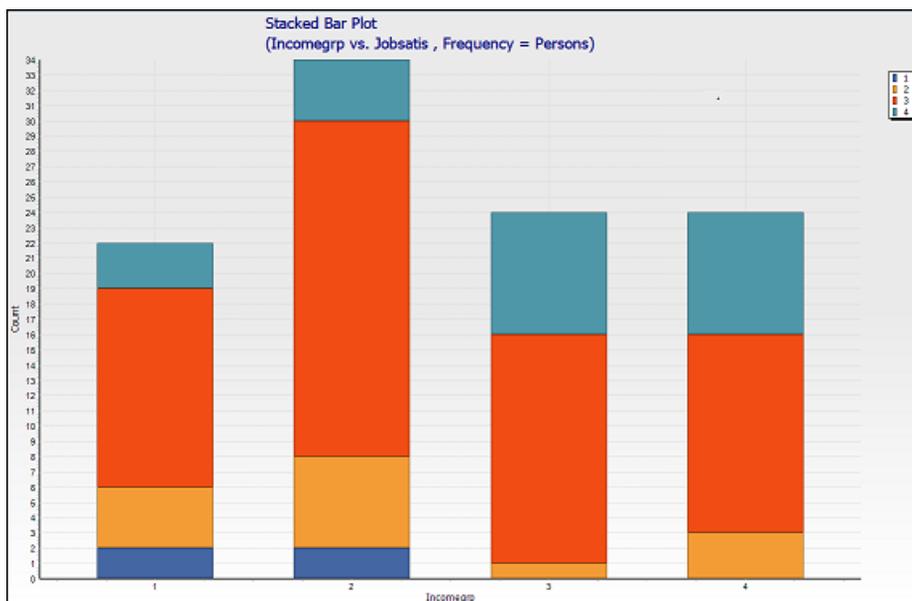
Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Categorical > Bar Chart > Stacked Bar
3. In the ensuing dialog box, select **Incomegrp** as the **Category Variable**, **Jobsatis**

as the **Stacked By Variable** and **Persons** as the **Frequency Variable (Optional)**



4. Click **OK**. The following **Stacked Bar** chart is displayed in the main window.



Horizontal Bar

Dataset: Job-Case.cydx as described in Section 72.3.

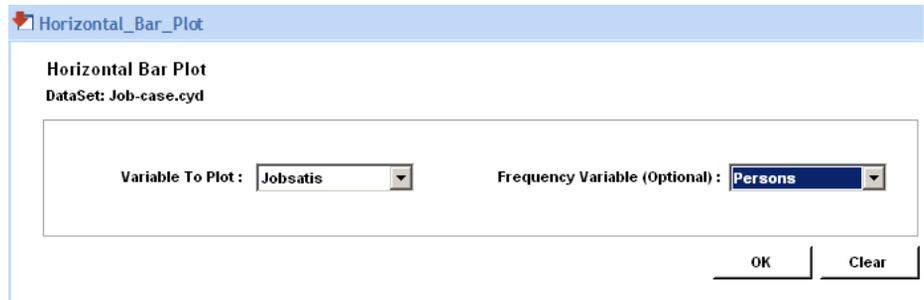
Purpose of Plot :

To generate a **Horizontal Bar** chart for the variable **Jobsatis**.

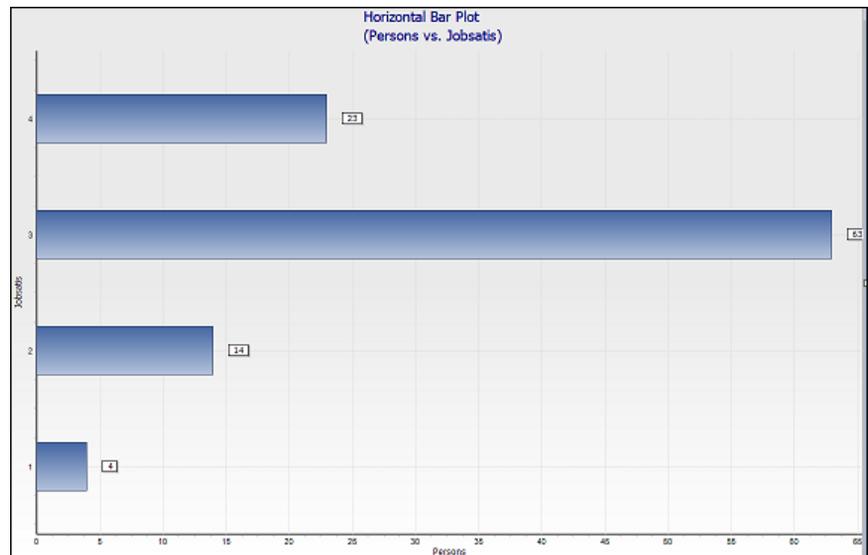
74 Analysis-Plots

Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Categorical > Bar Chart > Horizontal Bar
3. In the ensuing dialog box, select **Jobsatis** as the **X axis variable** and **Persons** as the **Frequency Variable (Optional)**.



4. Click **OK**. The following **Horizontal** Bar chart is displayed in the main window.

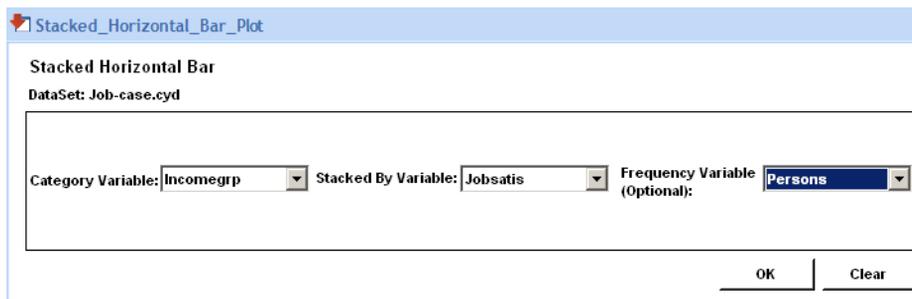


Horizontal Stacked Bar

Dataset: Job-Case.cyx as described in Section 72.3.

Purpose of the Analysis: To generate and display a **Horizontal Stacked Bar** chart for the variable **Incomegrp** stacked by **Jobsatis** based on the selected dataset. . **Steps:**

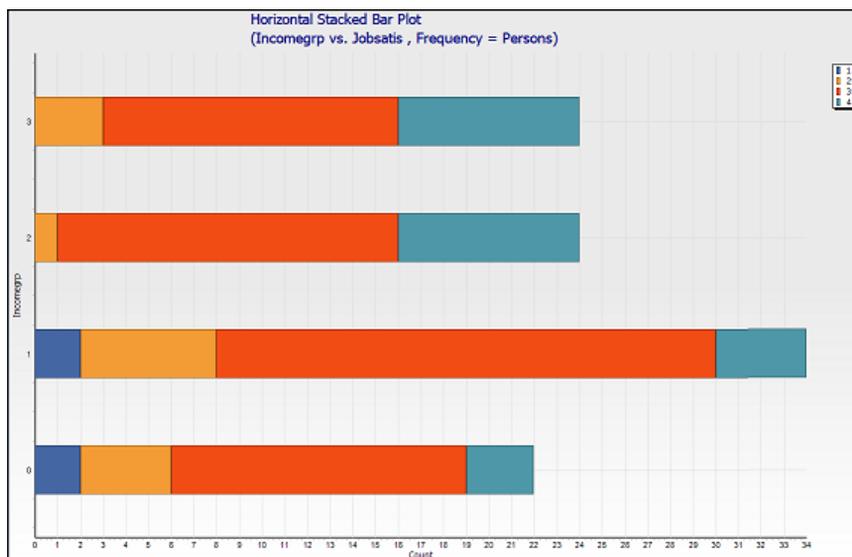
1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Categorical > Bar Chart > Horizontal Stacked Bar
3. In the ensuing dialog box, Select **Incomegrp** as the **Category Variable**, **Jobsatis** as the **Stacked By Variable** and **Persons** as the **Frequency Variable (Optional)**.



4. Click **OK**. The following **Horizontal Stacked Bar** chart is displayed in the

74 Analysis-Plots

main window.



74.2.2 Pie chart

Pie provides a circle graph divided into slices, each displaying the frequency of the category of a variable. The size of each slice is proportional to the relative frequency of the values.

Dataset: Socio.cydx

Data Description

This dataset contains measurements on 11 variables. There are 40 subjects in the study. The first 6 variables are concerned with the performance of the subject in the past while the last 5 variables reflect current performance and future plan.

Purpose of Plot :

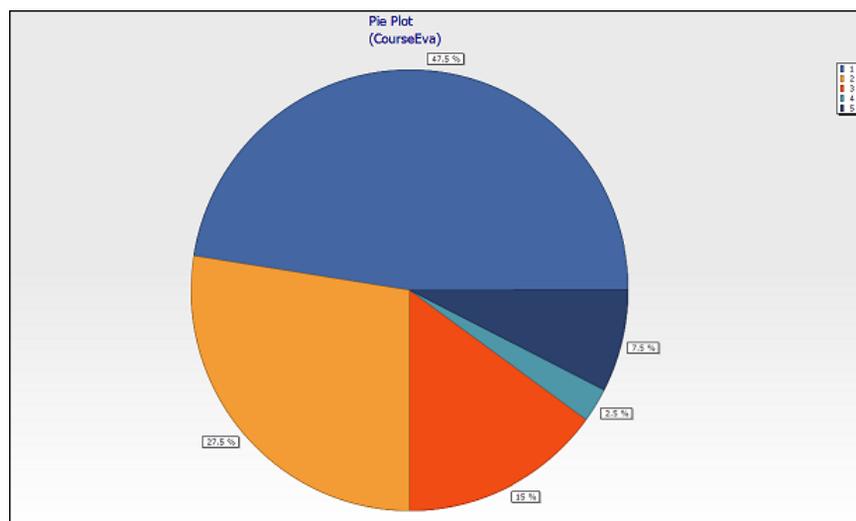
To generate and display a **Pie chart** for the variable **CourseEva**. **Steps:**

1. Open the dataset from **Samples** folder of the **East** Installation directory.
2. Choose the menu item:
Analysis > (Basic Plots) Categorical > Pie Chart
3. In the ensuing dialog box, select **CourseEva** as the **Variable To Plot** and leave

the **Summary Variable (optional)** blank.



4. Click **OK**. The following **Pie chart** is displayed in the main window.



74.3 Continuous

74.3.1 Area

74.3.2 Box

74.3.3 Bubble

74.3.4 Simple Scatter

74.3.5 Normality

74.3.1 Area

Area provides a graphical display of the trend of values of Y variable(s) over categories of an X variable. The display is in the form of shaded area(s) under the curve(s).

74 Analysis-Plots

Dataset: *Myeloma.cydx* as described in Section 72.1.1.

Purpose of Plot :

To generate and display an **Area chart** for the variable **haemoglobin** over **id**.

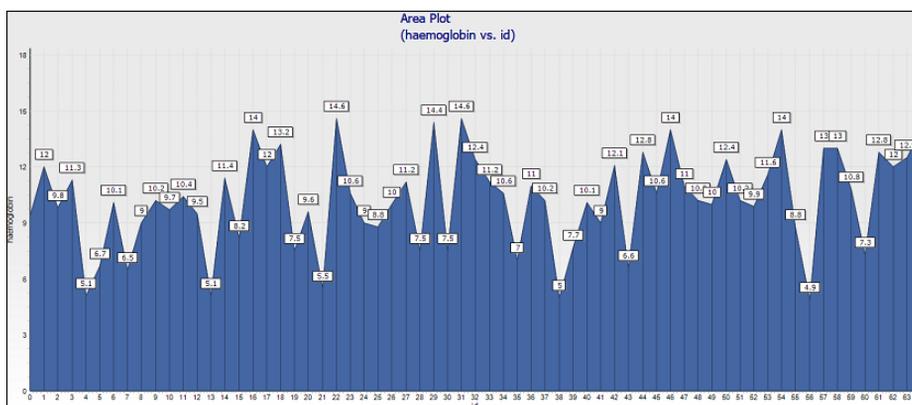
Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Continuous > Area
3. In the ensuing dialog box, select **id** as the **Variable To Plot** and **haemoglobin** as the **Frequency Variable (Optional)**.

Area Plot
DataSet: Myeloma.cydx

Variable To Plot : Frequency Variable (Optional) :

4. Click **OK**. The following **Area chart** is displayed in the main window.



74.3.2 Box

Box provides a data display that shows the 25th and 75th percentiles of the data (using the outline of the box), the median value (the large dashed line in the box), the mean value (smaller dashed line), and the largest and smallest data points (endpoints of the

vertical line going through the box).

Dataset: *Myeloma.cydx* as described in Section 72.1.1.

Purpose of Plot :

To generate and display a **Box chart** for the variable **haemoglobin** across different values of **status** based on the selected dataset.

Steps:

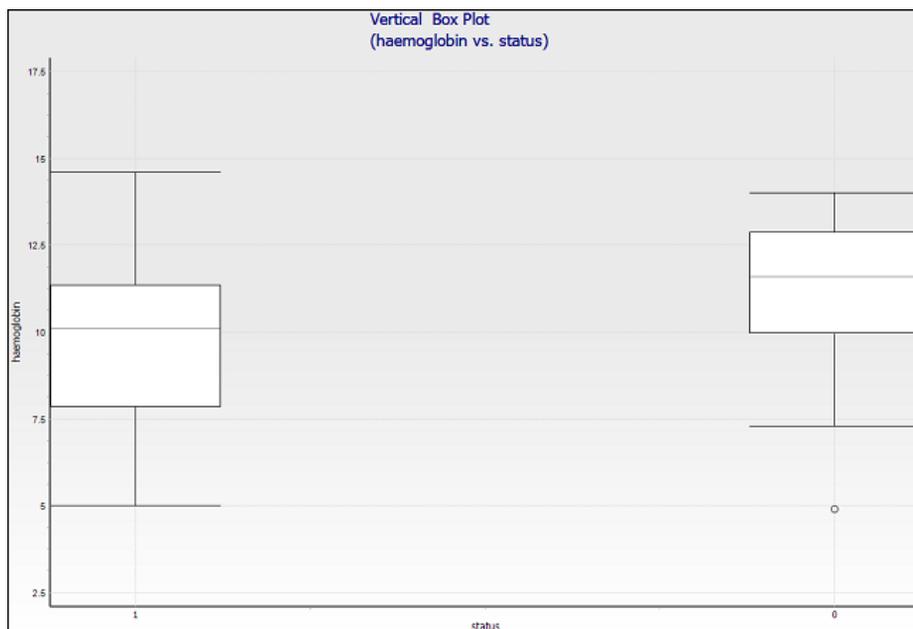
1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Continuous > Box
3. In the ensuing dialog box, select **status** as the **Category Axis (optional)** and **haemoglobin** as the **Variable Axis**.

The image shows a dialog box titled "Box Plot" with the following fields and options:

- DataSet:** Myeloma.cydx
- Category Axis (Optional):** status
- Variable Axis:** haemoglobin
- Frequency Variable (Optional):** (empty)
- Vertical Box:**
- Horizontal Box:**

74 Analysis-Plots

4. Click **OK**. The following **Box chart** is displayed in the main window.



74.3.3 Bubble

Bubble provides an X versus Y data display that shows the number of points at a particular x, y value with proportional size bubbles, to allow the user to gauge the relative amounts of information at discrete points.

Dataset: **Myeloma.cydx** as described in Section 72.1.1.

Purpose of Plot:

To generate and display a **Bubble chart** for **status** over **gender** based on the selected dataset.

Steps:

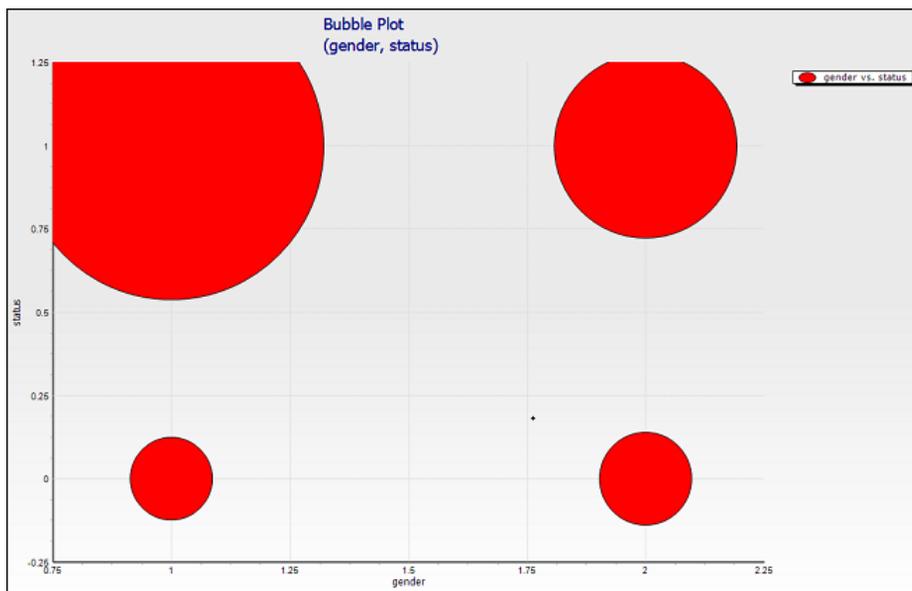
1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Continuous > Bubble
3. In the ensuing dialog box, select **gender** as the **Variable on X-Axis** and **status**

as the **Variable on Y-Axis** variable. Leave the **Frequency Variable (optional)** blank.

Bubble Plot
 DataSet: Myeloma.cyd

Variable on X-Axis: Variable on Y-Axis: Frequency Variable: (Optional)

4. Click **OK**. The following **Bubble chart** is displayed in the main window.



74.3.4 Simple Scatter

Simple Scatter provides an X versus Y scatter plot.

Dataset: Myeloma.cydx as described in Section 72.1.1.

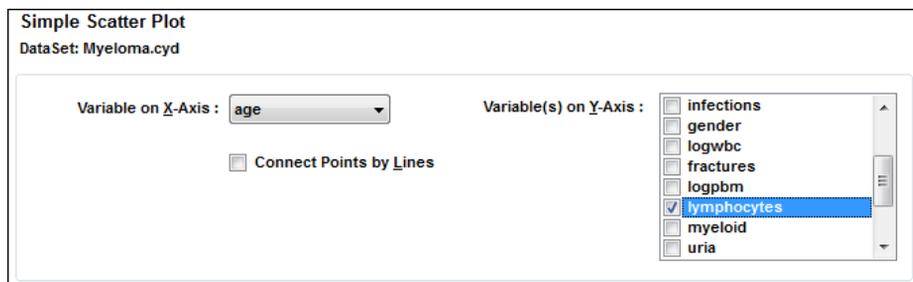
Purpose of Plot :

To generate and display a **Simple Scatter** chart for **lymphocytes** versus **age** based on the selected dataset.

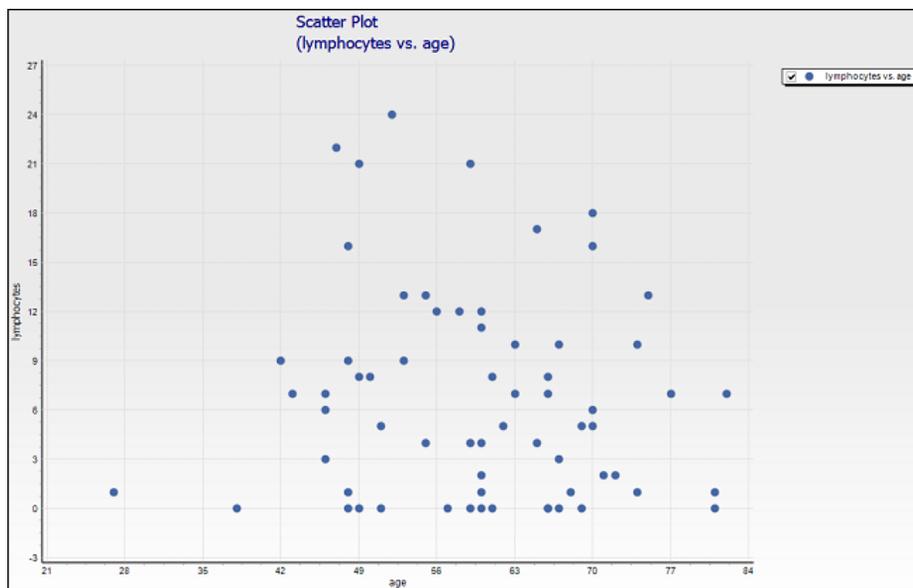
74 Analysis-Plots

Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Continuous > Simple Scatter
3. In the ensuing dialog box, select **age** as the **Variable on X-Axis** and **lymphocytes** as the **Variable on Y-Axis**.



4. Click **OK**. The following **Simple Scatter** chart is displayed in the main window.



74.3.5 Normality

PP Normal :

PP Normal provides a probability-probability (P-P) plot to see if the selected variable follows a normal distribution. The X-axis displays the observed cumulative probability and the Y-axis displays the expected cumulative probability. The plot should be approximately linear if the normal distribution is the correct model.

Dataset: **Socio.cydx** as described in Section 74.2.2.

Purpose of Plot :

To generate and display a **PP Normal** chart for the variable **FinalExam** based on the selected dataset.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Continuous > Normality > PP Normal
3. In the ensuing dialog box, select **FinalExam** as the **Variable To Plot**.

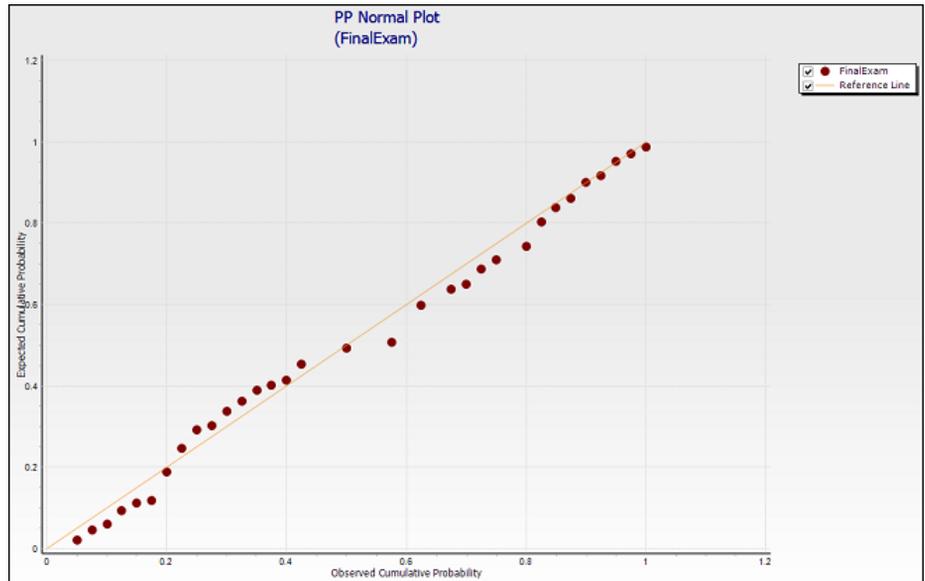
PP-Normal Plot

DataSet: Socio.cyd

Variable To Plot : FinalExam ▼

74 Analysis-Plots

4. Click **OK**. The following **PP Normal** chart is displayed in the main window.



QQ Normal :

QQ Normal provides a quantile-quantile (Q-Q) plot to see if the selected variable follows a normal distribution. The X-axis displays the observed normal value and the Y-axis displays the expected normal value. The plot should be approximately linear if the normal distribution is the correct model.

Dataset: **Socio.cyx** as described in Section 74.2.2.

Purpose of Plot :

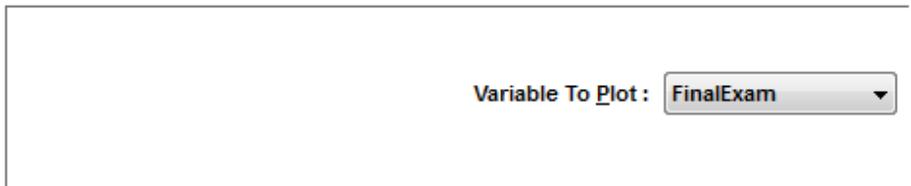
To generate and display a **QQ Normal** chart for the variable **FinalExam** based on the selected dataset.

Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Continuous > Normality > QQ Normal
3. In the ensuing dialog box, select **FinalExam** as the **Variable To Plot**.

QQ-Normal Plot

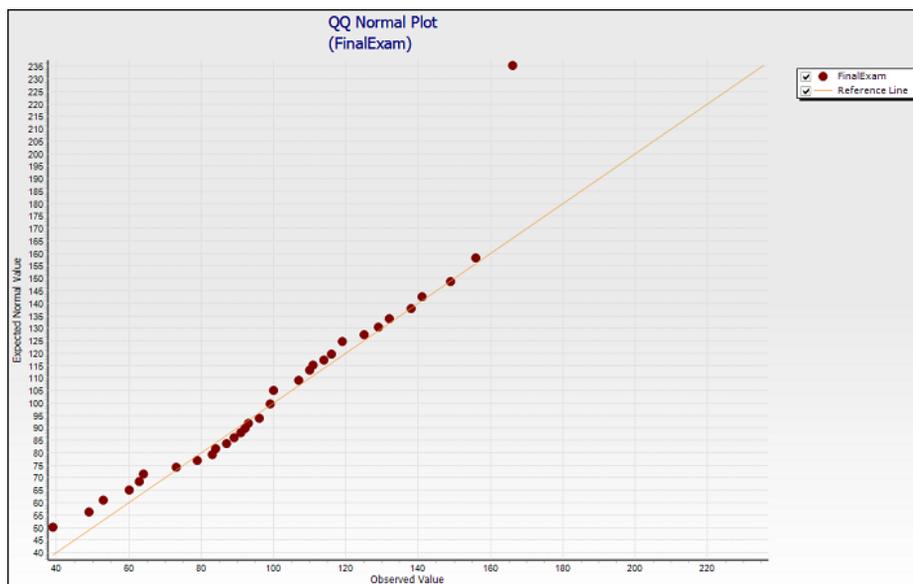
DataSet: **Socio.cyx**



Variable To Plot : FinalExam

74 Analysis-Plots

4. Click **OK**. The following **QQ Normal** chart is displayed in the main window.



74.4 Frequency Distribution

74.4.1 Cumulative Plot

74.4.2 Histogram

74.4.3 Stem and Leaf

74.4.4 Step Function

74.4.1 Cumulative Plot

Left cumulative

A left cumulative frequency plot is a way to display cumulative information graphically. It shows the number of observations that are less than or equal to particular values.

Dataset: Vari.cydx

Data Description:

A randomized clinical trial of Interferon and placebo was conducted on 44 children infected with childhood chicken pox (varicella) (Arvin, et al., 1982). One of the end points of the study was to determine whether Interferon is more effective than placebo in preventing adverse effects.

The dataset has three variables **Group**, **Category** and **Freq**. The **Group** variable

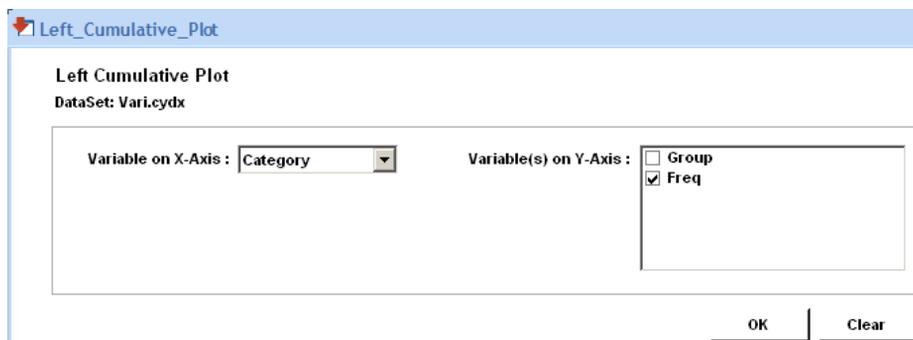
contains values 1 and 2 specifying the two groups **Interferon** and **Placebo**, respectively. The **Category** variable has four categories, representing the adverse effect starting from 'none' to 'death in less than a week' with values from 1 to 4 in increasing order. The number of children falling in each category, by treatment, is available in the variable **Freq**

Purpose of Plot :

To generate and display a **Left Cumulative** chart for **Freq** over the **Category** of adverse effects.

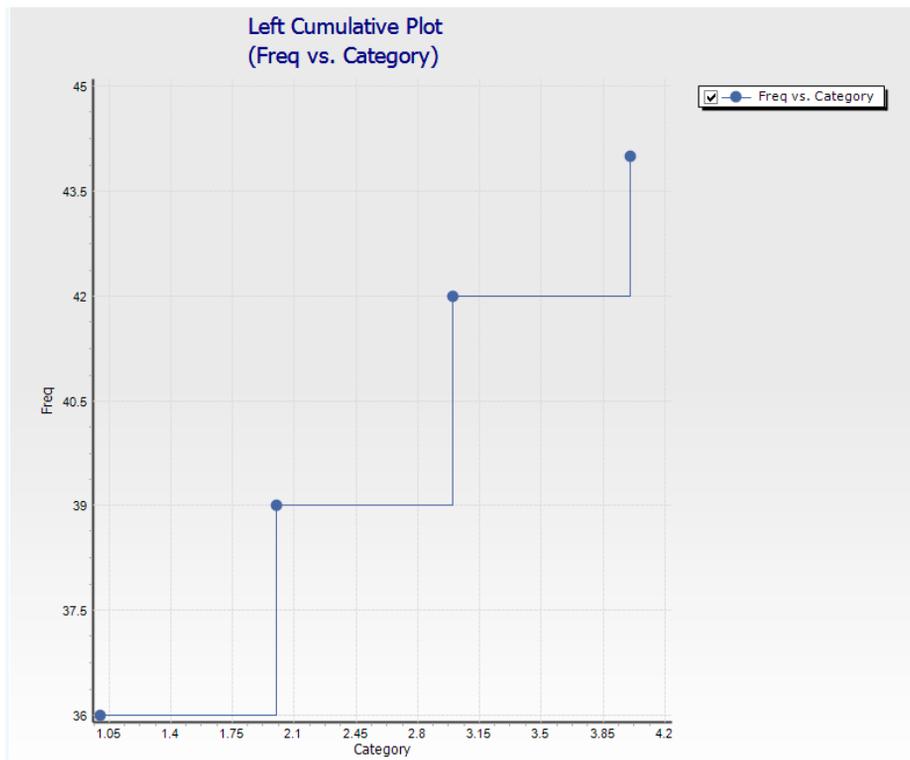
Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Frequency Distribution > Cumulative > Left Cumulative
3. In the ensuing dialog box, select **Category** as the **Variable on X-Axis** and **Freq** as the **Variable on Y-Axis**.



74 Analysis-Plots

4. Click **OK**. The following **Left Cumulative** chart is displayed.



Right cumulative

Dataset: **Vari.cydx** as described in Section [74.4.1](#).

Purpose of Plot :

To generate and display a **Right Cumulative** chart for **Freq** over the **Category** of adverse effects.

Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Frequency Distribution > Cumulative > Right Cumulative

3. In the ensuing dialog box, select **Category** as the **variable on X-Axis** and **Freq** as the **Variable on Y-Axis**.

Right_Cumulative_Plot

Right Cumulative Plot
DataSet: Vari.cydx

Variable on X-Axis : Category

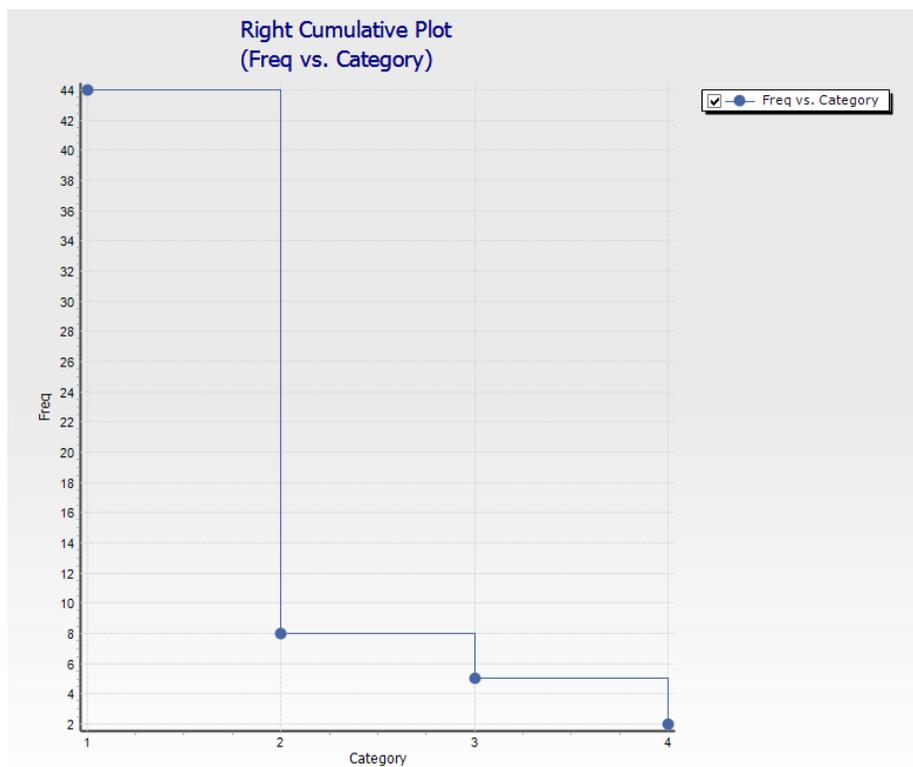
Variable(s) on Y-Axis : Group
 Freq

OK Clear

4. Click **OK**. The following **Right Cumulative** chart is displayed in the main

74 Analysis-Plots

window.



74.4.2 Histogram

Histogram provides a graphical display of the frequencies of the consecutive values of a variable. The display is in the form of contiguous bars, the height of the bars being proportional to the frequency of the values shown on the X-axis.

Dataset: [Myeloma.cydx](#) as described in Section [72.1.1](#).

Purpose of Plot :

To generate and display a **Histogram** chart for **age** based on the selected dataset.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

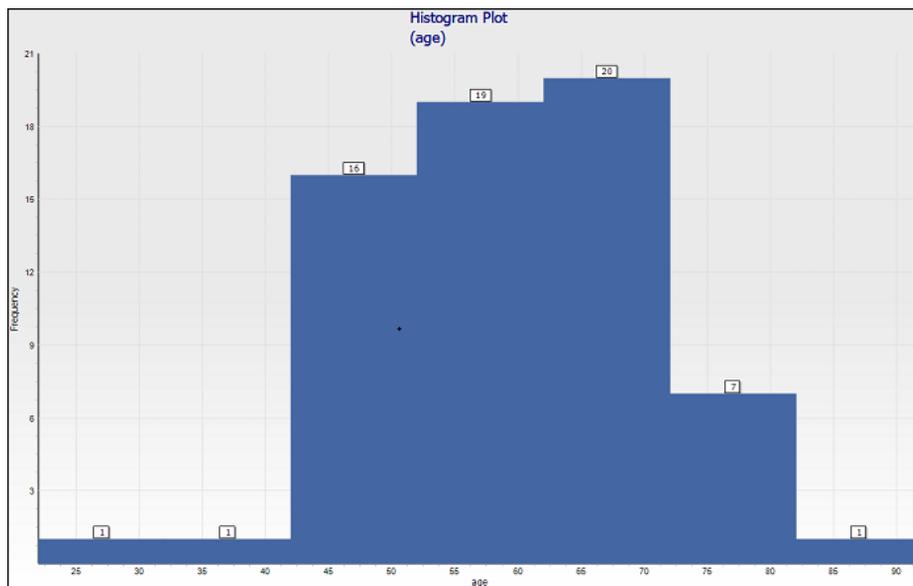
Analysis > (Basic Plots) Frequency Distribution > Histogram

- In the ensuing dialog box, select **age** as the **Variable To Plot**. Leave the **Frequency Variable** blank.

Histogram Plot
DataSet: Myeloma.cyd

Variable To Plot: age Frequency Variable (Optional):

- Click **OK**. The following **Histogram** chart is displayed in the main window.



74.4.3 Stem and Leaf

Stem and Leaf provides a way to form a diagrammatic display of data using data's number themselves.

Dataset: Myeloma.cydx as described in Section 72.1.1.

74 Analysis-Plots

Purpose of Plot :

To generate and display a **Stem and Leaf** chart for haemoglobin.

Steps:

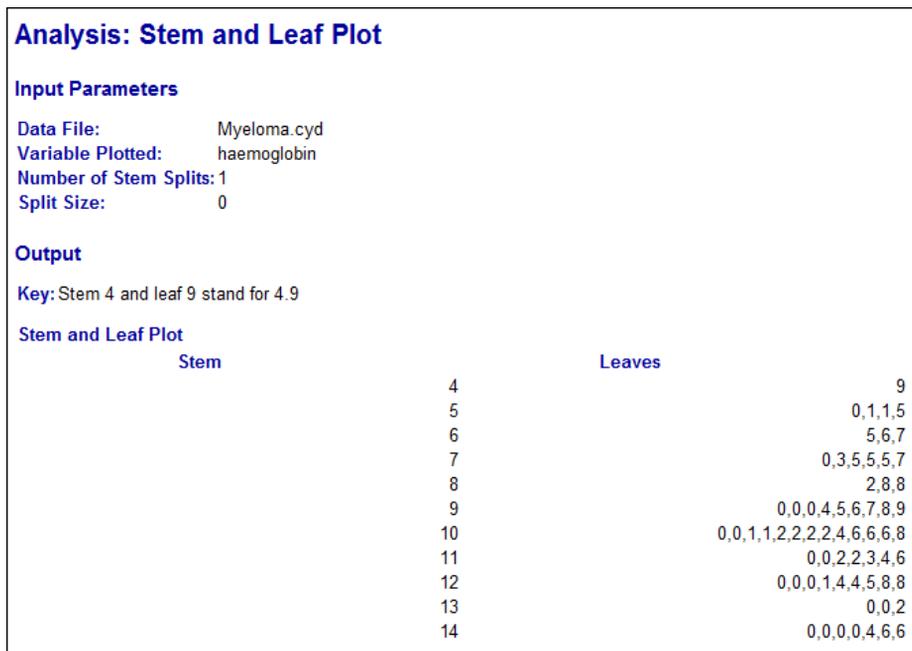
1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Frequency Distribution > Stem and Leaf
3. In the ensuing dialog box, select **haemoglobin** as the **Variable To Plot** and 1 as **Number of Stem Splits**. Enter 0 for **Stem Split Size**. Leave the **Frequency Variable (optional)** field blank.

Stem and Leaf Plot
DataSet: Myeloma.cyd

Variable To Plot: Frequency Variable (Optional):

Number Of Stem Splits: 1 2 5 Stem Split Size:

4. Click **OK**. The following **Stem and Leaf** chart is displayed in the main window.



5. Basically it is histogram type of plot, a histogram turned on its side. It resembles right half of a leaf with the stem on the left.

74.4.4 Step Function

Step Function provides a data display for a variable that changes its value at discrete intervals. **Dataset: Survival.cyx**

Purpose of Plot:

To generate and display a **Step Function** chart for the variable **SurvPer** based on the selected dataset.

Steps:

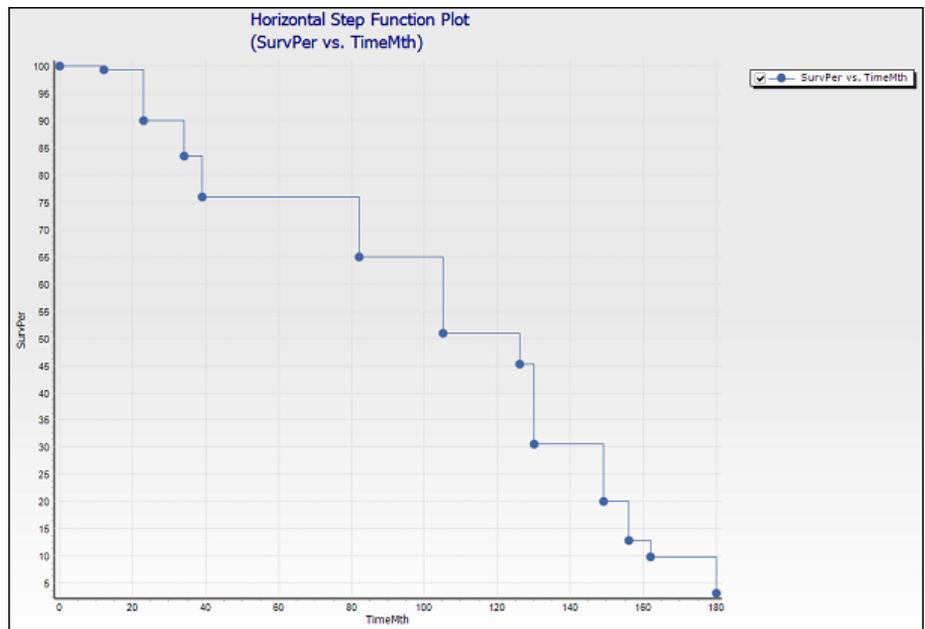
1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Basic Plots) Frequency Distribution > Step Function

74 Analysis-Plots

- In the ensuing dialog box, select **TimeMth** as the **Variable on X-Axis** and **SurvPer** as the **Variable on Y-Axis**.



- Click **OK**. The following **Step Function** chart is displayed in the main window.



74.5 Crossover Plots

In this section, we consider data obtained from a 2x2 cross over trial. We also assume that the response measured in each period of the trial has been recorded on a continuous scale. The data may either be in the form of a regular case data in bf East or a Crossover Patients Continuous Data generated using the crossover data editor. Plotting of crossover data helps in understanding the data well and also in getting an idea about the difference in treatment or period effects. Three important plots used specifically for crossover data are described here.

- Period_2 Vs. Period_1 Plot
- Subject Profile Plot
- Treatment-by-Periods Plot

We will first address drawing of these plots using case data related to a 2x2 crossover trial. To generate a crossover plot, open a case data file and then choose from the menu:

Analysis > (Crossover Plots) Subject Plots

Then you can choose any of the three plots, **Period_2 Vs. Period_1 Plot**, **Subject Profile Plot**, **Convex Hull**.

74.5.1 Period_2 Vs. Period_1 Plot

Period_2 Vs. Period_1 Plot provides a scatter plot of points for each patient where the response in period 1 is on X axis and the response in period 2 is taken on Y axis.

Dataset: PEFR.cyd

Data Description:

Data from a single-centre, randomized, placebo-controlled, double-blind study carried out to evaluate the efficacy and safety of an inhaled drug (A) given to patients with chronic obstructive pulmonary disease on mean morning expiratory flow rate (PEFR) compared with a placebo B. In all 56 patients were involved in the study, 27 in the < AB > group who received treatment A in the first period and B in the second and 29 in the < BA > group who received treatment B in the first period and A in the second. The data are taken from Jones and Kenward (1989).

Steps

- Open the data file **PEFR.CYD**.
- Choose the menu item:
Analysis > (Crossover Plots) Subject Plots > Period 2 Vs. Period 1 Plot

74 Analysis-Plots

- In the ensuing dialog box, select **Group_ID** as the **Group ID**, **Period_ID** as **Period ID**, **Patient_ID** as **Subject ID** and **PEFR** as **Response**. The dialog box will now look as shown below.

Period_2_Vs_Period_1_Plot

Period_2 vs. Period_1 Plot
DataSet: PEFR.eyd

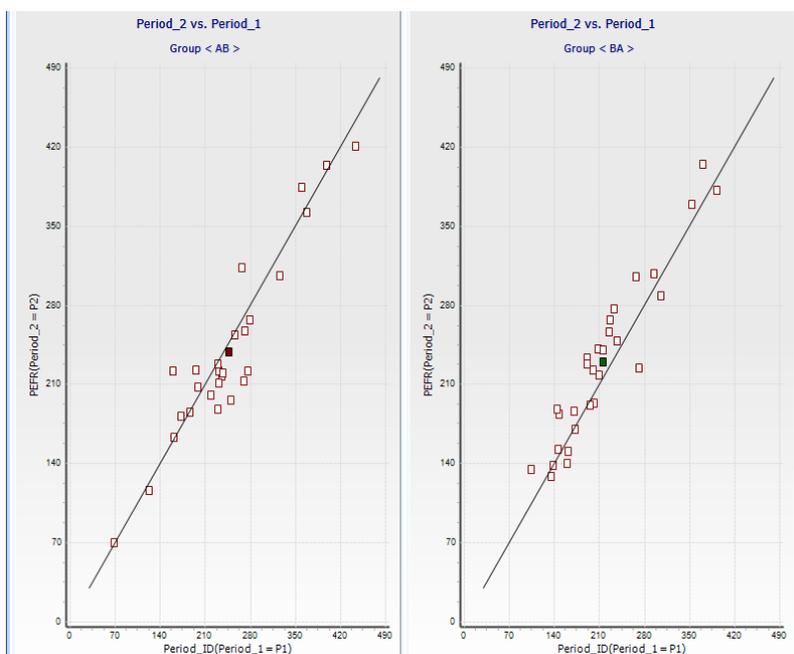
Group ID: Period ID:

Subject ID: Response:

Combine Groups Use Log Transform of Responses

OK Clear

- Click on **OK**. The following graph is produced.



The filled points represent the means of data called 'Centroids'. The line $Y=X$ is a line with slope 1 and intercept 0. Note that there is tendency for the plotted points to be

below the line in Group < AB > and above it in Group < BA >. Thus observations on treatment A tend to be greater than those on the placebo B. This is observed in both the groups. The points for each group are quite spread out, indicating high between-patient variability. We can also see that one of the patients has a very low mean PEFR values. You can take the cursor to the lowest point in Group < AB > and read the values which are (67.778, 70.278). The fact that the points from the two groups are almost symmetrically placed in relation to the diagonal is evidence for the absence of a period effect. To determine evidence for a direct treatment effect, we will plot a combined plot for both the groups. To do this, choose Period_2 Vs. Period_1 Plot from the **Cross Over Plots** menu and select the variables. This time select the check box for **Combined Groups**.

The dialog box will now look as shown below.

Period_2 Vs. Period_1 Plot

Period_2 vs. Period_1 Plot
DataSet: PEFR.cyd

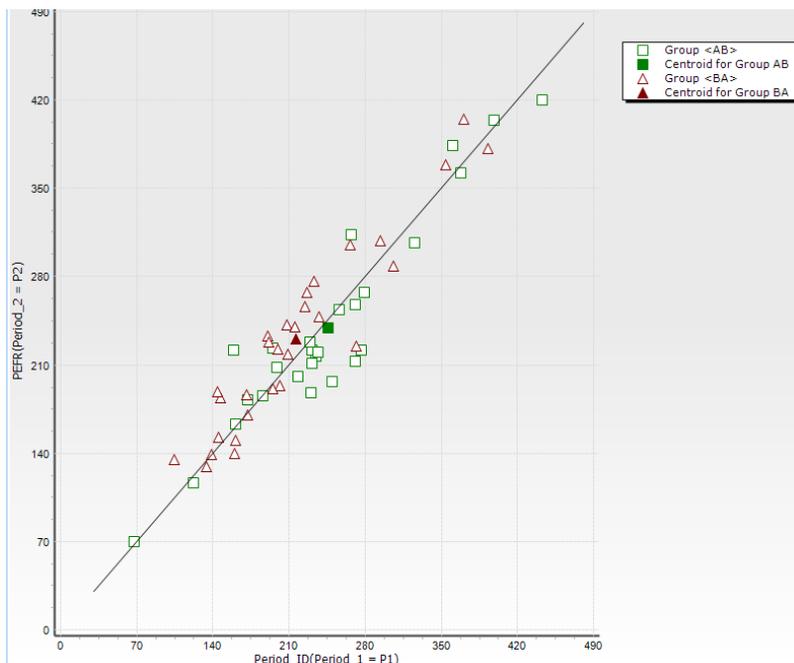
Group ID : Group_ID Period ID : Period_ID
Subject ID : Patient_ID Response : PEFR

Combine Groups Use Log Transform of Responses

OK Clear

74 Analysis-Plots

Click on **OK**. The following graph is produced.



Again, the filled points represent the centroids of the respective groups. The fact that the centroids are placed either side of the line with some vertical separation is evidence of a direct treatment effect.

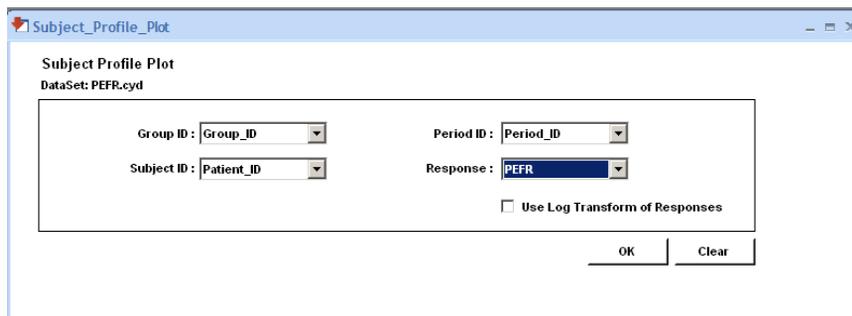
74.5.2 Subject Profile Plot

The objective of a crossover trial is to focus attention on within-patient treatment differences. A good plot for displaying these differences is the subject-profiles plot. In this plot, the change in each patient's response over the two treatment periods is plotted for each group. To draw the Subject Profile Plot, choose the menu item,

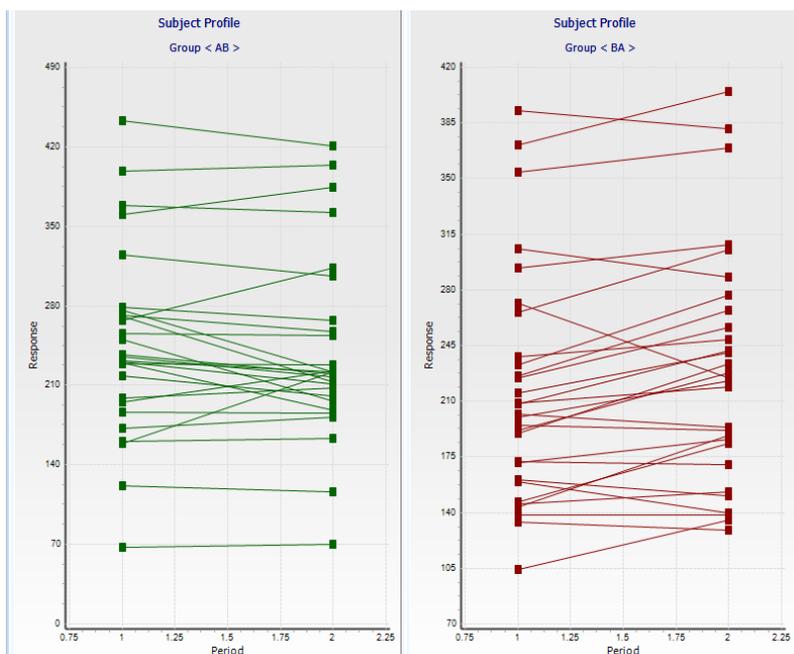
Analysis > (Crossover Plots) Subject Plots > Subject Profile

In the ensuing dialog box, select **Group_ID** as the **Group ID**, **Period_ID** as **Period ID**, **Patient_ID** as **Subject ID** and **PEFR** as **Response**. The dialog box will now

look as shown below.



Click on **OK**. The following graph is produced.



From the Subject Profile plot also the high between-patient variability is noticeable. The within patient changes are generally negative in Group < AB > and positive in Group < BA > with some exceptions. For Group < AB >, the slopes of lines are

74 Analysis-Plots

negative implying higher values for Period 1 where treatment A is applied. For Group $\langle BA \rangle$, the slopes of lines are positive showing higher values for Period 2 where A is applied. Thus the general trend implies a higher value of mean PEFr for treatment A rather than for placebo B. Most of the changes are smaller in magnitude barring some large ones.

74.5.3 Treatment-by-Periods Plot

Both the Period_2 Vs. Period_1 and Subject Profile Plots display values of Response for individual patients. To get the overall idea of the performance of both the treatments in two periods, a graph such as Treatment-by-Periods Plot is used. To draw this plot for the PEFr data:

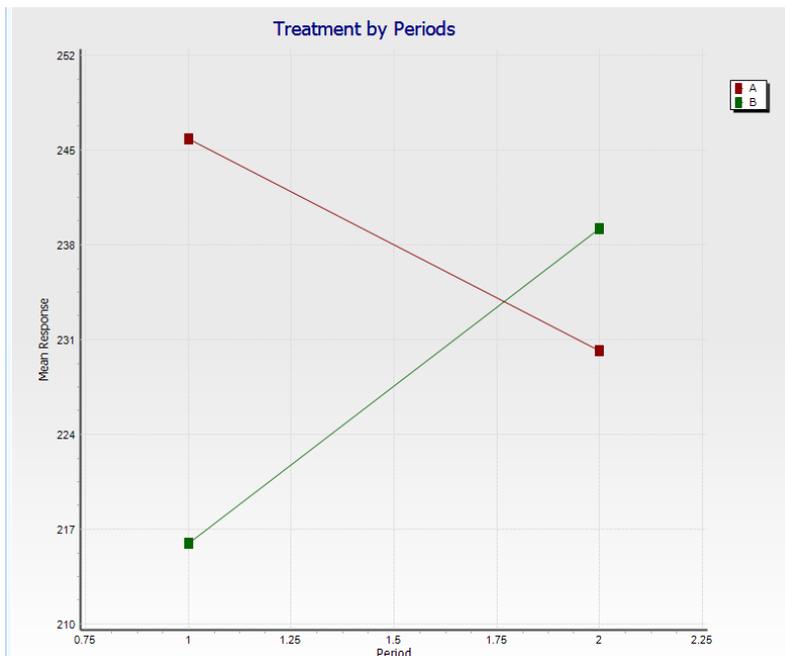
1. Choose from the menu:
Analysis > (Crossover Plots) Summary Plots > Treatment by Periods
2. In the ensuing dialog box, select **Group_ID** as the **Group ID**, **Period_ID** as **Period ID**, **Patient_ID** as **Subject ID** and **PEFR** as **Response**. In the text boxes for Treatment 1 and Treatment 2, type A and B respectively which are the treatments used in the study. The dialog box will now look as shown below.

Treatment_Period_Plot

Treatment by Periods Plot
DataSet: PEFr.cyd

| | | | | |
|---|---|-------------|--|---|
| Group ID : | <input type="text" value="Group_ID"/> | Period ID : | <input type="text" value="Period_ID"/> | Treatment Specifications Treatment 1 ((Group1,Period1) and (Group2,Period2)) <input type="text" value="A"/> Treatment 2 ((Group1,Period2) and (Group2,Period1)) <input type="text" value="B"/> |
| Subject ID : | <input type="text" value="Patient_ID"/> | Response : | <input type="text" value="PEFR"/> | |
| <input type="checkbox"/> Use Log Transform of Responses | | | | <input type="button" value="OK"/> <input type="button" value="Clear"/> |

3. Click on **OK**. The following graph is produced.

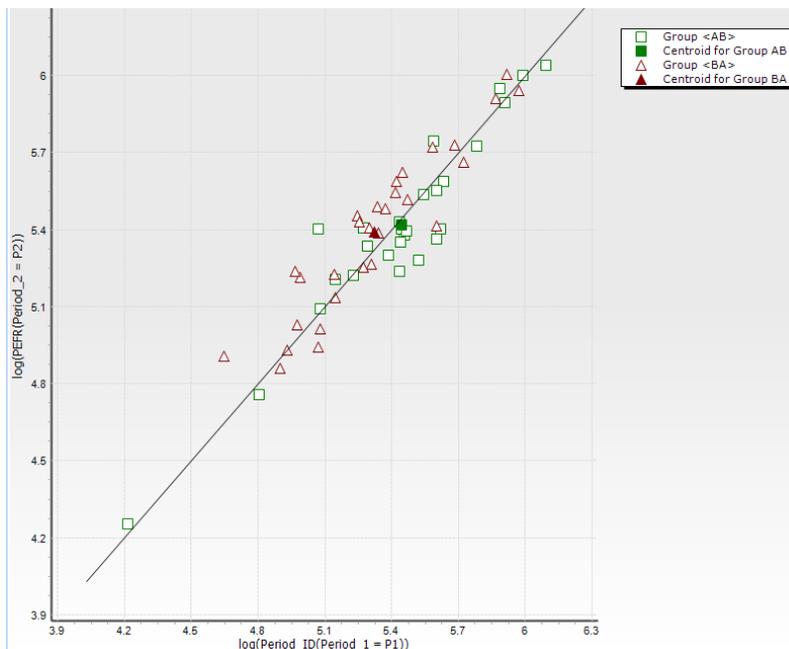


The points plotted are the means of the Response variable PEFR for Treatment and Period combination. As shown in the legend by the side of the plot, the lines join means for respective treatments for Period 1 and Period 2. If a cursor is taken to any of the points, it shows the label **Group ID** as well as the value of mean response for the corresponding treatment-by-period combination. Since no line is completely above the other, neither treatment gives higher mean response in both periods and the observed difference in means is smaller in the second period compared to the first. In Period 1, the difference is 29.847 whereas in Period 2 it is -9.041. To test whether this difference is statistically significant or not, the user is referred to the **Period Effect** test from the Crossover menu. All the above three plots could be drawn on log scale by checking the decision box for **'Use log(Response)'**. The response values are transformed to $\log(\text{Response})$ where natural logarithm of Response is plotted.

4. For instance, suppose we want to draw the Period_2 Vs. Period_1 Plot on log scale for the PEFR data, then choose from menu:
Analysis > (Crossover Plots) Subject Plots > Period 2 Vs. Period 1 Plot
5. In the ensuing dialog box, select **Group_ID** as the **Group ID**, **Period_ID** as **Period ID**, **Patient_ID** as **Subject ID** and **PEFR** as **Response**. Check

74 Analysis-Plots

the decision box for 'Use log(Response)' as well as for '**Combine Groups**'. Click on **OK**. The following graph is produced.

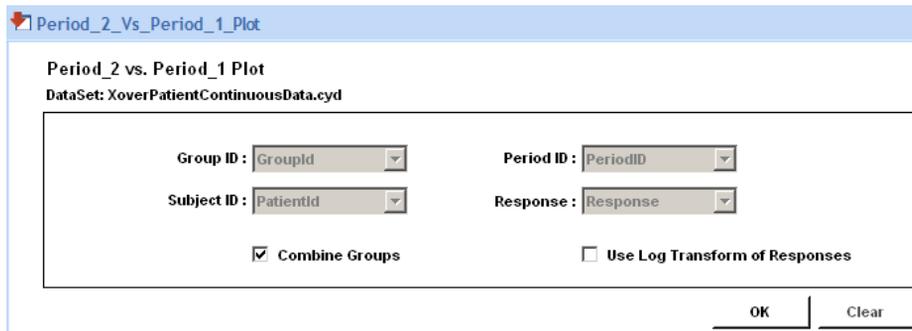


74.5.4 Crossover Plots using Crossover Patients Continuous Data

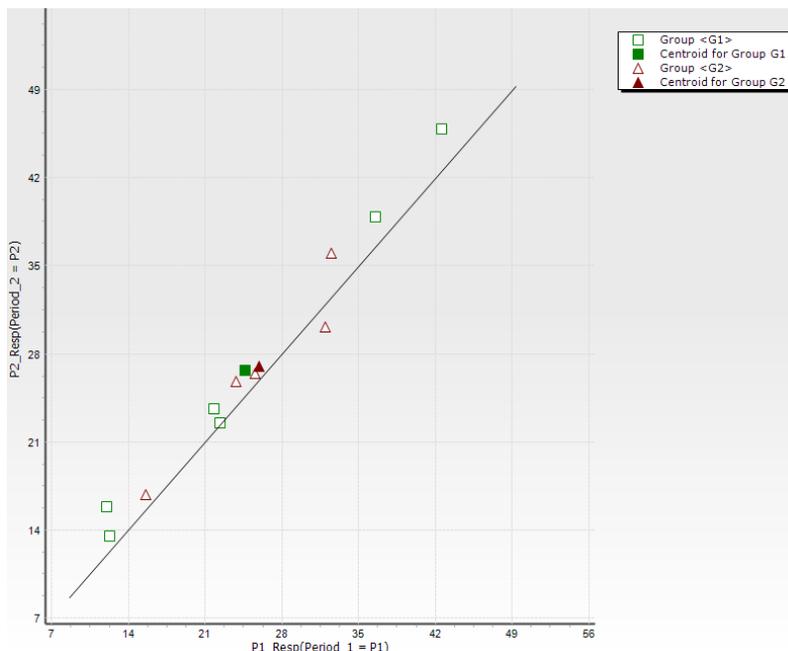
All the above graphs can be drawn on the crossover patients **continuous** data created by cross over data editor. To see this,

1. Open the data file, **XoverPatientContinuousData.cyd** available in the **Samples** directory of the Crossover installation directory.
2. Now choose the menu item:
Analysis > (Crossover Plots) Subject Plots > Period 2 Vs. Period 1 Plot

You will be presented with the following dialog box.



3. Check for **Combine Groups** as shown in the dialog box and click **OK**. The following graph is produced.



Note that when you are drawing these plots using Cross over Data editor, there is no need to select the variables etc, as they will be internally selected and the called plot will be drawn. For example, if you go in for the Subject Profile Plot on the same data,

74 Analysis-Plots

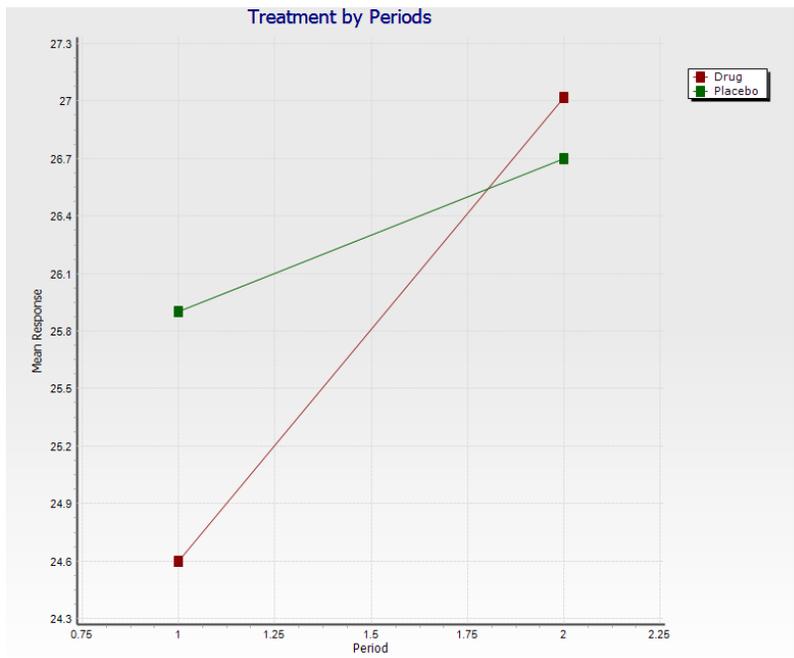
you will receive the dialog box as follows:

If you check **Use Log Transform of Responses**, the plot will be drawn on the log scale, otherwise on the original scale of the response variable. Similarly, the dialog box you get when you attempt to draw the **Treatment-by-Periods Plot** will be as follows:

| | PatientId | GroupId | P1_Resp | P2_Resp | var | var | var | var |
|----|-----------|---------|---------|---------|-----|-----|-----|-----|
| 1 | P1 | G1 | 12.3 | 13.5 | | | | |
| 2 | P3 | G1 | 12 | 15.8 | | | | |
| 3 | P5 | G1 | 36.5 | 38.9 | | | | |
| 4 | P7 | G1 | 21.8 | 23.6 | | | | |
| 5 | P9 | G1 | 22.4 | 22.5 | | | | |
| 6 | P11 | G1 | 42.6 | 45.9 | | | | |
| 7 | P2 | G2 | 15.6 | 16.8 | | | | |
| 8 | P4 | G2 | 25.6 | 26.4 | | | | |
| 9 | P6 | G2 | 23.8 | 25.8 | | | | |
| 10 | P8 | G2 | 32.5 | 36 | | | | |
| 11 | P10 | G2 | 32 | 30.1 | | | | |
| 12 | | | | | | | | |

You may specify the Treatment specifications of your choice such as **Drug** and **Placebo**, as is shown in the above dialog box. The plot will then have these

specifications in the legend as shown below:



If suppose the treatments are not specified in the text boxes provided in the dialog box, then the plot will have the default **Treatment 1** and **Treatment 2** specifications in the legend.

75 *Analysis-Normal Superiority One-Sample*

This chapter demonstrates how **East** can be used to perform inferences on data collected from a single-sample superiority study. This may consist of a random sample of observations from either a single treatment or paired observations from two treatments.

In chapter 7, the design, simulation and interim monitoring of these types of trials are discussed with reference to a Single Mean Test, a Test for the Difference of Paired Means and a Test for the Ratio of Paired Means.

East supports the analysis of all of these tests as well as the Wilcoxon Signed Rank Test. They are accessible from the **Analysis** menu and allow the validation of whether the data supports the null or alternative hypothesis of the study.

Analysis of a Single Mean Superiority Test is discussed in section 75.1, while the Two Paired Tests are discussed in section 75.2 and 75.3, respectively. Finally, the analysis of the non-parametric Wilcoxon Test is discussed in section 75.4.

75.1 *Example: Single Mean*

Consider the problem of comparing the mean of the distribution of observations from a single random sample to a specified constant. For example, when developing a new drug for treatment of a disease, there should be evidence of efficacy. In this example, the effect of a drug on children with mental retardation and ADHD is demonstrated.

For the single-sample problem, it may be desirable to compare the unknown mean response μ_t to a fixed value μ_0 . The null hypothesis $H_0: \mu_t = \mu_0$ is tested against either the two-sided alternative hypothesis $H_1: \mu_t \neq \mu_0$ or a one-sided alternative hypothesis $H_1: \mu_t < \mu_0$ or $H_1: \mu_t > \mu_0$.

Dataset: Methylphenidate.cyx

Data Description:

A trial was conducted to study the effect of Methylphenidate on cognitive functioning in children with mental retardation and ADHD. For the study details, refer to Pearson et al. (2003). For the twenty four children studied, the mean number of correct responses was observed for those receiving treatment (0.60 mg/kg of Methylphenidate) as well as those on placebo.

The first column of the dataset **D0** displays the number of correct responses after placebo, the second column **D60** shows the correct number of responses after

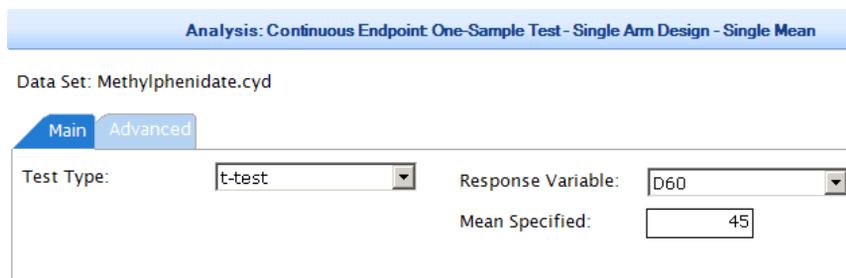
treatment (0.60 mg/kg of Methylphenidate), and the third column **diff** is the difference of the two measures.

Purpose of the Analysis:

To test whether the mean number of correct responses of children receiving treatment (0.60 mg/kg of Methylphenidate) is at least 45.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) One Sample > (Single Arm Design) Single Mean
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below.



4. Click **OK**. You will see the Analysis results as shown below.

Analysis: Continuous Response: Single Mean

Hypothesis

$H_0 : \mu_t = \mu_0$ Vs. $H_1 : \mu_t \neq \mu_0$ for 2-sided test

Either $H_1 : \mu_t > \mu_0$

Or $H_1 : \mu_t < \mu_0$ for 1-sided test

Input Parameters

Data File: Methylphenidate.cyd

Response Variable: D60

Test Type: t-test

Mean Specified: 45

Confidence Level: 0.95

75 Analysis-Normal Superiority One-Sample

Output

Response variable: D60
 Total no. of records: 24
 No. of records rejected: 0
 Summary of the Observed Data:

| Response | Min | Max | Median | Mean | Std.Dev | n |
|----------|-----|-----|--------|--------|---------|----|
| D60 | 29 | 77 | 42.5 | 44.708 | 12.32 | 24 |

Test of Hypothesis:

| n | Std. Effect Size | Std. Error | t | DF | 1-Sided | | 2-Sided | | 95% Confidence Interval(2-Sided) | |
|----|------------------|------------|--------|----|---------|------|---------|-------------|----------------------------------|--|
| | | | | | p-value | Tail | p-value | Lower Limit | Upper Limit | |
| 24 | -0.024 | 2.515 | -0.116 | 23 | 0.454 | L.E. | 0.909 | 39.506 | 49.911 | |

For this analysis, **East** displays the p-value associated with a left-tailed test because the observed sample mean is smaller than μ_0 . The two-sided 95% confidence interval is (39.506, 49.911). The lower limit is smaller than $\mu_0 = 45$, therefore $H_0: \mu_t \leq 45$ cannot be rejected in favor of $H_1: \mu_t > 45$ at one-sided 0.025 level of significance. The computed p-values also support this conclusion.

75.2 Example: Mean of Paired Differences

The paired t-test is often used to compare the means of two normal distributions. Here each observation from a random sample in one distribution is matched with a unique observation from the other distribution. A common application of this is when treatments are compared by using subjects who are matched using demographic and baseline characteristics. Another application is when two separate observations are made from the same subject under different experimental conditions, which will be the focus of the next example. **Dataset: Methylphenidate.cydx** as described in Section 75.1

Purpose of the Analysis:

To test the efficacy of Methylphenidate on cognitive functioning in children with mental retardation and ADHD. Let μ_0 and μ_t denote the mean number of correct responses under placebo and treatment, respectively, and $\delta = \mu_t - \mu_0$. A positive value of δ suggests efficacy of the treatment. Test the null hypothesis $H_0: \delta \leq 0$ against the alternative $H_1: \delta > 0$.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) One Sample > (Paired Design) Mean of Paired

Differences

- In the ensuing dialog box, under the **Main** tab choose the variables as shown below.

- Click **OK**. You will see the Analysis results as shown below.

Analysis: Continuous Response: Difference of Means for Paired Data

Hypothesis
 $H_0: \mu_t - \mu_c = 0$ Vs. $H_1: \mu_t - \mu_c \neq 0$ for 2-sided test
 Either $H_1: \mu_t - \mu_c < 0$
 Or $H_1: \mu_t - \mu_c > 0$ for 1-sided test

Input Parameters
 Data File: Methylphenidate.cyd
 Trial Type: Superiority
 Response Control: D0
 Response Treatment: D60
 Test Type: t-test
 Confidence Level: 0.95

Output
 Response variable: (D60-D0)
 Total no. of records: 24
 No. of records rejected: 0

Summary of the Observed Data:

| Response | Min | Max | Median | Mean | Std.Dev | n |
|----------|-----|-----|--------|--------|---------|----|
| D0 | 26 | 71 | 36 | 39.75 | 11.315 | 24 |
| D60 | 29 | 77 | 42.5 | 44.708 | 12.32 | 24 |

Test of Hypothesis:

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | 1-Sided | | 2-Sided | | 95% Confidence Interval(2-Sided) | |
|----|---------------------|------------------|------------|-------|----|---------|------|---------|-------------|----------------------------------|--|
| | | | | | | p-value | Tail | p-value | Lower Limit | Upper Limit | |
| 24 | 4.958 | 0.658 | 1.539 | 3.222 | 23 | 0.002 | G.E. | 0.004 | 1.775 | 8.141 | |

The two-sided 95% confidence interval is (1.775, 8.141), which does not include 0. The one sided p-value is 0.002 which supports this conclusion. Therefore, for this

75 Analysis-Normal Superiority One-Sample

example, it is reasonable to conclude that the use of Methylphenidate significantly increases mean number of correct responses as compared to placebo.

75.3 Example: Ratio of Paired Differences

The ratio of paired differences test is used to compare the means of two log normal distributions when each observation in the random sample from one distribution is matched with a unique observation from the other distribution. As with the previous example illustrating the mean of paired differences, a common application is when two observations are made from the same subject under different experimental conditions. Another is when treatments are compared using subjects who are matched by demographic and baseline characteristics, which will be the focus of the next example.

East is used to perform a log transformation on the original data, and a ratio of paired differences test on the log-transformed data.

Dataset: `Methylphenidate.cyx` as described in Section 75.1

Purpose of the Analysis:

To test the efficacy of Methylphenidate on cognitive functioning in children with mental retardation and ADHD. Define $\rho = \frac{\mu_t}{\mu_c}$. A value of $\rho > 1$ suggests efficacy of the treatment. Test the null hypothesis $H_0: \rho = 1$ against the alternative $H_1: \rho > 1$.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) One Sample > (Paired Design) Mean of Paired Ratios
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown

below.

Analysis: Continuous Endpoint One-Sample Test - Paired Design - Mean of Paired Ratios

Data Set: Methylphenidate.cyd

Main Advanced

Trial Type: Superiority Use Ratio

Test Type: t-test Response Control: D0

Response Treatment: D60

4. Click **OK**. You will see the Analysis results as shown below.

Analysis: Continuous Response: Ratio of Means for Paired Data

Hypothesis

$H_0 : \mu_1 / \mu_2 = 1$ Vs. $H_1 : \mu_1 / \mu_2 \neq 1$ for 2-Sided test
 Either $H_1 : \mu_1 / \mu_2 < 1$
 Or $H_1 : \mu_1 / \mu_2 > 1$ for 1-Sided test

Input Parameters

Data File: Methylphenidate.cyd
 Trial Type: Superiority
 Response Control: D0
 Response Treatment: D60
 Test Type: t-test
 Confidence Level: 0.95

Response Variable: (D60/D0)
 Total Number of Records: 24
 Number of Records Rejected: 0

Summary of the Observed Data:

| Response | Min | Max | Median | Mean | Std.Dev | n |
|----------|-----|-----|--------|--------|---------|----|
| D0 | 26 | 71 | 36 | 39.75 | 11.315 | 24 |
| D60 | 29 | 77 | 42.5 | 44.708 | 12.32 | 24 |

Test of Hypothesis for: ln(D60/D0)

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | 1-Sided | | 2-Sided |
|----|---------------------|------------------|------------|-------|----|---------|------|---------|
| | | | | | | p-value | Tail | p-value |
| 24 | 0.118 | 0.611 | 0.039 | 2.991 | 23 | 0.003 | G.E. | 0.007 |

Parameter Estimates:

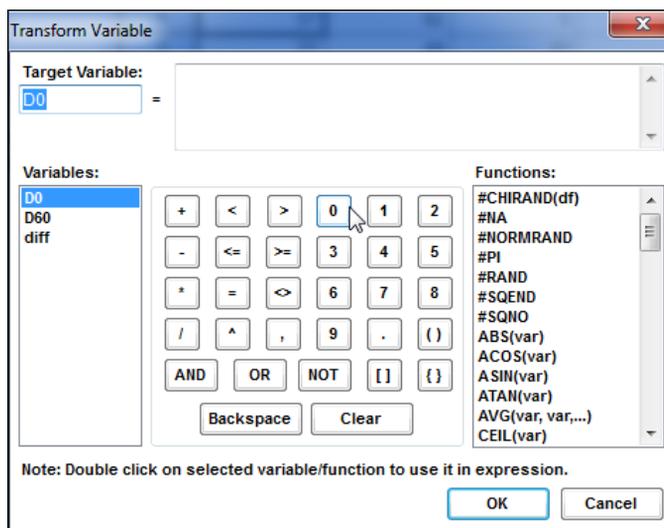
| Metric | Estimate | Std. Error | 95% Confidence Interval(2-Sided) | |
|------------|----------|------------|----------------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| ln (Ratio) | 0.118 | 0.039 | 0.036 | 0.199 |
| Ratio | 1.125 | 0.044 | 1.037 | 1.221 |

The observed value of test statistic is $t = 2.991$ and has $24 - 1 = 23$ degrees of freedom. The two-sided 95% confidence interval for $\ln(\rho)$ is (0.036, 0.199),

75 Analysis-Normal Superiority One-Sample

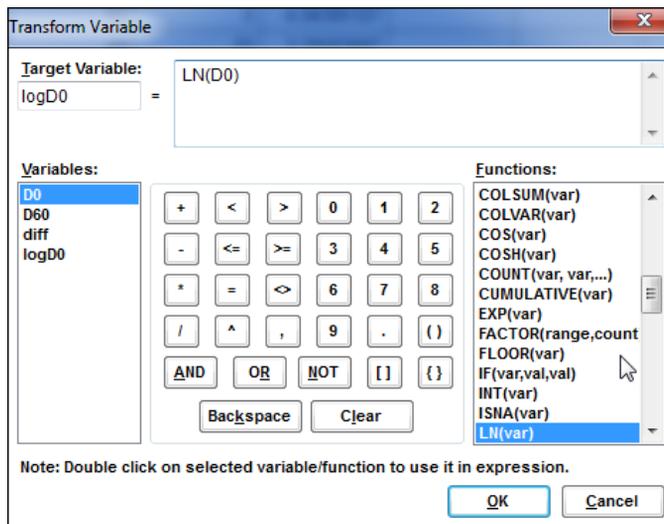
does not include 0, nor does the confidence interval for ρ (1.037, 1.221) contain 1. Therefore, $H_0: \rho = 1$ should be rejected in favor of $H_1: \rho \neq 1$, and the associated p-value of 0.007 supports this conclusion. The p-value for the one-sided test $H_0: \rho \leq 1$ versus $H_1: \rho > 1$ is 0.003. Again, for this example, it is reasonable to conclude that the use of Methylphenidate significantly increases mean number of correct responses as compared to placebo.

- Alternatively, a new log-transformed variable can be created directly in the dataset. Double click on **Methylphenidate.cyd** in the **Library** to display the data in the main window. Under the **Data Editor** tab, click the  icon in the **Variable** ribbon.



- Enter **logD0** in the **Target Variable** field. Type **LN(D0)** into the empty field on the right side of the equation, or select **D0** from the **Variables** list and **LN(var)**

from the **Functions** list:



7. Clicking **OK** will add a new column labeled `logD0` to the dataset. This contains log-transformed values of the entries in the `D0` column. In a similar manner, create a new variable `logD60` by transforming `D60` and perform a paired t-test

75 Analysis-Normal Superiority One-Sample

using logD0 and logD60.

Analysis: Continuous Response: Difference of Means for Paired Data

Hypothesis
 $H_0: \mu_t - \mu_c = 0$ Vs. $H_1: \mu_t - \mu_c \neq 0$ for 2-sided test
 Either $H_1: \mu_t - \mu_c < 0$
 Or $H_1: \mu_t - \mu_c > 0$ for 1-sided test

Input Parameters
 Data File: Methylphenidate.cyd
 Trial Type: Superiority
 Response Control: logD0
 Response Treatment: logD60
 Test Type: t-test
 Confidence Level: 0.95

Output
 Response variable: (logD60-logD0)
 Total no. of records: 24
 No. of records rejected: 0
 Summary of the Observed Data:

| Response | Min | Max | Median | Mean | Std.Dev | n |
|----------|-------|-------|--------|-------|---------|----|
| logD0 | 3.258 | 4.263 | 3.584 | 3.648 | 0.26 | 24 |
| logD60 | 3.367 | 4.344 | 3.749 | 3.766 | 0.263 | 24 |

Test of Hypothesis:

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | 1-Sided | | 2-Sided | | 95% Confidence Interval(2-Sided) | |
|----|---------------------|------------------|------------|-------|----|---------|------|---------|-------------|----------------------------------|--|
| | | | | | | p-value | Tail | p-value | Lower Limit | Upper Limit | |
| 24 | 0.118 | 0.611 | 0.039 | 2.991 | 23 | 0.003 | G.E. | 0.007 | 0.036 | 0.199 | |

Notice that the value of observed test statistic and the p-values are identical with those from the test for the ratio of means for paired data. In **East**, this test is equivalent to the paired t-test for log-transformed data.

75.4 Example: Wilcoxon Signed Rank Test

The non-parametric Wilcoxon signed rank test compares the median of the difference of two paired random variables. This test is equivalent to a nonparametric version of the paired t-test, and is preferred when the distribution of data deviates from normal.

Dataset: Methylphenidate.cydx as described in Section 75.1.

Purpose of the Analysis:

To test the null hypothesis $H_0: \lambda \leq 0$ against the alternative $H_1: \lambda > 0$ where λ is the median value of the paired difference. A positive value for λ suggests efficacy of the treatment.

Analysis Steps:

1. Open the dataset from **Samples** folder.

2. Choose the menu item:
Analysis > (Continuous) One Sample > (Paired Design) Wilcoxon Signed Rank
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below.

Continuous Endpoint: One-Sample Test - Paired Design - Wilcoxon Signed Rank

Data Set: Methylphenidate.cyd

Main Advanced

Trial Type: Superiority Use Diff. Variable

Response Control: D0

Response Treatment: D60

75 Analysis-Normal Superiority One-Sample

4. Click **OK**. You will see the Analysis results as shown below.

Analysis: Continuous Response: Wilcoxon Signed Rank Test

Hypothesis
 $H_0: \lambda = 0$ Vs. Either $H_1: \lambda > 0$ Or $H'_1: \lambda < 0$ for 2-sided test
 $H_0: \lambda = 0$ Vs. $H_1: \lambda > 0$ and $H'_1: \lambda < 0$ for 1-sided test
 where λ is the median of (Treatment Response – Control response.)

Input Parameters

Data File: Methylphenidate.cyd
 First Response: D0
 Second Response: D60
 Confidence Level: 0.95

Output

Response Variable: (D60-D0)
 Total Number of Records: 24
 Number of Records Rejected: 0
 Summary of the Observed Data:

| Response | Min | Max | Median | Mean | Std.Dev | n |
|----------|-----|------|--------|--------|---------|----|
| 26 | 71 | 36 | 39.75 | 11.315 | | 24 |
| 29 | 77 | 42.5 | 44.708 | 12.32 | | 24 |

Summary of Test Statistic:

| Minimum | Maximum | Mean | Std. dev. | Observed Statistics | Standardized |
|---------|---------|------|-----------|---------------------|--------------|
| 0 | 300 | 150 | 34.975 | 251.5 | 2.902 |

Test of Hypothesis:

| Estimate of Median Difference | Standardized Statistic | (1 Sided) | | (2 Sided) | 95% Confidence Interval(2-Sided)* | |
|-------------------------------|------------------------|-----------|---------|-----------|-----------------------------------|-------------|
| | | Tail | p-value | p-value | Lower Limit | Upper Limit |
| 5 | 2.902 | G.E. | 0.002 | 0.004 | 1.5 | 8 |

*The Nonparametric estimates are based on Hodges-Lehmann's formulation.

The **Estimate of Median Difference** has been calculated to be 5 and the observed **Standardized Statistic** is 2.902 with an associated 2-sided p-value of 0.004 and one-sided p-value of 0.002. The two-sided 95% confidence interval for λ is (1.5, 8) and does not include 0. Therefore, $H_0: \lambda \leq 0$ should be rejected in favor of $H_1: \lambda > 0$. The non-parametric Wilcoxon signed rank test also supports the reasonable conclusion that the use of Methylphenidate significantly increases mean number of correct responses as compared to placebo.

76 Analysis-Normal Noninferiority Paired-Sample

In this chapter, we explore how we can use **EasT** to perform inference on data collected from a paired-sample noninferiority study. Two common applications of paired sample designs are:

1. Comparison of two treatments using subjects who are matched by demographic and baseline characteristics, and
2. Two observations are made from the same subject under different experimental conditions.

Designing and simulation of such kind noninferiority trials are discussed in chapter 8. Analysis based on Paired Difference of Means is presented in section 76.1 and the Ratio of Paired Means is discussed in section 76.2.

76.1 Example: Mean of Paired Differences

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of an outcome variable, X, with means μ_t and μ_c , respectively, and with a standard deviation of paired difference as σ_D^2 . Let δ_0 be the noninferiority margin. For $\delta_0 < 0$, the null hypothesis $H_0: \mu_t - \mu_c \leq \delta_0$ is tested against the one-sided alternative hypothesis $H_1: \mu_t - \mu_c > \delta_0$. For $\delta_0 > 0$, the null hypothesis $H_0: \mu_t - \mu_c \geq \delta_0$ is tested against the one-sided alternative hypothesis $H_1: \mu_t - \mu_c < \delta_0$.

Dataset: Olestra.cyd

Data Description:

The dataset **Olestra.cyd** available in **Samples** folder contains paired observations from 28 subjects on two variables X and Y. Let μ_x and μ_y denote the population means of variables X and Y, respectively and $\delta = \mu_y - \mu_x$.

Purpose of the Analysis:

To test the null hypothesis $H_0: \delta \leq \delta_0$ against the alternative hypothesis $H_1: \delta > \delta_0$. For this example, we consider a non-inferiority margin of $\delta_0 = -0.5$.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) One Sample > (Paired Design) Mean of Paired Differences

76 Analysis-Normal Noninferiority Paired-Sample

- In the ensuing dialog box, under the **Main** tab choose the variables as shown below.

Analysis: Continuous Endpoint One-Sample Test - Paired Design - Mean of Paired Differences

Data Set: Olestra.cyd

Main Advanced

Trial Type: Noninferiority Use Diff. Variable Frequency

Test Type: t-test Response Control: X

Response Treatment: Y

Noninferiority Margin: -0.5

- In the **Advanced** tab, enter 0.975 for **Confidence Level**.

Analysis: Continuous Endpoint One-Sample Test - Paired Design - Mean of Paired Differences

Data Set: Olestra.cyd

Main **Advanced**

By Variable 1: Confidence Level: 0.975

By Variable 2:

- Click **OK** to analyze the data. Upon completion of analysis, a new node with label **Analysis: Continuous Response: Difference of Means for Paired Data** will be added in the **Library** and the output will be displayed in the main

window.

Analysis: Continuous Response: Difference of Means for Paired Data

Hypothesis

If $\delta_0 < 0$ then, $H_0 : \mu_t - \mu_c \leq \delta_0$ Vs. $H_1 : \mu_t - \mu_c > \delta_0$
 If $\delta_0 > 0$ then, $H_0 : \mu_t - \mu_c \geq \delta_0$ Vs. $H_1 : \mu_t - \mu_c < \delta_0$

Input Parameters

Data File: Olestra.cyd
 Trial Type: Noninferiority
 Response Control: X
 Response Treatment: Y
 Test Type: t-test
 Noninferiority Margin (δ_0): -0.5
 Confidence Level: 0.975

Response Variable: (Y-X)
 Total Number of Records: 28
 Number of Records Rejected: 0
 Summary of the Observed Data:

| Response | Min | Max | Median | Mean | Std.Dev | n |
|----------|------|-------|--------|-------|---------|----|
| X | 4.13 | 12.45 | 6.75 | 7.195 | 2.063 | 28 |
| Y | 3.8 | 11.19 | 6.93 | 7.486 | 2.072 | 28 |

Test of Hypothesis:

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | 1-Sided | | 97.5% Confidence Interval(1-Sided) | |
|----|---------------------|------------------|------------|-------|----|---------|------|------------------------------------|-------------|
| | | | | | | p-value | Tail | Lower Limit | Upper Limit |
| 28 | 0.291 | 0.533 | 0.28 | 2.822 | 27 | 0.004 | G.E. | -0.284 | INF |

The observed value of test statistic is 2.822 and it has $28 - 1 = 27$ degrees of freedom. The lower-limit of 1- sided 97.5% confidence interval of $\delta = \mu_t - \mu_c$ is -0.284. Since this is greater than the non-inferiority margin of -0.5, we can reject $H_0 : \delta \leq \delta_0$ in favor of $H_1 : \delta > \delta_0$ at one-sided 2.5% level of significance. The p-value associated with this rejection is 0.004.

76.2 Example: Ratio of Paired Means

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of an outcome variable, X, with means μ_t and μ_c , respectively, and let σ_t^2 and σ_c^2 denote the respective variances. Let ρ_0 be the noninferiority margin. With $\rho_0 < 1$, the null hypothesis $H_0 : \mu_t/\mu_c \leq \rho_0$ is tested against the one-sided alternative hypothesis $H_1 : \mu_t/\mu_c > \rho_0$. With $\rho_0 > 1$, the null hypothesis $H_0 : \mu_t/\mu_c \geq \rho_0$ is tested against the one-sided alternative hypothesis $H_1 : \mu_t/\mu_c < \rho_0$. Let, $\rho = \mu_t/\mu_c$.

76 Analysis-Normal Noninferiority Paired-Sample

Dataset: *Olestra.cyd* as described in Section 76.1.

Purpose of the Analysis:

To test the null hypothesis $H_0: \rho \leq \rho_0$ against the alternative hypothesis $H_1: \rho > \rho_0$. For this illustrative example, we consider a non-inferiority margin (ρ_0) of 0.8.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) One Sample > (Paired Design) Mean of Paired Ratios
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below.

Analysis: Continuous Endpoint One-Sample Test - Paired Design - Mean of f

Data Set: Olestra.cyd

Main Advanced

Trial Type: Noninferiority Use Ratio

Test Type: t-test Response Control: X

Response Treatment: Y

Noninferiority Margin: 0.8

4. In the **Advanced** tab, enter 0.975 for **Confidence Level**.

Analysis: Continuous Endpoint One-Sample Test - Paired Design

Data Set: Olestra.cyd

Main **Advanced**

By Variable 1: Confidence Level:

By Variable 2:

5. Click **OK** to analyze the data. Upon completion of analysis following output will be displayed in the main window.

76 Analysis-Normal Noninferiority Paired-Sample

Analysis: Continuous Response: Ratio of Means for Paired Data

Hypothesis

If $\rho_0 < 1$ then $H_0 : \mu_t / \mu_c \leq \rho_0$ Vs. $H_1 : \mu_t / \mu_c > \rho_0$

If $\rho_0 > 1$ then $H_0 : \mu_t / \mu_c \geq \rho_0$ Vs. $H_1 : \mu_t / \mu_c < \rho_0$

where ρ_0 is the non inferiority margin

Input Parameters

Data File: Olestra.cyd
 Trial Type: Noninferiority
 Response Control: X
 Response Treatment: Y
 Test Type: t-test
 Noninferiority Margin (ρ_0): 0.8
 Confidence Level: 0.975

Output

Response variable: (Y/X)
 Total no. of records: 28
 No. of records rejected: 0
 Summary of the Observed Data:

| Response | Min | Max | Median | Mean | Std.Dev | n |
|----------|------|-------|--------|-------|---------|----|
| X | 4.13 | 12.45 | 6.75 | 7.195 | 2.063 | 28 |
| Y | 3.8 | 11.19 | 6.93 | 7.486 | 2.072 | 28 |

Test of Hypothesis for:

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | 1-Sided | |
|----|---------------------|------------------|------------|-------|----|----------|------|
| | | | | | | p-value | Tail |
| 28 | 0.038 | 1.233 | 0.04 | 6.526 | 27 | 2.679E-7 | G.E. |

Parameter Estimates:

| Metric | Estimate | Std. Error | 97.5% Confidence Interval(1-Sided) | |
|-----------|----------|------------|------------------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| ln(Ratio) | 0.038 | 0.04 | -0.044 | INF |
| Ratio | 1.038 | 0.041 | 0.957 | INF |

The observed value of test statistic is 6.526 and it has $28 - 1 = 27$ degrees of freedom. The lower limit of one-sided 97.5% confidence interval for $\rho = \mu_y / \mu_x$ is 0.957. This is greater than the non-inferiority margin $\rho_0 = 0.8$. Therefore, we can reject $H_1 : \rho \leq 0.9$ in favor of $H_1 : \rho > 0.8$. The p-value associated with this rejection is very close to 0.

77 Analysis-Normal Equivalence Paired-Sample

This section demonstrates how **East** can be used to perform inference on data collected from a paired-sample equivalence study. Independent sample experimental design in some applications (e.g., bioanalytical cross-validation study) may confound statistical tests because of a possible large pooled variance that is actually due to the intersample variability, especially for incurred biological samples obtained from clinical or animal studies (Feng et. al., 2006). This problem can be overcome by applying paired sample analysis. Two common applications of paired sample designs are as follows:

- Comparison of two treatments using subjects who are matched by demographic and baseline characteristics.
- Two observations are made from the same subject under different experimental conditions.

Chapter 9 deals with design, and simulation of these types of equivalence trials. The type of endpoint for paired equivalence design could be the difference of means or ratio of means.

Analysis based on Paired Difference of Means as endpoint is presented in section 77.1 and the Ratio of Paired Means is discussed in section 77.2.

77.1 Example: Mean of Paired Differences

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of an outcome variable, X, with means μ_t and μ_c , respectively, and with a standard deviation of paired difference as σ_D^2 . Let δ_L and δ_U be the lower and upper equivalence limits respectively. We wish to test the hypothesis

$$H_0 : \mu_t - \mu_c \leq \delta_L$$

or

$$\mu_t - \mu_c \geq \delta_U$$

against

77 Analysis-Normal Equivalence Paired-Sample

$$H_1 : \delta_L < \mu_t - \mu_c < \delta_U$$

Dataset: FengData.cyx

Data Description:

Feng et al. (2006) reported the data on 12 quality control (QC) samples. Each sample were analyzed first by Lab1 and then by Lab2. The value in the columns Lab1 and Lab2 represent the measured concentration (in $\mu\text{g ML}^{-1}$) by Lab1 and Lab2.

Purpose of the Analysis:

To ensure that comparable results can be achieved between two laboratories Lab1 and Lab2, in other words to establish statistical equivalence between the two laboratories. In this example, we consider Lab1 as the standard laboratory (C) and Lab2 as the one, which needs to be, validated (T). Denote the mean concentrations from Lab1 and Lab2 by μ_c and μ_t . Considering an equivalence limit of $(-10, 10)$ we can state our hypothesis for test as

$$H_0 : \mu_t - \mu_c \leq -10$$

(or)

$$H_0 : \mu_t - \mu_c \geq 10$$

against

$$H_1 : -10 < \mu_t - \mu_c < 10$$

To reject H_0 with type I error rate not exceeding 0.025.

Analysis Steps:

1. Open the dataset from **Samples** folder. .
2. Choose the menu item:
Analysis > (Continuous) One Sample > (Paired Design) Mean of Paired Differences

3. In the ensuing dialog box (under the **Main** tab) choose the variables as shown below.

Analysis: Continuous Endpoint One-Sample Test - Paired Design - Mean

Data Set: QCSample

Main Advanced

Trial Type: Equivalence Use Diff. Variable

Test Type: t-test Response Control: Lab1

Response Treatment: Lab2

Lower Equiv. Limit: -10

Upper Equiv. Limit: 10

4. In the **Advanced** tab, enter 0.975 for **Confidence Level**.

Analysis: Continuous Endpoint One-Sample Test - Paired Design - Mean of Paired Differences

Data Set: QCSample

Main Advanced

By Variable 1: Confidence Level: 0.975

By Variable 2: Confidence Level (Equivalence): 0.95

5. Click **OK** to analyze the data. Following output will be displayed in the main window.

Analysis: Continuous Response: Difference of Means for Paired Data

Hypothesis

$H_{01} : \mu_t - \mu_c \leq \delta_L$ Or $H_{02} : \mu_t - \mu_c \geq \delta_U$

Vs.

$H_1 : \delta_L < \mu_t - \mu_c < \delta_U$

Input Parameters

Data File: FengData.cydx
 Trial Type: Equivalence
 Response Control: Lab1
 Response Treatment: Lab2
 Test Type: t-test
 Lower Equivalence Limit (δ_L): -10
 Upper Equivalence Limit (δ_U): 10
 Confidence Level (for Equivalence): 0.9

77 Analysis-Normal Equivalence Paired-Sample

Response Variable: (Lab2-Lab1)
 Total Number of Records: 12
 Number of Records Rejected: 0
 Summary of the Observed Data:

| Response | Min | Max | Median | Mean | Std.Dev | n |
|----------|------|------|--------|--------|---------|----|
| Lab1 | 83.7 | 102 | 93.5 | 94.242 | 6.136 | 12 |
| Lab2 | 81.3 | 95.2 | 91.25 | 89.883 | 4.825 | 12 |

Test of Hypothesis:

| n | Difference of Means | Std. Effect Size | | Std. Error | t | | DF | 1-Sided | 2-Sided | 90% Confidence Interval(2-Side | |
|----|---------------------|------------------|-----------|------------|---------|---------|----|-------------|-------------|--------------------------------|-------|
| | | Under H01 | Under H02 | | p-value | p-value | | Lower Limit | Upper Limit | | |
| 12 | -4.358 | 0.69 | -1.756 | 2.36 | 2.39 | -6.084 | 11 | 0.018 | 3.959E-5 | -8.597 | -0.12 |

The observed values of two test statistics are 2.39 and -6.084, and both of them have $12 - 1 = 11$ degrees of freedom. The 2-sided 95% confidence interval for $\delta = \mu_t - \mu_c$ is (-9.553, 0.836). This confidence interval is within the equivalence interval of (-10, 10), therefore, we can reject $H_0: \mu_t - \mu_c \leq -10$ or $\mu_t - \mu_c \geq 10$ in favor of $H_1: -10 < \mu_t - \mu_c < 10$ at 2.5% level of significance.

77.2 Example: Ratio of Paired Means

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of an outcome variable, X, with means μ_t and μ_c , and let σ_t^2 and σ_c^2 denote the respective variances. Here, the null hypothesis $H_0: \mu_t/\mu_c \leq \rho_L$ or $\mu_t/\mu_c \geq \rho_U$ is tested against the alternative hypothesis $H_1: \rho_L < \mu_t/\mu_c < \rho_U$. Let $\rho = \mu_t/\mu_c$ denotes the ratio of two means. Then the null hypothesis can be expressed as $H_0: \rho \leq \rho_L$ or $\rho \geq \rho_U$ and the alternative can be expressed as $H_1: \rho_L < \rho < \rho_U$. In practice, ρ_L and ρ_U are often chosen such that $\rho_L = 1/\rho_U$. The two one-sided tests (TOST) procedure of Schuirmann (1987) is commonly used for this analysis, and is employed in this section for a parallel-group study.

We can perform the test for difference as discussed in section 77.1 on the log-transformed data.

77.2.1 Example: Ratio of Paired Means

Dataset:FengData.cyd as described in section 77.1.

Purpose of the Analysis:

To test

$$H_0: \mu_t/\mu_c \leq 0.85 \text{ or } \mu_t/\mu_c \geq 1.15 \text{ against } H_1: 0.85 < \mu_t/\mu_c < 1.15$$

We want to reject H_0 with probability of type I error not exceeding 0.025.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) One Sample > (Paired Design) Mean of Paired Ratios
3. In the ensuing dialog box (under the **Main** tab) choose the variables as shown below.

Analysis: Continuous Endpoint One-Sample Test - Paired Design

Data Set: QCSample

Main Advanced

Trial Type: Equivalence Use Ratio

Test Type: t-test Response Control: Lab1

Response Treatment: Lab2

Lower Equiv. Limit: 0.85

Upper Equiv. Limit: 1.15

4. In the **Advanced** tab, enter 0.975 for **Confidence Level**.

Analysis: Continuous Endpoint One-Sample Test - Paired Design - Mean of Paired Ratios

Data Set: QCSample

Main Advanced

By Variable 1: Confidence Level: 0.975

By Variable 2: Confidence Level (Equivalence): 0.95

5. Click **OK** to analyze the data. Following output will be displayed in the main

77 Analysis-Normal Equivalence Paired-Sample

window.

Analysis: Continuous Response: Ratio of Means for Paired Data

Hypothesis

$H_{01} : \mu_t / \mu_c \leq \rho_L$ Or $H_{02} : \mu_t / \mu_c \geq \rho_U$

Vs.

$H_1 : \rho_L < \mu_t / \mu_c < \rho_U$

Input Parameters

Data File: FengData.cydx
 Trial Type: Equivalence
 Response Control: Lab1
 Response Treatment: Lab2
 Test Type: t-test
 Lower Equivalence Limit (ρ_L): 0.85
 Upper Equivalence Limit (ρ_U): 1.15
 Confidence Level (for Equivalence): 0.95

Response Variable: (Lab2/Lab1)

Total Number of Records: 12

Number of Records Rejected: 0

Summary of the Observed Data:

| Response | Min | Max | Median | Mean | Std.Dev | n |
|----------|------|------|--------|--------|---------|----|
| Lab1 | 83.7 | 102 | 93.5 | 94.242 | 6.136 | 12 |
| Lab2 | 81.3 | 95.2 | 91.25 | 89.883 | 4.825 | 12 |

Test of Hypothesis for: ln(Lab2/Lab1)

| n | Difference of Means | Std. Effect Size | | Std. Error | t | | DF | p-value | |
|----|---------------------|------------------|-----------|------------|-----------|-----------|----|-----------|-----------|
| | | Under H01 | Under H02 | | Under H01 | Under H02 | | Under H01 | Under H02 |
| 12 | -0.047 | 1.308 | -2.107 | 0.026 | 4.53 | -7.298 | 11 | 4.286E-4 | 7.738E-6 |

Parameter Estimates:

| Metric | Estimate | Std. Error | 95% Confidence Interval(2-Sided) | |
|-----------|----------|------------|----------------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| ln(Ratio) | -0.047 | 0.026 | -0.103 | 0.01 |
| Ratio | 0.954 | 0.024 | 0.902 | 1.01 |

The observed values of two test statistics are 4.53 and -7.298 and both of them have $12 - 1 = 11$ degrees of freedom. The 2-sided 95% confidence interval of $\rho = \mu_t / \mu_c$ is (0.902, 1.01). This confidence interval is within the equivalence interval of (0.85, 1.15), therefore, we can reject $H_0 : \mu_t / \mu_c \leq 0.85$ or $\mu_t / \mu_c \geq 1.15$ in favor of $H_1 : 0.85 < \mu_t / \mu_c < 1.15$ with 2.5% level of significance.

78 *Analysis-Normal Superiority Two-Sample*

To demonstrate the superiority of a new treatment over the control, it is often necessary to randomize subjects to the control and treatment arms, and contrast the group-dependent means of the outcome variables.

In chapter 10, designing, simulation and interim monitoring of such kind of trials are discussed in details.

In this chapter, we explore how we can use **East** to analyze data that comes from two independent samples and crossover superiority studies.

78.1 *Example: Difference of Means*

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of a normally distributed outcome variable, X , with means μ_t and μ_c , respectively, and with a common variance σ^2 .

Define the treatment difference to be $\delta = \mu_t - \mu_c$. The null hypothesis $H_0: \delta = 0$ is tested against the two-sided alternative hypothesis $H_1: \delta \neq 0$ or a one-sided alternative hypothesis $H_1: \delta < 0$ or $H_1: \delta > 0$.

Dataset: Myeloma.cyd as described in section 72.1.1

Purpose of the Analysis:

To compare the mean **haemoglobin** level between two groups indicated by the variable **status** (0-alive, 1-dead).

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Two Samples > (Parallel Design) Difference of Means

78 Analysis-Normal Superiority Two-Sample

3. In the ensuing dialog box choose the variables as shown below:

Analysis: Continuous Endpoint Two-Sample Test - Parallel Design - Difference of Means

Data Set: Myeloma.cyd

Main Advanced

Trial Type: Superiority Test Type: t-test

Variance Type: Equal

Population Id: status Response Variable: haemoglobin

Control: 0

Treatment: 1

4. Click **OK** to analyze the data. Following output will be displayed in the main window.

Analysis: Continuous Response: Difference of Means for Independent Data

$H_0: \mu_1 - \mu_0 = 0$ Vs. $H_1: \mu_1 - \mu_0 \neq 0$ for 2-Sided test
 Either $H_1: \mu_1 - \mu_0 < 0$
 Or $H_1: \mu_1 - \mu_0 > 0$ for 1-Sided test

Input Parameters

Data File: Myeloma.cyd
 Trial Type: Superiority
 Population Id: status(Treatment=0, Control=1)
 Response Variable: haemoglobin
 Variance Type: Equal
 Test Type: t-test
 Confidence Level: 0.95

Output

Response Variable: haemoglobin
 Total Number of Records: 65
 Number of Records Rejected: 0
 Summary of the Observed Data:

| Treatment Group | Min | Max | Median | Mean | Std.Dev | n |
|-----------------|-----|------|--------|--------|---------|----|
| 1 | 5 | 14.6 | 10.1 | 9.91 | 2.564 | 48 |
| 0 | 4.9 | 14 | 11.6 | 11.024 | 2.425 | 17 |

Test of Hypothesis:

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | 1-Sided | | 2-Sided | | 95% Confidence Interval(2-Sided) | |
|----|---------------------|------------------|------------|-------|----|---------|------|---------|-------------|----------------------------------|--|
| | | | | | | p-value | Tail | p-value | Lower Limit | Upper Limit | |
| 65 | 1.113 | 0.44 | 0.714 | 1.559 | 63 | 0.062 | G.E. | 0.124 | -0.313 | 2.54 | |

The observed value of test statistic is 1.559 and it has $48 + 17 - 2$ or 63 degrees of

freedom. The p-value for two-sided test is 0.124. This is the p-value associated with rejecting $H_0: \delta = 0$ in favor of alternative hypothesis $H_1: \delta \neq 0$. The p-value for right tailed test is 0.062. This p-value is associated with the rejection of $H_0: \delta \leq 0$ in favor of the alternative hypothesis $H_1: \delta > 0$. **East** displays the p-value associated with right tailed test on this occasion because $\hat{\delta} > 0$. The two-sided 95% confidence interval is (-0.313, 2.54). Since the 2-sided confidence interval includes 0 or the p-value for two-sided test is 0.124, we cannot reject $H_0: \delta = 0$ at 5% level of significance.

78.2 Example: Ratio of Means

The statistical analysis regarding the ratio of means of two independent log-normal distributions is often of interest in biomedical research. Ratio of means as endpoint should be preferred when underlying distribution is skewed and therefore a lognormal distribution is a better fit than normal. Sometimes goal of the experiment can be better represented using ratio of means instead of their difference.

Let μ_t and μ_c denote the means of the observations from the experimental treatment (T) and the control treatment (C), respectively, and let σ_t^2 and σ_c^2 denote the corresponding variances of the observations. It is assumed that $\sigma_t/\mu_t = \sigma_c/\mu_c$, i.e. the coefficient of variation $CV = \sigma/\mu$ is the same for T and C .

Define the treatment ratio to be $\rho = \mu_t/\mu_c$. The null hypothesis $H_0: \rho = 1$ is tested against the two-sided alternative hypothesis $H_1: \rho \neq 1$ or a one-sided alternative hypothesis $H_1: \rho < 1$ or $H_1: \rho > 1$.

Dataset: Myeloma.cyd as described in section 72.1.1.

Purpose of the Analysis:

To compare the mean **haemoglobin** level between two groups indicated by the variable **status** (0-alive, 1-dead).

Here, we are interested in testing the null hypothesis $H_0: \rho = 1$ against the alternative hypothesis $H_1: \rho > 1$.

Since we can translate the ratio hypothesis into difference hypothesis using log transformation, **East** performs the test for difference on log-transformed data as discussed in section 78.1 to draw inference on ρ .

78 Analysis-Normal Superiority Two-Sample

Analysis Steps:

1. Choose the menu item:
Analysis > (Continuous) Two Samples > (Parallel Design) Ratio of Means
2. In the ensuing dialog box (under the **Main**) select the variables as shown below:

The screenshot shows a software dialog box titled "Analysis: Continuous Endpoint Two-Sample Test - Parallel Design - Ratio of Means". The data set is "Myeloma.cyd". The "Main" tab is selected. The settings are as follows:

| | | | |
|----------------|-------------|--------------------|-------------|
| Trial Type: | Superiority | Test Type: | t-test |
| Variance Type: | Equal | Response Variable: | haemoglobin |
| Population Id: | status | Control: | 1 |
| | | Treatment: | 0 |

3. Click **OK** to analyze the data. Following output will be displayed in the main

window.

Analysis: Continuous Response: Ratio of Means for Independent Data

Hypothesis

$H_0: \mu_t/\mu_c = 1$ Vs. $H_1: \mu_t/\mu_c \neq 1$ for 2-Sided test

Either $H_1: \mu_t/\mu_c < 1$

Or $H_1: \mu_t/\mu_c > 1$ for 1-Sided test

Input Parameters

Data File: Myeloma.cyd
 Trial Type: Superiority
 Population Id: status(Treatment=0, Control=1)
 Response Variable: haemoglobin
 Variance Type: Equal
 Test Type: t-test
 Confidence Level: 0.95

Output

Response Variable: haemoglobin
 Total Number of Records: 65
 Number of Records Rejected: 0
 Summary of the Observed Data:

| Treatment Group | Min | Max | Median | Mean | Std.Dev | n |
|-----------------|-----|------|--------|--------|---------|----|
| 1 | 5 | 14.6 | 10.1 | 9.91 | 2.564 | 48 |
| 0 | 4.9 | 14 | 11.6 | 11.024 | 2.425 | 17 |

Test of Hypothesis for: ln(haemoglobin)

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | 1-Sided | | 2-Sided |
|----|---------------------|------------------|------------|-------|----|---------|------|---------|
| | | | | | | p-value | Tail | p-value |
| 65 | 0.114 | 0.41 | 0.078 | 1.453 | 63 | 0.076 | G.E. | 0.151 |

Parameter Estimates:

| Metric | Estimate | Std. Error | 95% Confidence Interval(2-Sided) | |
|-----------|----------|------------|----------------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| ln(Ratio) | 0.114 | 0.078 | -0.043 | 0.27 |
| Ratio | 1.121 | 0.088 | 0.958 | 1.311 |

The observed value of test statistic is 1.453 and it has $48 + 17 - 2 = 63$ degrees of freedom. The two-sided 95% confidence interval for $\ln \rho$ is (-0.043, 0.27) and for ρ is (0.958, 1.311). The former confidence interval includes 0 and the latter includes 1. Therefore, we cannot reject $H_0: \rho = 1$ in favor of $H_1: \rho \neq 1$. The p-value for comparing $H_0: \rho \leq 1$ in favor of $H_1: \rho > 1$ is 0.076. Therefore, we cannot reject

78 Analysis-Normal Superiority Two-Sample

$H_0: \rho \leq 1$ against $H_1: \rho > 1$ either at 5% level of significance.

78.3 Example: Difference of Means in Crossover design

In a 2×2 crossover design each subject is randomized to one of two sequence groups. Subjects in the sequence group 1 receive the test drug (T) formulation in the first period, have their outcome variable, X , recorded, wait out a washout period to ensure that the drug is cleared from their system, then receive the control drug formulation (C) in period 2 and finally have the measurement on X again. In sequence group 2, the order in which the T and C are assigned is reversed. The table below summarizes this type of trial design.

| Group | Period 1 | Washout | Period 2 |
|-------|----------|---------|----------|
| 1(TC) | Test | — | Control |
| 2(CT) | Control | — | Test |

The resulting data are commonly analyzed using a statistical linear model. The response y_{ijk} in period j on subject k in sequence group i , where $i = 1, 2$, $j = 1, 2$, and $k = 1, \dots, n_i$ is modeled as a linear function of an overall mean response μ , formulation effect τ_t and τ_c , period effects π_1 and π_2 , and sequence effects λ_1 and λ_2 . The fixed effects model can be displayed as:

| Group | Period 1 | Washout | Period 2 |
|-------|-----------------------------------|---------|------------------------------------|
| 1(TC) | $\mu + \tau_t + \pi_1 + \gamma_1$ | — | $\mu + \tau_c + \pi_2 + \lambda_1$ |
| 2(CT) | $\mu + \tau_c + \pi_1 + \gamma_2$ | — | $\mu + \tau_t + \pi_2 + \lambda_2$ |

For superiority trial, **East** can test following null hypotheses:

- Test1: $H_0 : \tau_t - \tau_c = 0$. for treatment effect
- Test2: $H_0 : \pi_1 - \pi_2 = 0$. for period effect
- Test1: $H_0 : \lambda_1 - \lambda_2 = 0$. For carryover effect

Dataset: CrossoverCaseData.cyd

Data Description:

Jones and Kenward (2003) presented data from a 2×2 crossover trial where the primary objective was to evaluate the efficacy and safety of an inhaled drug given to patients with chronic obstructive pulmonary disease. Eligible patients were randomized to either treatment sequence AB or BA (A: Drug; B=Placebo). There was

4 weeks of gap between two periods. The main comparison of efficacy was based on the mean morning peak expiratory flow rate (PEFR). The data of this trial are available in [CrossoverCaseData.cyd](#).

This dataset contains 112 observations and 7 variables. The columns **GroupID**, **PeriodID** and **subjectID** contain the information about group sequence, period and subject id, respectively. The column **Response** contains the measurements on the PEFR.

Purpose of the Analysis:

To test if there is significant carryover effect from period 1 to period 2.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Two Samples > (Crossover Design) Difference of Means
3. In the ensuing dialog box, there are two tabs in this window – **Main** and **Advanced**. In the **Main** tab, select different variables as shown below:

Analysis: Continuous Endpoint Two-Sample Test - Crossover Design - Difference of Means

Data Set: CrossoverCaseData.cyd

Main Advanced

Trial Type: Superiority

Effect Type: Carryover

Treatment Specifications

Treatment 1: Treatment 1

Treatment 2: Treatment 2

Period ID: PeriodID

Subject ID: subjectID

Group ID: GroupID

Response Variable: Respon

4. Click **OK** to start analysis. Upon completion of analysis following output will be displayed in the main window.

78 Analysis-Normal Superiority Two-Sample

Analysis: Continuous Response: Difference of Means test for Crossover Data

Hypothesis

$H_0 : \lambda_1 - \lambda_2 = 0$ Vs. $H_1 : \lambda_1 - \lambda_2 \neq 0$ for 2-Sided test

Either $H_1 : \lambda_1 - \lambda_2 < 0$

Or $H_1 : \lambda_1 - \lambda_2 > 0$ for 1-Sided test

Data File: CrossoverCaseData.cyd
Trial Type: Superiority
Response Variable: Response
Group ID: GroupID
Subject ID: subjectID
Period ID: PeriodID
Test Type: t-Test
Effect Type: Carryover
Confidence Level: 0.95
Treatment Assignment:

| Group ID | Period ID | |
|----------|-------------|-------------|
| | 1 | 2 |
| 1 | Treatment 1 | Treatment 2 |
| 2 | Treatment 2 | Treatment 1 |

Response Variable: (Response)

Total Number of Records: 112

Number of Records Rejected: 0

Summary of the Observed Data:

| Group ID | Statistic | Period ID | | Group Summary |
|----------------|-----------------|-----------|---------|---------------|
| | | 1 | 2 | |
| 1 | No. of Subjects | 27 | 27 | 27 |
| | Mean | 245.839 | 239.203 | 242.621 |
| | Std. Dev. | 82.78 | 81.697 | 81.529 |
| | Median | 235 | 221.842 | 228.421 |
| | Minimum | 67.778 | 70.278 | 67.778 |
| 2 | No. of Subjects | 29 | 29 | 29 |
| | Mean | 215.992 | 230.162 | 223.077 |
| | Std. Dev. | 72.629 | 73.942 | 72.993 |
| | Median | 201.905 | 225 | 212.068 |
| | Minimum | 104.444 | 128.947 | 104.444 |
| Period Summary | No. of Subjects | 56 | 56 | |
| | Mean | 230.382 | 234.521 | |
| | Std. Dev. | 78.43 | 77.197 | |
| | Median | 225.595 | 222.005 | |
| | Minimum | 67.778 | 70.278 | |
| | Maximum | 443.25 | 420.5 | |

Treatment Summary:

| Statistic | Treatment 1 | Treatment 2 |
|-----------------|-------------|-------------|
| No. of Subjects | 56 | 56 |
| Mean | 237.72 | 227.183 |
| Std. Dev. | 78.008 | 77.315 |
| Median | 231.157 | 214.272 |
| Minimum | 67.778 | 70.278 |
| Maximum | 443.25 | 420.5 |

Inference for: (Response)

Test of Hypothesis for:

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | (1-Sided) | | (2-Sided) | 95% Confidence Interval(2-Sided) | |
|----|---------------------|------------------|------------|-------|----|-----------|------|-----------|----------------------------------|-------------|
| | | | | | | p-value | Tail | p-value | Lower Limit | Upper Limit |
| 56 | 38.888 | 0.254 | 41.008 | 0.948 | 54 | 0.174 | G.E. | 0.347 | -43.328 | 121.105 |

The observed value of test statistic is 0.948 and it has $27 + 24 - 2 = 54$ degrees of freedom. The p-value for two sided test is 0.347. Therefore, the carryover effect is not significant at 5% level of significance.

Let us further examine if there is a significant **Treatment effect**:

78 Analysis-Normal Superiority Two-Sample

Dataset: CrossoverCaseData.cyd as described in Section 78.3.

Purpose of the Analysis:

To test if there is significant treatment effect.

Analysis Steps:

1. Open the dataset.
2. Choose the menu item:
Analysis > (Continuous) Two Samples > (Crossover Design) Difference of Means
3. In the ensuing dialog box, complete the **Main** tab as before except for **Effect Type**. Select **Treatment** as **Effect Type** as we are interested in testing the treatment effect.

Analysis: Continuous Endpoint Two-Sample Test - Crossover Design - Difference of Means

Data Set: CrossoverCaseData.cyd

Main Advanced

Trial Type: Superiority Period ID: PeriodID Response Variable: Response

Effect Type: Treatment Subject ID: subjectID

Treatment Specifications

Treatment 1: Treatment 1 Group ID: GroupID

Treatment 2: Treatment 2

4. Click **OK** to start analysis. Upon completion of analysis, a new node with label **Analysis: Continuous Response: Difference of Means test for Crossover Data2** will be added to the **Library** and the output will be displayed in the main window. Scroll down to the end of the output. Output for statistical test of treatment effect is displayed in the last table.

Treatment Summary:

| Statistic | Treatment 1 | Treatment 2 |
|-----------------|-------------|-------------|
| No. of Subjects | 56 | 56 |
| Mean | 237.72 | 227.183 |
| Std. Dev. | 78.008 | 77.315 |
| Median | 231.157 | 214.272 |
| Minimum | 67.778 | 70.278 |
| Maximum | 443.25 | 420.5 |

Inference for: (Response)
Test of Hypothesis for:

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | (1-Sided) | | (2-Sided) | | 95% Confidence Interval(2-Sided) | |
|----|---------------------|------------------|------------|-------|----|-----------|------|-----------|-------------|----------------------------------|--|
| | | | | | | p-value | Tail | p-value | Lower Limit | Upper Limit | |
| 56 | 10.403 | 0.407 | 3.416 | 3.046 | 54 | 0.002 | G.E. | 0.004 | 3.555 | 17.25 | |

The observed value of test statistic is 3.046 and it has $27 + 24 - 2 = 54$ degrees of freedom. The p-value for two sided test is 0.004. Therefore, the treatment effect is significant at 5% level of significance.

78.4 Example: Ratio of Means in Crossover design

In this chapter, we show how we can use **East** to test for ratio of means from a superiority 2×2 crossover trial. We have already discussed 2×2 crossover design in section 78.3. However, unlike section 78.3, here we are interested in ratio of means. Let μ_t and μ_c denote the means of the observations from the experimental treatment (T) and the control treatment (C), respectively. **East** can test following null hypotheses:

- Test1: $H_0 : \mu_t / \mu_c = 1$. For treatment effect
- Test2: $H_0 : \pi_1 / \pi_2 = 1$. For period effect
- Test1: $H_0 : \lambda_1 / \lambda_2 = 1$. For carryover effect

Since we can translate the ratio hypothesis into difference hypothesis using log transformation, **East** performs the test for difference on log-transformed data as discussed in section 78.3.

Dataset: CrossoverCaseData.cyd as described in section 78.3.

Purpose of the Analysis:

To test the null hypothesis $H_0: \rho = 1$ against the alternative hypothesis $H_1: \rho \neq 1$.

Analysis Steps:

1. Choose the menu item:
Analysis > (Continuous) Two Samples > (Crossover Design) Ratio of Means
2. In the ensuing dialog box, select different variables as shown below:

The screenshot shows the 'Analysis: Continuous Endpoint, Two-Sample Test - Crossover Design - Ratio of Means' dialog box. The 'Data Set' is 'CrossoverCaseData.cyd'. The 'Main' tab is selected. The settings are as follows:

- Trial Type: Superiority
- Effect Type: Carryover
- Treatment Specifications: Treatment 1, Treatment 2
- Period ID: PeriodID
- Subject ID: subjectID
- Group ID: GroupID
- Response Variable: Response

78 Analysis-Normal Superiority Two-Sample

- Click **OK** to start analysis. Upon completion of analysis, a new node with label **Analysis: Continuous Response: Ratio of Means for Crossover Data1** will be added to the **Library** and the output will be displayed in the main window. Scroll down to the end of the output. Output for statistical test of treatment effect is displayed in the last table.

| Treatment Summary: | | | | | | | | | |
|--|---------------------|------------------|----------------------------------|-------------|----|-----------|------|-----------|--|
| Statistic | | Treatment 1 | Treatment 2 | | | | | | |
| No. of Subjects | | 56 | 56 | | | | | | |
| Mean | | 237.72 | 227.183 | | | | | | |
| Std. Dev. | | 78.008 | 77.315 | | | | | | |
| Median | | 231.157 | 214.272 | | | | | | |
| Minimum | | 67.778 | 70.278 | | | | | | |
| Maximum | | 443.25 | 420.5 | | | | | | |
| Inference for: ln(Response) | | | | | | | | | |
| Test of Hypothesis for: | | | | | | | | | |
| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | (1-Sided) | | (2-Sided) | |
| | | | | | | p-value | Tail | p-value | |
| 56 | 0.046 | 0.388 | 0.016 | 2.904 | 54 | 0.003 | G.E. | 0.005 | |
| Parameter Estimates: | | | | | | | | | |
| Metric | Estimate | Std. Error | 95% Confidence Interval(2-Sided) | | | | | | |
| | | | Lower Limit | Upper Limit | | | | | |
| ln(Ratio) | 0.046 | 0.016 | 0.014 | 0.078 | | | | | |
| Ratio | 1.047 | 0.017 | 1.014 | 1.081 | | | | | |

East performs the analysis based on the log-transformed data. The observed value of test statistic based on log-transformed data is 2.904 and it has $27 + 24 - 2 = 54$ degrees of freedom. The p-value for two sided test is 0.005. Therefore, the treatment effect is significant at 5% level of significance.

Now we will perform the test for difference of means for crossover data based on log-transformed data. The **CrossoverCaseData.cyd** has a column labeled as **LnResp** which contains the log-transformed values of the entries in the *Response* column. The result for test of treatment effect based on **LnResp** as response variable (using

difference of means for crossover data) is as follows:

| Treatment Summary: | | | | | | | | | | | |
|-------------------------|---------------------|------------------|------------|-------|----|-----------|------|-----------|----------------------------------|-------------|--|
| Statistic | Treatment 1 | Treatment 2 | | | | | | | | | |
| No. of Subjects | 56 | 56 | | | | | | | | | |
| Mean | 5.416 | 5.369 | | | | | | | | | |
| Std. Dev. | 0.347 | 0.346 | | | | | | | | | |
| Median | 5.443 | 5.367 | | | | | | | | | |
| Minimum | 4.216 | 4.252 | | | | | | | | | |
| Maximum | 6.094 | 6.041 | | | | | | | | | |
| Inference for: (LnResp) | | | | | | | | | | | |
| Test of Hypothesis for: | | | | | | | | | | | |
| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | (1-Sided) | | (2-Sided) | 95% Confidence Interval(2-Sided) | | |
| | | | | | | p-value | Tail | p-value | Lower Limit | Upper Limit | |
| 56 | 0.046 | 0.388 | 0.016 | 2.904 | 54 | 0.003 | G.E. | 0.005 | 0.014 | 0.078 | |

Compare the value of observed test statistic and the p-values with those from test for ratio of crossover means. They are identical. This is because the test for ratio of crossover means in **East** is equivalent to test for difference of crossover means based on log-transformed data.

79 Analysis-Normal Noninferiority Two-Sample

In a noninferiority trial, the goal is to establish that an experimental treatment is **no worse than** the standard treatment, rather than attempting to establish that it is superior. A therapy that is demonstrated to be non-inferior to the current standard therapy for a particular indication might be an acceptable alternative if, for instance, it is easier to administer, cheaper, or less toxic.

In chapter 12, designing, simulation and interim monitoring of such kind of trials are discussed in details.

In this chapter, we explore how we can use **East** to perform analysis of data that comes from two independent samples and crossover noninferiority studies.

79.1 Example: Difference of Means

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of a normally distributed outcome variable, X , with means μ_t and μ_c , respectively, and with a common variance σ^2 .

Define the treatment difference to be $\delta = \mu_t - \mu_c$ and δ_0 be the non inferiority margin. When $\delta_0 < 0$, **East** tests the null hypothesis $H_0: \delta \leq \delta_0$ against the alternative hypothesis $H_1: \delta > \delta_0$. When $\delta_0 > 0$, the null hypothesis $H_0: \delta \geq \delta_0$ is tested against the alternative hypothesis $H_1: \delta < \delta_0$.

Let \bar{X}_t and \bar{X}_c be the mean responses of the experimental and control groups, respectively, based on n_t observations from T and n_c observations from C. Then the estimate of δ is $\hat{\delta} = \bar{X}_t - \bar{X}_c$. Test statistic can be defined as

$$Z = \frac{\hat{\delta} - \delta_0}{se(\hat{\delta})} \tag{79.1}$$

where $se(\hat{\delta})$ is the standard error of $\hat{\delta}$ based on $n_t + n_c$ observations. Z is distributed as variable that follows t distribution with $n_t + n_c - 2$ degrees of freedom or standard normal variate.

Dataset: `Werner.cyd` as described in section 73.4.2.

Purpose of the Analysis:

The purpose here is to compare the mean cholesterol levels between the birthpill users and nonusers. Let μ_t and μ_c be the mean cholesterol level in birthpill user group and

non-user group, respectively, and $\delta = \mu_t - \mu_c$. We want to test the null hypothesis $H_0: \delta \geq 25$ against the alternative hypothesis $H_1: \delta < 25$. For this analysis, we consider one-sided type I error rate of 0.025.

Analysis Steps:

1. To open the dataset from Samples folder
2. In case multiple workbooks are currently open, then this will bring up the **Keep in** dialog box. You can select either one of the existing workbooks or you can create new workbook. Suppose you want to create a new workbook labeled as “Birthpill Noninferiority”. In order to do this, select the radio button **New Workbook** and type in **Birthpill Noninferiority** in the field next to it.
3. Choose the menu item:
Analysis > (Continuous) Two Samples > (Parallel Design) Difference of Means
4. In the ensuing dialog box (under the **Main**) tab select **Noninferiority** as **Trial Type**, **Equal** as **Variance Type** and **t-test** as **Test Type**. Select **Birthpill** as **Population Id** variable. As you select variable for **Population Id** field, a new box will appear below where you have to specify the levels of the **Population Id** variable for control and treatment group. Choose **0** for **Control**. By doing this, **East** will treat the subjects with BIRTHPILL=0 as they are in the control group and remaining subjects in the treatment group. Select **Response Variable** as **CHOLESTEROL** and enter **25** for **Noninferiority Margin**. Leave the **Frequency Variable** field blank.

Analysis: Continuous Endpoint Two-Sample Test - Parallel Design - Differe

Data Set: Werner.cyd

Main Advanced

Trial Type: Noninferiority Test Type: t-test

Variance Type: Equal

Population Id: BIRTHPILL Response Variable: CHOLESTEROL

Control: 0 Noninferiority Margin: 25

Treatment: 1

79 Analysis-Normal Noninferiority Two-Sample

5. In the **Advanced** tab, enter 0.975 for **Confidence Level**.

Analysis: Continuous Endpoint Two-Sample Test - Parallel Design - Wilcoxon-Mann-Whitney

Data Set: Werner.cyd

Main **Advanced**

By Variable 1: Confidence Level:

By Variable 2:

6. Click **OK** to analyze the data. Following output will be displayed in the main window.

Analysis: Continuous Response: Difference of Means for Independent Data

Hypothesis

If $\delta_0 < 0$ then, $H_0: \mu_t - \mu_c \leq \delta_0$ Vs. $H_1: \mu_t - \mu_c > \delta_0$
 If $\delta_0 > 0$ then, $H_0: \mu_t - \mu_c \geq \delta_0$ Vs. $H_1: \mu_t - \mu_c < \delta_0$

Input Parameters

Data File: Werner.cyd
 Trial Type: Noninferiority
 Population Id: BIRTHPILL(Treatment=1, Control=0)
 Response Variable: CHOLESTEROL
 Variance Type: Equal
 Test Type: t-test
 Noninferiority Margin (δ_0): 25
 Confidence Level: 0.975

Output

Response Variable: CHOLESTEROL
 Total Number of Records: 188
 Number of Records Rejected: 0
 Summary of the Observed Data:

| Treatment Group | Min | Max | Median | Mean | Std.Dev | n |
|-----------------|-----|-----|--------|---------|---------|----|
| 0 | 155 | 335 | 230 | 232.968 | 43.492 | 94 |
| 1 | 50 | 600 | 235 | 240.585 | 58.924 | 94 |

Test of Hypothesis:

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | 1-Sided | | 97.5% Confidence Interval(1-Sided) | |
|-----|---------------------|------------------|------------|--------|-----|---------|------|------------------------------------|-------------|
| | | | | | | p-value | Tail | Lower Limit | Upper Limit |
| 188 | 7.617 | -0.336 | 7.554 | -2.301 | 186 | 0.011 | L.E. | -INF | 22.519 |

There are 94 observations in each group. The mean (standard deviation) cholesterol levels are 232.97 (43.492) and 240.59 (58.924) in birthpill non-user and user groups, respectively. Estimated treatment difference is $\hat{\delta} = 7.617$ with $se(\hat{\delta}) = 7.554$. The

effect size is -0.336 . This can be verified by plugging the value of $\hat{\delta} = 7.617$, $\delta_0 = 25$ and $\hat{\sigma} = 7.554/\sqrt{1/94 + 1/94} = 51.788$ in the following formula of effect size

$$\frac{\hat{\delta} - \delta_0}{\hat{\sigma}}$$

The observed value of test statistic is -2.301 (see eq. 79.1) and it has $94 + 94 - 2$ or 186 degrees of freedom. The p-value for one-sided test is 0.011. This is the p-value associated with rejecting $H_0: \delta \geq 25$ in favor of alternative hypothesis $H_1: \delta < 25$. The one-sided 97.5% confidence interval is $(-\infty, 22.519)$. Since the upper limit of the confidence interval is smaller than the noninferiority margin of 25, we can reject $H_0: \delta \geq 25$ at one-sided 2.5% level of significance.

79.2 Example: Ratio of Means

The statistical analysis regarding the ratio of means of two independent log-normal distributions is often of interest in biomedical research. Ratio of means as endpoint should be preferred when underlying distribution is skewed and therefore a lognormal distribution is a better fit than normal. Sometimes goal of the experiment can be better represented using ratio of means instead of their difference.

Let μ_t and μ_c denote the means of the observations from the experimental treatment (T) and the control treatment (C), respectively, and let σ_t^2 and σ_c^2 denote the corresponding variances of the observations. It is assumed that $\sigma_t/\mu_t = \sigma_c/\mu_c$, i.e. the coefficient of variation $CV = \sigma/\mu$ is the same for T and C . Finally, let $\rho = \mu_t/\mu_c$.

Let ρ_0 be the noninferiority margin. For $\rho_0 < 1$, **East** tests the null hypothesis $H_0: \rho \leq \rho_0$ against the alternative hypothesis $H_1: \rho > \rho_0$. When $\rho_0 > 1$, the null hypothesis $H_0: \rho \geq \rho_0$ is tested against the alternative hypothesis $H_1: \rho < \rho_0$.

Since we can translate the ratio hypothesis into difference hypothesis using log transformation, **East** performs the test for difference on log-transformed data as discussed in section 79.1 to draw inference on ρ .

Dataset: We will again use **Werner.cyd** dataset as described in section 73.4.2.

Purpose of the Analysis:

Let μ_t and μ_c be the mean cholesterol level in birthpill user and nonuser groups, respectively, and $\rho = \mu_t/\mu_c$. Here, we are interested in testing the null hypothesis $H_0: \rho \geq 1.10$ is tested against the alternative hypothesis $H_1: \rho < 1.10$. For this analysis, we consider one-sided type I error rate of 0.05.

79 Analysis-Normal Noninferiority Two-Sample

Analysis Steps:

1. Choose the menu item:
Analysis > (Continuous) Two Samples > (Parallel Design) Ratio of Means
2. If the dataset is not displayed in your main window, this will bring up the **Select Dataset** dialog box with the list of available workbooks and datasets available under each workbook. If the dataset is already displayed in your main window, **East** will skip this step and the dataset in the main window will be used in the analysis. In case **East** brings up the **Select Dataset** dialog box, choose **Werner.cyd** dataset under workbook **BirthpillNon inferiority** and click **OK**.
3. In the ensuing dialog box (under the **Main**) select the variables as shown below:

Analysis: Continuous Endpoint Two-Sample Test - Parallel Design

Data Set: Werner.cyd

Main Advanced

Trial Type: Noninferiority Test Type: t-test

Variance Type: Equal

Population Id: BIRTHPILL Response Variable: CHOLESTEROL

Control: 0 Noninferiority Margin: 1.10

Treatment: 1

- Under the **Advanced** tab enter 0.975 for **Confidence Level**.

Analysis: Continuous Endpoint Two-Sample Test - Parallel Design - Ratio of Means

Data Set: Werner.cyd

Main **Advanced**

By Variable 1: Confidence Level:

By Variable 2:

- Click **OK** to analyze the data. Following output will be displayed in the main

79 Analysis-Normal Noninferiority Two-Sample

window.

Analysis: Continuous Response: Ratio of Means for Independent Data

Hypothesis
 If $p_0 < 1$ then $H_0 : \mu_t / \mu_c \leq p_0$ Vs. $H_1 : \mu_t / \mu_c > p_0$
 If $p_0 > 1$ then $H_0 : \mu_t / \mu_c \geq p_0$ Vs. $H_1 : \mu_t / \mu_c < p_0$
 where p_0 is the non inferiority margin

Input Parameters

Data File: Werner.cyd
 Trial Type: Noninferiority
 Population Id: BIRTHPILL(Treatment=1, Control=0)
 Response Variable: CHOLESTEROL
 Variance Type: Equal
 Test Type: t-test
 Noninferiority Margin (p_0): 1.1
 Confidence Level: 0.975

Output

Response variable: (CHOLESTEROL)
 Total no. of records: 188
 No. of records rejected: 0

Summary of the Observed Data:

| Treatment Group | Min | Max | Median | Mean | Std.Dev | n |
|-----------------|-----|-----|--------|---------|---------|----|
| 0 | 155 | 335 | 230 | 232.968 | 43.492 | 94 |
| 1 | 50 | 600 | 235 | 240.585 | 58.924 | 94 |

Test of Hypothesis for: ln(CHOLESTEROL)

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | 1-Sided | |
|-----|---------------------|------------------|------------|------|-----|---------|------|
| | | | | | | p-value | Tail |
| 188 | 0.021 | -0.336 | 0.032 | -2.3 | 186 | 0.011 | L.E. |

Parameter Estimates:

| Metric | Estimate | Std. Error | 97.5% Confidence Interval(1-Sided) | |
|-----------|----------|------------|------------------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| ln(Ratio) | 0.021 | 0.032 | -INF | 0.085 |
| Ratio | 1.021 | 0.033 | 0 | 1.088 |

In the **Output** section, the first part provides descriptive statistics for the two groups. The second table labeled with **Test of Hypothesis for:ln(CHOLESTEROL)** provides details about the test result. Note the word “ln(CHOLESTEROL)”; this emphasize that

the analysis is performed on log-transformed data. In this table, we see the **Difference of Means** as 0.021. This is the estimated treatment difference in terms of log-transformed data on CHOLESTEROL. The estimated effect size is -0.336.

The observed value of test statistic is -2.3 and it has $94 + 94 - 2 = 186$ degrees of freedom. The one-sided 97.5% confidence interval for $\ln \rho$ is $(-\infty, 0.085)$ and for ρ is $(0, 1.088)$. The upper limit of one-sided 97.5% confidence interval for ρ is smaller than the noninferiority margin $\rho_0 = 1.10$. Therefore, we reject $H_0: \rho \geq 1.10$ in favor of $H_1: \rho < 1.10$ at one-sided 0.025 level of significance. The p-value associated with this rejection is 0.011.

79.3 Example: Difference of Means in Cross-over design

In a 2×2 crossover design each subject is randomized to one of two sequence groups. Subjects in the sequence group 1 receive the test drug (T) formulation in a first period, have their outcome variable, X recorded, wait out a washout period to ensure that the drug is cleared from their system, then receive the control drug formulation (C) in period 2 and finally have the measurement on X again. In sequence group 2, the order in which the T and C are assigned is reversed. The table below summarizes this type of trial design.

| Group | Period 1 | Washout | Period 2 |
|-------|----------|---------|----------|
| 1(TC) | Test | — | Control |
| 2(CT) | Control | — | Test |

The resulting data are commonly analyzed using a statistical linear model. The response y_{ijk} in period j on subject k in sequence group i , where $i = 1, 2$, $j = 1, 2$, and $k = 1, \dots, n_i$ is modeled as a linear function of an overall mean response μ , formulation effect τ_t and τ_c , period effects π_1 and π_2 , and sequence effects λ_1 and λ_2 . The fixed effects model can be displayed as:

| Group | Period 1 | Washout | Period 2 |
|-------|-----------------------------------|---------|------------------------------------|
| 1(TC) | $\mu + \tau_t + \pi_1 + \gamma_1$ | — | $\mu + \tau_c + \pi_2 + \lambda_1$ |
| 2(CT) | $\mu + \tau_c + \pi_1 + \gamma_2$ | — | $\mu + \tau_t + \pi_2 + \lambda_2$ |

Let $\mu_t = \mu + \tau_t$ and $\mu_c = \mu + \tau_c$. For noninferiority crossover trial, **East** tests only for treatment effect. With δ_0 as noninferiority margin, **East** tests $H_0: \mu_t - \mu_c \leq \delta_0$ when $\delta_0 < 0$ and $H_0: \mu_t - \mu_c \geq \delta_0$ when $\delta_0 > 0$.

79 Analysis-Normal Noninferiority Two-Sample

In **East** we use following test statistic to test the above null hypothesis:

$$T_L = \frac{(\bar{y}_{11} - \bar{y}_{12} - \bar{y}_{21} + \bar{y}_{22})/2 - \delta_0}{\sqrt{\frac{\hat{\sigma}^2}{2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where, \bar{y}_{ij} is the mean of the observations from group i and period j and $\hat{\sigma}^2$ is the estimate of error variance (obtained as mean-squared error from ANOVA including period, treatment and sequence as source of variation in the model). T_τ is distributed with Student's t distribution with $(n_1 + n_2 - 2)$ degrees of freedom.

Dataset: pkfood.cyd

Data Description:

Here we will use pharmacokinetic data from 2×2 crossover trial available in **pkfood.cyd**. The dataset consists of observations from 20 subjects on AUC , C_{max} and T_{max} evaluated under two regimens A and B. For this example, we will consider regimen B as reference and regimen A as test drug and AUC as response variable.

Purpose of the Analysis:

Let μ_c and μ_t denote the mean AUC in regimen B and regimen A, respectively and $\delta = \mu_t - \mu_c$. We are interested in testing $H_0: \mu_t - \mu_c \leq \delta_0$ against $H_1: \mu_t - \mu_c > \delta_0$. Here we set the noninferiority margin, δ_0 as -5000 . For this analysis, one-sided type I error of 0.025 is considered.

Analysis Steps:

1. Choose the menu item:
Home > Open > Data to open the dataset from **Samples** folder.
2. In case multiple workbooks are currently open, then this will bring up the **Keep in** dialog box. You can select either one of the existing workbooks or you can create new workbook. Suppose you want to create a new workbook labeled as "Crossover noninferiority". In order to do this, select the radio button **New Workbook** and type in **Crossover noninferiority** in the field next to it. Click **OK**. This will open the **pkfood.cyd** dataset in the main window of under the **Data Editor**.
3. Choose the menu item:
Analysis > (Continuous) Two Samples > (Crossover Design) Difference of Means

4. In the ensuing dialog box (under the **Main**) select/enter the different variables as shown below.

The screenshot shows a software dialog box titled "Analysis: Continuous Endpoint, Two-Sample Test - Crossover Design - Difference of Means". The "Data Set" is "pkfood.cyd". The "Main" tab is active. The "Trial Type" is set to "Noninferiority". The "Period ID" is "PERIOD", "Subject ID" is "SUBJECT", and "Group ID" is "SEQUENCE". The "Response Variable" is "AUC" and the "Noninferiority Margin(δ_0)" is "-5000". Under "Treatment Specifications", "Treatment 1" is "Test" and "Treatment 2" is "Reference".

5. In the **Advanced** tab, leave the fields **By Variable 1** and **By Variable 2** blank and enter 0.975 for **Confidence Level**.
6. Click **OK** to analyze the data. Following output will be displayed in the main

79 Analysis-Normal Noninferiority Two-Sample

window.

Analysis: Continuous Response: Difference of Means test for Crossover Data

Hypothesis

If $\delta_0 < 0$ then $H_0 : \mu_1 - \mu_2 \leq \delta_0$ Vs. $H_1 : \mu_1 - \mu_2 > \delta_0$
 If $\delta_0 > 0$ then $H_0 : \mu_1 - \mu_2 \geq \delta_0$ Vs. $H_1 : \mu_1 - \mu_2 < \delta_0$
 where δ_0 is the non inferiority margin

Data File: pkfood.cyd
 Trial Type: Noninferiority
 Response Variable: AUC
 Group ID: SEQUENCE
 Subject ID: SUBJECT
 Period ID: PERIOD
 Test Type: t-Test
 Noninferiority Margin(δ_0): -5000
 Confidence Level: 0.975
 Treatment Assignment:

| Group ID | Period ID | |
|----------|-----------|-----------|
| | 1 | 2 |
| AB | Test | Reference |
| BA | Reference | Test |

Response Variable: (AUC)
 Total Number of Records: 40
 Number of Records Rejected: 0

Summary of the Observed Data:

| Group ID | Statistic | Period ID | | Group Summary |
|----------------|-----------------|-----------|----------|---------------|
| | | 1 | 2 | |
| AB | No. of Subjects | 10 | 10 | 10 |
| | Mean | 8542 | 9582.9 | 9062.45 |
| | Std. Dev. | 2978.945 | 3820.535 | 3376.8 |
| | Median | 8641 | 7859.5 | 8012 |
| | Minimum | 4686 | 5648 | 4686 |
| | Maximum | 14059 | 15765 | 15765 |
| BA | No. of Subjects | 10 | 10 | 10 |
| | Mean | 10620.3 | 8369.4 | 9494.85 |
| | Std. Dev. | 3468.083 | 3313.29 | 3497.233 |
| | Median | 9852.5 | 7607.5 | 9165.5 |
| | Minimum | 5911 | 5118 | 5118 |
| | Maximum | 18864 | 16510 | 18864 |
| Period Summary | No. of Subjects | 20 | 20 | |
| | Mean | 9581.15 | 8976.15 | |
| | Std. Dev. | 3322.27 | 3535.775 | |
| | Median | 9641.5 | 7859.5 | |
| | Minimum | 4686 | 5118 | |
| | Maximum | 18864 | 16510 | |

79 Analysis-Normal Noninferiority Two-Sample

Treatment Summary:

| Statistic | Test | Reference |
|-----------------|----------|-----------|
| No. of Subjects | 20 | 20 |
| Mean | 8455.7 | 10101.6 |
| Std. Dev. | 3067.804 | 3590.908 |
| Median | 8091.5 | 9604.5 |
| Minimum | 4686 | 5648 |
| Maximum | 16510 | 18864 |

Inference for: (AUC)

Test of Hypothesis for:

| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | (1-Sided) | | 97.5% Confidence Interval(1-Sided) | |
|----|---------------------|------------------|------------|-------|----|-----------|------|------------------------------------|-------------|
| | | | | | | p-value | Tail | Lower Limit | Upper Limit |
| 20 | -1645.9 | 0.95 | 789.591 | 4.248 | 18 | 2.42E-4 | G.E. | -3304.769 | INF |

In the **Output** section, the first part provides descriptive statistics for the two groups. The second table provides the treatment summary. The third table labeled as **Test of Hypothesis for** provides results for statistical test of treatment effect. The estimated effect size is 0.95. The observed value of test statistic is 4.248 and it has $10 + 10 - 2 = 18$ degrees of freedom. The p-value for one-sided test is 0. This is the p-value associated with rejecting $H_0: \delta \leq -5000$ in favor of alternative hypothesis $H_1: \delta > -5000$. The one-sided 97.5% confidence interval is $(-3304.769, -\infty)$. Since the lower limit of the confidence interval is greater than the noninferiority margin of -5000, we can reject $H_0: \delta \leq -5000$ at one-sided 2.5% level of significance.

79.4 Example: Ratio of Means in Crossover design

In this chapter, we show how we can use **East** to test for ratio of means from a noninferiority 2×2 crossover trial. We have already discussed 2×2 crossover design in section 79.3. However, unlike section 79.3, here we are interested in ratio of means. Let μ_t and μ_c denote the means of the observations from the experimental treatment (T) and the control treatment (C), respectively. For noninferiority trial, **East** tests only for treatment effect. With ρ_0 as noninferiority margin, **East** tests $H_0: \mu_t/\mu_c \leq \rho_0$ when $\rho_0 < 1$ and $H_0: \mu_t/\mu_c \geq \rho_0$ when $\rho_0 > 1$.

Since we can translate the ratio hypothesis into difference hypothesis using log transformation, **East** performs the test for difference on log-transformed data as discussed in section 79.3.

Dataset: We will again use **pkfood.cyd** dataset as described in section 79.3.

Purpose of the Analysis:

Here, we are interested in testing the null hypothesis $H_0: \rho \leq 0.8$ is tested against the alternative hypothesis $H_1: \rho > 0.8$. For this analysis, we consider one-sided type I error of 0.025.

1. Choose the menu item:
Analysis > (Continuous) Two Samples > (Crossover Design) Ratio of Means
2. If the dataset is not displayed in your main window, this will bring up the **Select Dataset** dialog box with the list of available workbooks and datasets available under each workbook. If the dataset is already displayed in your main window, **East** will skip this step and the dataset in the main window will be used in the analysis. In case **East** brings up the **Select Dataset** dialog box, choose **pkfood.cyd** dataset under workbook **Crossover noninferiority** and click **OK**.
3. In the ensuing dialog box (under the **Main**) select the variables as shown below:

Analysis: Continuous Endpoint: Two-Sample Test - Crossover Design - Ratio of Means

Data Set: pkfood.cyd

Main Advanced

Trial Type: Noninferiority

Period ID: PERIOD

Response Variable: AUC

Subject ID: SUBJECT

Noninferiority Margin(ρ_0): 0.8

Group ID: SEQUENCE

Treatment Specifications

Treatment 1: Test

Treatment 2: Reference

4. In **Advanced** tab specify confidence interval as **0.975**. Click **OK** and the output will be displayed in the main window. Scroll down to the end of the output.

79 Analysis-Normal Noninferiority Two-Sample

Output for statistical test of treatment effect is displayed in the last table.

| Treatment Summary: | | | | | | | |
|-------------------------|---------------------|------------------|------------------------------------|-------------|----|-----------|------|
| Statistic | Test | Reference | | | | | |
| No. of Subjects | 20 | 20 | | | | | |
| Mean | 8455.7 | 10101.6 | | | | | |
| Std. Dev. | 3067.804 | 3590.908 | | | | | |
| Median | 8091.5 | 9604.5 | | | | | |
| Minimum | 4686 | 5648 | | | | | |
| Maximum | 16510 | 18864 | | | | | |
| Inference for: ln(AUC) | | | | | | | |
| Test of Hypothesis for: | | | | | | | |
| n | Difference of Means | Std. Effect Size | Std. Error | t | DF | (1-Sided) | |
| | | | | | | p-value | Tail |
| 20 | -0.179 | 0.126 | 0.079 | 0.561 | 18 | 0.291 | G.E. |
| Parameter Estimates: | | | | | | | |
| Metric | Estimate | Std. Error | 97.5% Confidence Interval(1-Sided) | | | | |
| | | | Lower Limit | Upper Limit | | | |
| ln(Ratio) | -0.179 | 0.079 | -0.345 | INF | | | |
| Ratio | 0.836 | 0.066 | 0.708 | INF | | | |

East performs the analysis based on the log-transformed data. The observed value of test statistic based on log-transformed data is 0.561 and it has $10 + 10 - 2 = 18$ degrees of freedom. The p-value associated with rejection $H_0: \rho \leq 0.8$ is 0.291. The one-sided 97.5% confidence interval for ρ is $(0.708, \infty)$. Since the lower limit of the confidence interval is smaller than the noninferiority margin of 0.8, we cannot reject $H_0: \rho \leq 0.8$ at one-sided 2.5% level of significance.

80 *Analysis-Normal Equivalence Two-Sample*

In many cases, the goal of a clinical trial is neither superiority nor non-inferiority, but equivalence. Chapter 13 deals with the design and simulation of these types of trials. This chapter explains how we can use **East** to perform analysis of data that comes from two independent samples and crossover equivalence studies.

80.1 *Example: Difference of Means*

Dataset: Iris.cyd

Data Description:

Iris flower dataset (Fisher, 1936) consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured for each sample: the length and the width of the sepals and petals, in centimeters. In this example we will consider sepal widths from I. virginica and I. versicolor respectively. The purpose here is to compare the mean sepal widths between I. virginica and I. versicolor.

Purpose of the Analysis:

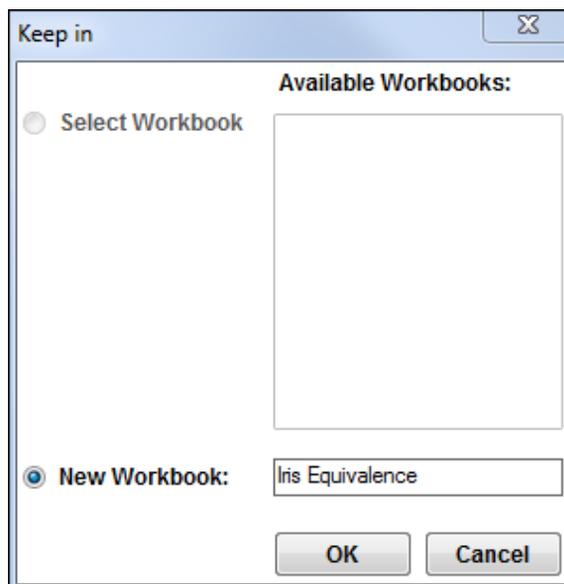
Let μ_t and μ_c be the mean sepal widths in I. virginica and I. versicolor, and $\delta = \mu_t - \mu_c$. We want to test the null hypothesis $H_0: \delta \leq -5$ or $\delta \geq 5$ against the alternative hypothesis $H_1: -5 < \delta < 5$. We want to reject H_0 with probability of type I error not exceeding 0.025.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. In case multiple workbooks are currently open, then this will bring up the **Keep in** dialog box. You can select either one of the existing workbooks or you can create new workbook. Suppose you want to create a new workbook labeled as “Iris Equivalence”. In order to do this, select the radio button **New Workbook**

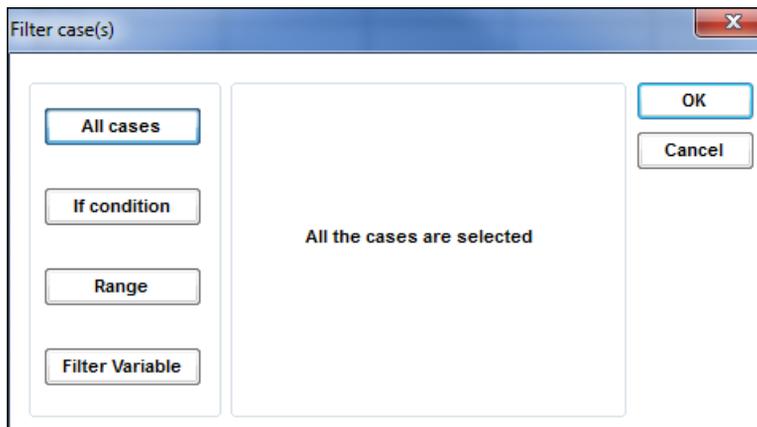
80 Analysis-Normal Equivalence Two-Sample

and type in **Iris Equivalence** in the field next to it.

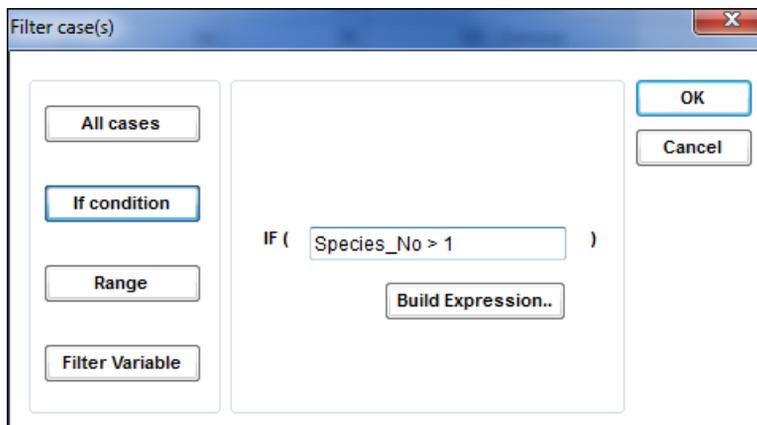


3. Click **OK**. The **Iris.cyd** dataset opens in the main window under the **Data Editor** menu. The dataset has observation from 150 subjects from the 3 species. The columns **Species_na** and **Sepal_widt** contains the information on name of species and width of sepals. We are considering **I. virginica** and **I. versicolor** only in this example. Therefore, we need to keep the data only from these two datasets and remove the remaining observations.
4. Under the **Data Editor** menu, click  icon in the **Data** ribbon. This shows

the **Filter case(s)** dialog box.



5. Click **If condition** button and enter **Species_No > 1**.



6. You can also use **Build Expression** to formulate a conditional expression for the **IF ()** field instead of directly writing the expression. Click **OK**. The observations pertaining to species **Setosa** are highlighted. Select these highlighted observations and click  icon under the **Data Editor** menu. The dataset will now have only 100 observations pertaining to *I. virginica* and *I. versicolor*.
7. Choose the menu item:
Analysis > (Continuous) Two Samples > (Parallel Design) Difference of

80 Analysis-Normal Equivalence Two-Sample

Means

- In the **Main** tab, select **Equivalence** as **Trial Type**, **Equal** as **Variance Type** and **t-test** as **Test Type**. For the **Population Id** field you have to choose a dichotomous variable. The variable selected in this field is the population identifier. Select **Species_na** as **Population Id** variable. As you select variable for **Population Id** field, a new box will appear below where you have to specify the levels of the **Population Id** variable for control and treatment group. Choose **Versicolor** for **Control**. **East** will treat the Versicolor as control and Verginica the treatment.
Select **sepal_widt** as **Response Variable** and enter -5 and 5 for **Lower Equiv. Limit** and **Upper Equiv. Limit**. The Frequency Variable allows the user to specify a variable that represents a frequency, or weighted value. For the current example, leave the **Frequency Variable** field blank.

Analysis: Continuous Endpoint Two-Sample Test - Parallel Desig

Data Set: Iris.cyd

Main Advanced

| | |
|--|--|
| Trial Type: <input type="text" value="Equivalence"/> | Test Type: <input type="text" value="t-test"/> |
| Variance Type: <input type="text" value="Equal"/> | |
| Population Id: <input type="text" value="Species_na"/> | Response Variable: <input type="text" value="Sepal_widt"/> |
| Control: <input type="text" value="Versicolor"/> | Lower Equiv. Limit: <input type="text" value="-5"/> |
| Treatment: <input type="text" value="Verginica"/> | Upper Equiv. Limit: <input type="text" value="5"/> |

- In the **Advanced** tab, leave the fields **By Variable 1** and **By Variable 2** blank. Enter 0.975 for **Confidence Level**.
- Click **OK** to start the analysis. Upon completion of analysis, a new node with the label **Analysis: Continuous Response: Difference of Means for Independent Data** will be added in the **Library** and the output is displayed in

the main window.

Analysis: Continuous Response: Difference of Means for Independent Data

Hypothesis
 $H_{01} : \mu_1 - \mu_c \leq \delta_L$ Or $H_{02} : \mu_1 - \mu_c \geq \delta_U$
 Vs.
 $H_1 : \delta_L < \mu_1 - \mu_c < \delta_U$

Input Parameters
 Data File: Iris.cyd
 Trial Type: Equivalence
 Population Id: Species_na(Treatment=Verginica, Control=Versicolor)
 Response Variable: Sepal_widt
 Variance Type: Equal
 Test Type: t-test
 Lower Equivalence Limit (δ_L): -5
 Upper Equivalence Limit (δ_U): 5
 Confidence Level (for Equivalence): 0.95

Output
 Response variable: (Sepal_widt)
 Total no. of records: 100
 No. of records rejected: 0
 Summary of the Observed Data:

| Treatment Group | Min | Max | Median | Mean | Std.Dev | n |
|-----------------|-----|-----|--------|-------|---------|----|
| Versicolor | 20 | 34 | 28 | 27.64 | 3.141 | 50 |
| Verginica | 22 | 38 | 30 | 29.74 | 3.225 | 50 |

Test of Hypothesis:

| n | Difference of Means | Std. Effect Size | | Std. Error | t | | DF | 95% Confidence Interval(2-Sided) | | p-value | |
|-----|---------------------|------------------|-----------|------------|-----------|-----------|----|----------------------------------|-------------|-----------|-----------|
| | | Under H01 | Under H02 | | Under H01 | Under H02 | | Lower Limit | Upper Limit | Under H01 | Under H02 |
| 100 | 2.1 | 2.23 | -0.911 | 0.637 | 11.152 | -4.555 | 98 | 0.837 | 3.363 | 1.110E-16 | 7.552E-6 |

The result of the analysis is divided in three sections. The **Hypothesis** section states the null and alternative hypothesis for 2-sided and 1-sided tests.

The **Input Parameters** section displays the name of the data file, the response variable, type of test performed, type I error set for the analysis and other parameter(s) used in the analysis. It is important to review this section to ensure correct and complete input parameters are specified.

The last section is **Output**. First part of the output section is about the descriptive statistics about the response variable. There are 50 observations in each group. Mean sepal lengths (standard deviation) are 27.64 (3.141) and 29.74 (3.225) in I. versicolor and I. verginica groups. Estimated treatment difference is $\hat{\delta} = 2.1$ with $se(\hat{\delta}) = 0.637$. There are two effect sizes - 2.23 (under H_{01}) and -0.911 (under H_{02}). These values can be verified by plugging the value of $\hat{\delta} = 2.1$, $\delta_L = -5$, $\delta_U = 5$ and $\hat{\sigma} = 0.637 / \sqrt{1/50 + 1/50} = 3.185$ in the following formula of effect size under H_{01} and H_{02} .

80 Analysis-Normal Equivalence Two-Sample

$$\frac{\hat{\delta} - \delta_L}{\hat{\sigma}} \quad \text{and} \quad \frac{\hat{\delta} - \delta_U}{\hat{\sigma}}$$

The observed value of two test statistics are 11.152 and -4.555 and both of them have $50 + 50 - 2 = 98$ degrees of freedom. The two-sided 95% confidence interval of $\delta = \mu_t - \mu_c$ is (0.837, 3.363). This confidence interval is within the equivalence interval of (-5, 5), therefore, we can reject $H_0: \mu_t - \mu_c \leq -5$ or $\mu_t - \mu_c \geq 5$ in favor of $H_1: -5 < \mu_t - \mu_c < 5$ with 5% level of significance.

80.2 Example: Log-ratio of Means

Dataset: We will again use dataset **Iris.cyd** here.

Data Description:

Description of this dataset is given in subsection 80.1.

Purpose of the Analysis:

Let μ_t and μ_c be the mean sepal widths of I. virginica and I. versicolor, and $\rho = \mu_t / \mu_c$. Here, we are interested in testing the null hypothesis

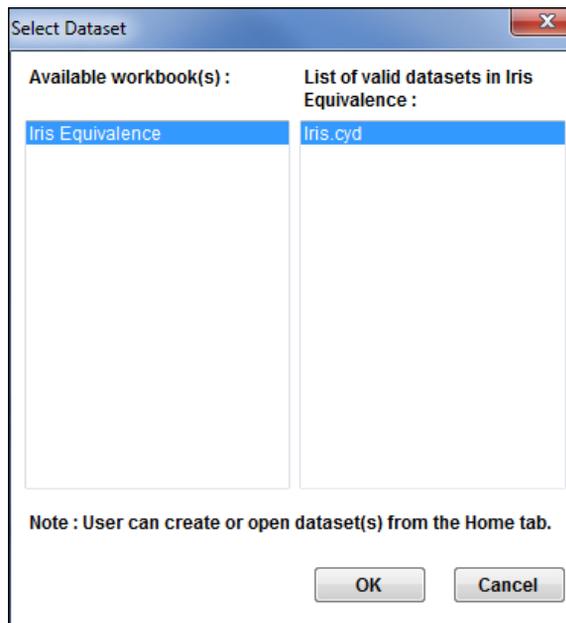
$H_0: \rho \leq 0.8$ or $\rho \geq 1.25$ is tested against the alternative hypothesis

$H_1: 0.8 < \rho < 1.25$. We want to reject H_0 with type I error not exceeding 0.025.

Analysis Steps:

1. Choose the menu item:
Analysis > (Continuous) Two Samples > (Parallel Design) Ratio of Means
2. If the dataset is not displayed in your main window, this will bring up the **Select Dataset** dialog box with the list of available workbooks and datasets available under each workbook. If the dataset is already displayed in your main window, **East** will skip this step and the dataset in the main window will be used in the analysis. In case **East** brings up the **Select Dataset** dialog box, choose **Iris.cyd**

dataset under workbook **Iris Equivalence** and click **OK**.



3. In the ensuing dialog box (under the **Main** tab) select the variables as shown below:

Data Set: Iris.cyd

| Main | | Advanced | |
|----------------|-------------|---------------------|------------|
| Trial Type: | Equivalence | Test Type: | t-test |
| Variance Type: | Equal | | |
| Population Id: | Species_na | Response Variable: | Sepal_widt |
| Control: | Versicolor | Lower Equiv. Limit: | 0.8 |
| Treatment: | Verginica | Upper Equiv. Limit: | 1.25 |

4. In the **Advanced** tab, leave **By Variable 1** and **By Variable 2** blank and enter

80 Analysis-Normal Equivalence Two-Sample

0.975 for **Confidence Level**.

Analysis: Continuous Endpoint Two-Sample Test - Parallel Design - Wilcoxon-Mann-Whitney

Data Set: Iris.cyd

Main **Advanced**

| | | | |
|----------------|----------------------|---------------------------------|------------------------------------|
| By Variable 1: | <input type="text"/> | Confidence Level: | <input type="text" value="0.975"/> |
| By Variable 2: | <input type="text"/> | Confidence Level (Equivalence): | <input type="text" value="0.95"/> |

5. Click **OK** to start the analysis. Upon completion of the analysis, following

output is displayed in the main window.

Analysis: Continuous Response: Ratio of Means for Independent Data

Hypothesis

$H_{01} : \mu_t / \mu_c \leq p_L$ Or $H_{02} : \mu_t / \mu_c \geq p_U$

Vs.

$H_1 : p_L < \mu_t / \mu_c < p_U$

Input Parameters

Data File: Iris.cyd
 Trial Type: Equivalence
 Population Id: Species_na(Treatment=Verginica, Control=Versicolor)
 Response Variable: Sepal_widt
 Variance Type: Equal
 Test Type: t-test
 Lower Equivalence Limit (p_L): 0.8
 Upper Equivalence Limit (p_U): 1.25
 Confidence Level (for Equivalence): 0.95

Output

Response variable: (Sepal_widt)
 Total no. of records: 100
 No. of records rejected: 0

Summary of the Observed Data:

| Treatment Group | Min | Max | Median | Mean | Std.Dev | n |
|-----------------|-----|-----|--------|-------|---------|----|
| Versicolor | 20 | 34 | 28 | 27.64 | 3.141 | 50 |
| Verginica | 22 | 38 | 30 | 29.74 | 3.225 | 50 |

Test of Hypothesis for: ln(Sepal_widt)

| n | Difference of Means | Std. Effect Size | | Std. Error | t | | DF | p-value | |
|-----|---------------------|------------------|-----------|------------|-----------|-----------|----|-----------|-----------|
| | | Under H01 | Under H02 | | Under H01 | Under H02 | | Under H01 | Under H02 |
| 100 | 0.074 | 2.636 | -1.322 | 0.023 | 13.181 | -6.611 | 98 | 1.110E-16 | 1.008E-9 |

Parameter Estimates:

| Metric | Estimate | Std. Error | 95% Confidence Interval(2-Sided) | |
|-----------|----------|------------|----------------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| ln(Ratio) | 0.074 | 0.023 | 0.029 | 0.119 |
| Ratio | 1.077 | 0.024 | 1.03 | 1.126 |

First two sections display the information about the hypothesis tested and the inputs specified. In the **Output** section, the first part provides descriptive statistics for the two groups. The second table labeled with **Test of Hypothesis for:ln(Sepal_widt)** provides details about the test result. Note the word “ln(Sepal_widt)”; this emphasizes that the analysis is performed on log-transformed data. In this table, the **Difference of Means** is 0.074. This is the estimated treatment difference in terms of log-transformed data on **Sepal_widt**. In this example, the two effects sizes are 2.636 and -1.322. The observed value of two test statistics are 13.181 and -6.611 and both of them have 98 degrees of

80 Analysis-Normal Equivalence Two-Sample

freedom. The two-sided 95% confidence interval of $\rho = \mu_t/\mu_c$ is (1.03, 1.126). This confidence interval is within the equivalence interval of (0.80, 1.25), therefore, we can reject $H_0: \mu_t/\mu_c \leq 0.80$ or $\mu_t/\mu_c \geq 1.25$ in favor of $H_1: 0.80 < \mu_t/\mu_c < 1.25$ with 5% level of significance..

80.3 Example: Difference of Means in Crossover Designs

Crossover trials are widely used in clinical and medical research and in other diversified areas such as veterinary science, psychology, sports science, dairy science, and agriculture. Crossover design is often preferred over parallel design because each subject receives all the treatments and thus each subject acts as their own control. In this section, we show how **East** supports the design and simulation of such experiments with endpoint as difference of means.

Dataset: pkfood.cyd

Data Description:

Here we will use pharmacokinetic data from 2×2 crossover trial available in **pkfood.cyd**. The dataset consists of observations from 20 subjects on AUC , C_{max} and T_{max} evaluated under two regimens A and B. For this example, we will consider regimen B as reference and regimen A as test drug and AUC as response variable.

Purpose of the Analysis:

Let μ_c and μ_t denote the mean AUC in regimen B and regimen A and $\delta = \mu_t - \mu_c$. Here we set the bioequivalence limits (δ_L, δ_U) as $(-5000, 5000)$. We are interested in testing $H_0: \delta \leq -5000$ or $\delta \geq 5000$ against $H_1: -5000 < \delta < 5000$. For this analysis, probability of type I error of 0.05 is considered.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. In case multiple workbooks are currently open, then this will bring up the **Keep in** dialog box. You can select either one of the existing workbooks or you can create new workbook. Suppose you want to create a new workbook labeled as “Crossover Equivalence”. In order to do this, select the radio button **New Workbook** and type in **Crossover Equivalence** in the field next to it. Click **OK**. This will open the **pkfood.cyd** dataset in the main window of under the **Data Editor**.
3. Choose the menu item:
Analysis > (Continuous) Two Samples > (Crossover Design) Difference of Means

4. In the ensuing dialog box (under the **Main** tab) select/enter the different variables as shown below.

5. In the **Advanced** tab, leave **By Variable 1** and **By Variable 2** blank and enter 0.95 for **Confidence Level**.

6. Click **OK** to analyze the data. Following output will be displayed in the main

80 Analysis-Normal Equivalence Two-Sample

window.

Analysis: Continuous Response: Difference of Means test for Crossover Data

$H_{01} : \mu_1 - \mu_2 \leq \delta_L$ Or $H_{02} : \mu_1 - \mu_2 \geq \delta_U$
 Vs.
 $H_1 : \delta_L < \mu_1 - \mu_2 < \delta_U$

Data File: pkfood.cyd
 Trial Type: Equivalence
 Response Variable: AUC
 Group ID: SEQUENCE
 Subject ID: SUBJECT
 Period ID: PERIOD
 Test Type: t-Test
 Lower Equivalence Limit(δ_L): -5000
 Upper Equivalence Limit(δ_U): 5000
 Confidence Level (for Equivalence): 0.9
 Treatment Assignment:

| Group ID | Period ID | |
|----------|-----------|-----------|
| | 1 | 2 |
| AB | Test | Reference |
| BA | Reference | Test |

Output

Response Variable: (AUC)
 Total Number of Records: 40
 Number of Records Rejected: 0

Summary of the Observed Data:

| Group ID | Statistic | Period ID | | Group Summary |
|----------------|-----------------|-----------|----------|---------------|
| | | 1 | 2 | |
| AB | No. of Subjects | 10 | 10 | 10 |
| | Mean | 8542 | 9582.9 | 9062.45 |
| | Std. Dev. | 2978.945 | 3820.535 | 3376.8 |
| | Median | 8641 | 7859.5 | 8012 |
| | Minimum | 4686 | 5648 | 4686 |
| | Maximum | 14059 | 15765 | 15765 |
| BA | No. of Subjects | 10 | 10 | 10 |
| | Mean | 10620.3 | 8369.4 | 9494.85 |
| | Std. Dev. | 3468.083 | 3313.29 | 3497.233 |
| | Median | 9852.5 | 7607.5 | 9165.5 |
| | Minimum | 5911 | 5118 | 5118 |
| | Maximum | 18864 | 16510 | 18864 |
| Period Summary | No. of Subjects | 20 | 20 | |
| | Mean | 9581.15 | 8976.15 | |
| | Std. Dev. | 3322.27 | 3535.775 | |
| | Median | 9641.5 | 7859.5 | |
| | Minimum | 4686 | 5118 | |
| | Maximum | 18864 | 16510 | |

Treatment Summary:

| Statistic | Test | Reference |
|-----------------|----------|-----------|
| No. of Subjects | 20 | 20 |
| Mean | 8455.7 | 10101.6 |
| Std. Dev. | 3067.804 | 3590.908 |
| Median | 8091.5 | 9604.5 |
| Minimum | 4686 | 5648 |
| Maximum | 16510 | 18864 |

Inference for: (AUC)
Test of Hypothesis for:

| n | Difference of Means | Std. Effect Size | | Std. Error | t | | DF | (1-Sided) | (2-Sided) | 90% Confidence Interval(2-Side | |
|----|---------------------|------------------|-----------|------------|---------|---------|----|-------------|-------------|--------------------------------|----------|
| | | Under H01 | Under H02 | | p-value | p-value | | Lower Limit | Upper Limit | | |
| 20 | -1645.9 | 0.95 | -1.882 | 789.591 | 4.248 | -8.417 | 18 | 2.42E-4 | 5.9E-8 | -3015.101 | -276.699 |

In the **Output** section, the first part provides descriptive statistics for the two groups. The second table provides the treatment summary. The third table labeled as **Test of Hypothesis for** provides results for statistical test of treatment effect. The observed values of two test statistics are 4.248 and -8.417 and both of them have 18 degrees of freedom. The 2-sided 90% confidence interval of $\delta = \mu_t - \mu_c$ is (-3015, -277). This confidence interval is well within the equivalence interval of (-5000, 5000), therefore, we can reject $H_0: \mu_t - \mu_c \leq -5000$ or $\mu_t - \mu_c \geq 5000$ in favor of $H_1: -5000 < \mu_t - \mu_c < 5000$ with 5% level of significance.

80.4 Example: Ratio of Means in Crossover Designs

Often in crossover designs, equivalence hypothesis is tested in terms of ratio of means. This type of trial is very popular in establishing bioequivalence and bioavailability between two formulations in terms of pharmacokinetic parameters (FDA guideline on BA/BE studies for orally administered drug products, 2003). In particular, FDA considers two products bioequivalent if the 90% confidence interval of the ratio of two means lie within (0.8, 1.25). This chapter, shows how **East** is used to analyze data from such experiments with endpoint as ratio of means.

Since the ratio hypothesis is translated into difference hypothesis using log transformation, **East** performs two one sided tests (TOST) on the log-transformed data as discussed in section 80.3.

Dataset: We will again use **pkfood.cyd** dataset here.

Data Description:

Description of this dataset is given in subsection 80.3.

Purpose of the Analysis:

Here we are interested in ratio of means. Let μ_t and μ_c denote the means of the observations from the experimental treatment (T) and the control treatment (C). In equivalence trial with endpoint as ratio of means, the goal is to establish $\rho_L < \rho < \rho_U$, where ρ_L and ρ_U are specified values used to define equivalence. In practice, ρ_L and ρ_U are often chosen such that $\rho_L = 1/\rho_U$

The null hypothesis $H_0: \rho \leq \rho_L$ or $\rho \geq \rho_U$ is tested against the two-sided alternative

80 Analysis-Normal Equivalence Two-Sample

hypothesis $H_1: \rho_L < \rho < \rho_U$ at level α , using two one-sided tests. Schuirmann (1987) proposed working this problem out on the natural logarithm scale. Thus we are interested in the parameter $\delta = \ln(\rho) = \ln(\mu_t) - \ln(\mu_c)$ and the null hypothesis $H_0: \delta \leq \delta_L$ or $\delta \geq \delta_U$ is tested against the 2-sided alternative hypothesis $H_1: \delta_L < \delta < \delta_U$ at level α , using two one-sided t-tests. Here $\delta_L = \ln(\rho_L)$ and $\delta_U = \ln(\rho_U)$.

Here, we are interested in testing the null hypothesis $H_0: \rho \leq 0.8$ or $\rho \geq 1.25$ against the alternative hypothesis $H_1: 0.8 < \rho < 1.25$. For this analysis, consider type I error rate of 0.05.

Analysis Steps:

1. Choose the menu item:
Analysis > (Continuous) Two Samples > (Crossover Design) Ratio of Means
2. If the dataset is not displayed in your main window, this will bring up the **Select Dataset** dialog box with the list of available workbooks and datasets available under each workbook. If the dataset is already displayed in your main window, **East** will skip this step and the dataset in the main window will be used in the analysis. In case **East** brings up the **Select Dataset** dialog box, choose **pkfood.cyd** dataset under workbook **Crossover Equivalence** and click **OK**.
3. In the ensuing dialog box (under the **Main** tab) select/enter the variables as shown below:

Analysis: Continuous Endpoint Two-Sample Test - Crossover Design - Ratio of Means

Data Set: pkfood.cyd

Main Advanced

Trial Type: Equivalence Period ID: PERIOD Response Variable: AUC

Subject ID: SUBJECT Lower Equiv. Limit(ρ_L): 0.8

Treatment Specifications Group ID: SEQUENCE Upper Equiv. Limit(ρ_U): 1.25

Treatment 1: Test

Treatment 2: Reference

4. In the **Advanced** tab specify confidence interval as 0.95.
5. Click **OK** to start the analysis. Upon completion of the analysis, a new node with label **Analysis: Continuous Response: Ratio of Means for Crossover Data1** is added to the **Library** and the output is displayed in the main window. Scroll down to the end of the output. Output for statistical test of treatment effect

is displayed in the last two tables.

| Inference for: ln(AUC) | | | | | | | | |
|-------------------------|---------------------|------------------|----------------------------------|-------------|-----------|-----------|-----------|-----------|
| Test of Hypothesis for: | | | | | | | | |
| n | Difference of Means | Std. Effect Size | | Std. Error | t | | p-value | |
| | | Under H01 | Under H02 | | Under H01 | Under H02 | Under H01 | Under H02 |
| 20 | -0.179 | 0.126 | -1.137 | 0.079 | 0.561 | -5.086 | 0.291 | 3.854E-5 |
| Parameter Estimates: | | | | | | | | |
| Metric | Estimate | Std. Error | 90% Confidence Interval(2-Sided) | | | | | |
| | | | Lower Limit | Upper Limit | | | | |
| ln(Ratio) | -0.179 | 0.079 | -0.316 | -0.042 | | | | |
| Ratio | 0.836 | 0.066 | 0.729 | 0.959 | | | | |

East performs the analysis based on the log-transformed data. The observed values of test statistics based on log-transformed data are 0.561 and -5.086 and they are distributed with $10 + 10 - 2 = 18$ degrees of freedom. The 2-sided 95% confidence interval of $\rho = \mu_t/\mu_c$ is (0.729, 0.959). This confidence interval is NOT within the equivalence interval of (0.80, 1.25), therefore, we can reject $H_0: \mu_t/\mu_c < 0.80$ or $\mu_t/\mu_c > 1.25$ in favor of $H_1: 0.80 \leq \mu_t/\mu_c \leq 1.25$ NOT at 5% level of significance.

81 Analysis-Nonparametric Two-Sample

The Wilcoxon-Mann-Whitney nonparametric test is commonly used for the comparison of two distributions when the observations cannot be assumed to come from normal distributions. It is used when the distributions differ only in a location parameter and is especially useful when the distributions are not symmetric.

East supports analysis using Wilcoxon-Mann-Whitney nonparametric test for parallel as well as crossover designs. The former is discussed in Section 81.1, 81.2 and 81.3 and the later in Section 81.4, 81.5 and 81.6.

81.1 Test for Superiority

81.1.1 Example

Let X_1, \dots, X_{n_t} be the n_t observations from the treatment (T) with distribution function F_t and Y_1, \dots, Y_{n_c} be the n_c observations from the control (C) with distribution function F_c . F_t and F_c are assumed to be continuous with corresponding densities f_t and f_c , respectively.

The primary objective in Wilcoxon-Mann-Whitney test is to investigate whether there is a shift of location, which indicates the presence of the treatment effect. Let θ represents the treatment effect. That is, $F_t(z) = F_c(z + \theta)$.

In a superiority trial, we test the null hypothesis $H_0: \theta = 0$ against the two-sided alternative $H_1: \theta \neq 0$ or a one-sided alternative hypothesis $H_1: \theta < 0$ or $H_1: \theta > 0$. The test statistic is the sum of the ranks for the treatment in the pooled sample minus $n_t(n_t + 1)/2$ or equivalently the number of pairs (X_i, Y_j) such that $X_i < Y_j$. Usually, the test statistic is denoted by W . Asymptotically, this is distributed with following mean and variance

$$E(W) = \frac{n_t(n_t + n_c + 1)}{2} \quad \text{var}(W) = \frac{n_t n_c (n_t + n_c + 1)}{12}$$

The standardized test statistic, Z , is obtained as

$$Z = \frac{W - E(W)}{\sqrt{\text{var}(W)}}$$

The p-value is calculated assuming Z is distributed as standard normal variate.

81.1.1 Example

Dataset : Myeloma.cyd as described in Section 72.1.1.

Purpose of the Analysis:

The purpose here is to compare the values of the variable haemoglobin level between two groups indicated by the variable status (0-alive, 1-dead). Let θ be the median difference between the two groups. We will use θ_t and θ_c to denote the median haemoglobin in the alive and dead groups, respectively. We are interested in testing the null hypothesis $H_0: \theta = 0$ with type I error not exceeding 5% level of significance, where $\theta = \theta_t - \theta_c$.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
**Analysis > (Continuous) Two Samples > (Parallel Design)
Wilcoxon-Mann-Whitney**
3. In the ensuing dialog box choose the variables as shown below:

Analysis: Continuous Endpoint Two-Sample Test - Parallel Design - Wilcoxon-Mann-Whitney

Data Set: Myeloma.cyd

Main Advanced

Trial Type: Superiority Response Variable: haemoglobin

Population Id: status

Control: 1

Treatment: 0

4. Click **OK** to start analysis. The output will be displayed in the main window

81 Analysis-Nonparametric Two-Sample

now.

Analysis: Continuous Response: Wilcoxon Mann Whitney Test for Independent Data

Hypothesis
 $H_0 : \theta_t - \theta_c = 0$ Vs. $H_1 : \theta_t - \theta_c \neq 0$ for 2-sided test
 Either $H_1 : \theta_t - \theta_c < 0$
 Or $H_1 : \theta_t - \theta_c > 0$ for 1-sided test
 where θ_t and θ_c are the medians of the Treatment Group and Control Group.

Input Parameters

Data File: Myeloma.cyd
 Trial Type: Superiority
 Population Id: status(Treatment=0, Control=1)
 Response Variable: haemoglobin
 Confidence Level: 0.95

Output

Response Variable: (haemoglobin)
 Total Number of Records: 65
 Number of Records Rejected: 0
 Summary of the Observed Data:

| Treatment Group | Min | Max | Median | Mean | Std.Dev | n |
|-----------------|------|------|--------|-------|---------|---|
| 5 | 14.6 | 10.1 | 9.91 | 2.564 | 48 | |
| 4.9 | 14 | 11.6 | 11.024 | 2.425 | 17 | |

Summary of Test Statistic:

| Minimum | Maximum | Mean | Std. dev. | Observed Statistics | Standardized |
|---------|---------|------|-----------|---------------------|--------------|
| 153 | 968 | 561 | 66.959 | 672 | 1.658 |

Test of Hypothesis:

| Estimate of Median Difference | Standardized Statistic | (1-Sided) | | (2-Sided) | 95% Confidence Interval(2-Sided)* | |
|-------------------------------|------------------------|-----------|---------|-----------|-----------------------------------|-------------|
| | | Tail | p-value | p-value | Lower Limit | Upper Limit |
| 1.2 | 1.658 | G.E. | 0.049 | 0.097 | -0.2 | 2.7 |

*The Nonparametric estimates are based on Hodges-Lehmann's formulation.

The last section is the **Output**. First part of the output is about the descriptive statistics about the response variable. There are 65 observations. The mean (standard deviation) hemoglobin levels are 9.91 (2.564) and 11.024 (2.425) in control and treatment groups, respectively. Estimated median difference between the two groups is 1.2. The observed test statistic is $W=672$. The value of standardized statistic is 1.658 and this is obtained according to Eq. 81.1. The 2-sided p-value for comparison of two groups is 0.097. We conclude that based on Wilcoxon Mann Whitney test, the medians in two groups are not significantly different at 5% significance level.

81.2 Test for Non-inferiority

81.2.1 Example

As before, we assume that X_1, \dots, X_{n_t} be the n_t observations from the treatment (T) with distribution function F_t and Y_1, \dots, Y_{n_c} be the n_c observations from the control (C) with distribution function F_c . F_t and F_c are assumed to be continuous with corresponding densities f_t and f_c , respectively. Let θ be the shift of location such that, $F_t(z) = F_c(z + \theta)$.

In a non-inferiority trial, we test the null hypothesis $H_0: \theta \leq \delta_0$ against the alternative hypothesis $H_1: \theta > \delta_0$ if $\delta_0 < 0$ or $H_0: \theta \geq \delta_0$ against the alternative hypothesis $H_1: \theta < \delta_0$ if $\delta_0 > 0$. **East** first subtracts δ_0 from X_1, \dots, X_{n_t} and then the value of test statistic, standardized test statistic and p-value are calculated as done in superiority trial.

81.2.1 Example

Dataset: `Werner.cyd` as described in Section 73.4.2

Purpose of the Analysis:

The purpose here is to compare the median cholesterol level in birthpill user group (T) with the non-user group (C) with non-inferiority margin (δ_0) of 25 and one-sided type I error of 0.025. Let θ_t and θ_c be the median cholesterol levels in birthpill user and non-user groups, respectively. Since $\delta_0 = 25 > 0$, we are testing $H_0: \theta \geq \delta_0$ against the alternative hypothesis $H_1: \theta < \delta_0$, where $\theta = \theta_t - \theta_c$.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
**Analysis > (Continuous) Two Samples > (Parallel Design)
 Wilcoxon-Mann-Whitney**

81 Analysis-Nonparametric Two-Sample

3. In the ensuing dialog box choose the variables as shown below:

Analysis: Continuous Endpoint Two-Sample Test - Parallel Design - Wilcoxon-Mann-Whitney

Data Set: Werner.cyd

Main Advanced

Trial Type: Noninferiority Response Variable: CHOLESTEROL

Population Id: BIRTHPILL Noninferiority Margin: 25

Control: 0

Treatment: 1

4. Now, click on **Advanced** tab and enter 0.975 for **Confidence Level**.

Analysis: Continuous Endpoint Two-Sample Test - Parallel Design - Wilcoxon-Mann-Whitney

Data Set: Werner.cyd

Main Advanced

By Variable 1: Confidence Level: 0.975

By Variable 2:

5. Click **OK** to start analysis. The output will be displayed in the main window

now.

Analysis: Continuous Response: Wilcoxon Mann Whitney Test for Independent Data

If $\delta_0 < 0$ then $H_0 : \theta_t - \theta_c \leq \delta_0$ Vs. $H_1 : \theta_t - \theta_c > \delta_0$

If $\delta_0 > 0$ then $H_0 : \theta_t - \theta_c \geq \delta_0$ Vs. $H_1 : \theta_t - \theta_c < \delta_0$

where θ_t, θ_c are the medians of the Treatment Group and Control Group.

Data File: Werner.cyd
 Trial Type: Noninferiority
 Population Id: BIRTHPILL(Treatment=1, Control=0)
 Response Variable: CHOLESTEROL
 Noninferiority Margin (δ_0): 25
 Confidence Level: 0.975

Output

Response Variable: (CHOLESTEROL)
 Total Number of Records: 188
 Number of Records Rejected: 0
 Summary of the Observed Data:

| Treatment Group | Min | Max | Median | Mean | Std.Dev | n |
|-----------------|-----|-----|--------|---------|---------|----|
| 0 | 155 | 335 | 230 | 232.968 | 43.492 | 94 |
| 1 | 50 | 600 | 235 | 240.585 | 58.924 | 94 |

Summary of Test Statistic:

| Minimum | Maximum | Mean | Std. dev. | Observed Statistics | Standardized |
|---------|---------|------|-----------|---------------------|--------------|
| 4465.5 | 13300.5 | 8883 | 372.95 | 7781.5 | -2.953 |

Test of Hypothesis:

| Estimate of Median Difference | Standardized Statistic | (1-Sided) | | 97.5% Confidence Interval(1-Sided)* | |
|-------------------------------|------------------------|-----------|---------|-------------------------------------|-------------|
| | | Tail | p-value | Lower Limit | Upper Limit |
| 5 | -2.953 | L.E. | 0.002 | -INF | 19 |

*The Nonparametric estimates are based on Hodges-Lehmann's formulation.

The last section is the **Output**. First part of the output is about the descriptive statistics about the response variable. There are 25 observations. Estimate of $\theta = \theta_t - \theta_c$ (i.e., median difference) is 5. The observed value of the test statistic (W) and standardized test statistic (Z) are 7781.5 and -2.953, respectively. The p-value for this non-inferiority test is 0.002. Therefore, we conclude that the Birthpill user group is non-inferior to the non-user group in terms of cholesterol level with non-inferiority margin of 25.

81.3 Test for Equivalence

81.3.1 Example

As before, we assume that X_1, \dots, X_{n_t} be the n_t observations from the treatment (T) with distribution function F_t and Y_1, \dots, Y_{n_c} be the n_c observations from the control (C) with distribution function F_c . F_t and F_c are assumed to be continuous with

81 Analysis-Nonparametric Two-Sample

corresponding densities f_t and f_c , respectively. Let θ be the shift of location such that, $F_t(z) = F_c(z + \theta)$.

The null hypothesis $H_0: \theta \leq \delta_L$ or $\theta \geq \delta_U$ is tested against the two-sided alternative hypothesis $H_1: \delta_L < \theta < \delta_U$ at level α , using the following two one-sided tests (TOST).

- Test1: $H_{0L}: \theta \leq \delta_L$ against $H_{1L}: \theta > \delta_L$ at level α
- Test2: $H_{0U}: \theta \geq \delta_U$ against $H_{1U}: \theta < \delta_U$ at level α

East subtracts δ_L and δ_U from X_1, \dots, X_{n_t} for Test1 and Test2, respectively. Then the value of test statistic, standardized test statistic and p-value are calculated separately as done in superiority trial. To declare equivalence, both H_{0L} and H_{0U} need to be rejected.

81.3.1 Example

Dataset: **Iris.cyd** as described in Section 80.1

Purpose of the Analysis:

The purpose here is to compare the median sepal widths between I. virginica and I. versicolor with equivalence limits (δ_L, δ_U) as $(-5, 5)$. Let θ_t and θ_c denote the median sepal widths in I. virginica and I. versicolor, respectively, and $\theta = \theta_t - \theta_c$.

We want to test the null hypothesis $H_0: \theta \leq -5$ or $\theta \geq 5$ against the alternative hypothesis $H_1: -5 < \theta < 5$. We want to reject H_0 with type I error rate not exceeding 0.05.

Analysis Steps:

1. Open the **Iris.cyd** from the **Samples** folder and keep only the observations pertaining to I. virginica and I. versicolor as described in subsection 80.1.
2. Choose the menu item:
**Analysis > (Continuous) Two Samples > (Parallel Design)
Wilcoxon-Mann-Whitney**

3. In the ensuing dialog box choose the variables as shown below:

Data Set: Iris.cyd

Main **Advanced**

Trial Type: Response Variable:

Population Id: Lower Equiv. Limit:

Control: Upper Equiv. Limit:

Treatment:

4. Now click on **Advanced** tab. Enter 0.975 for **Confidence Level**.

Analysis: Continuous Endpoint Two-Sample Test - Parallel Design - Wilcoxon-Mann-Whitney

Data Set: Iris.cyd

Main **Advanced**

By Variable 1: Confidence Level:

By Variable 2: Confidence Level (Equivalence):

5. Click **OK** to start analysis. The output will be displayed in the main window

81 Analysis-Nonparametric Two-Sample

now.

Analysis: Continuous Response: Wilcoxon Mann Whitney Test for Independent Data

Hypothesis
 $H_{01} : \theta_t - \theta_c \leq \delta_L$ Or $H_{02} : \theta_t - \theta_c \geq \delta_U$
 Vs. $H_1 : \delta_L < \theta_t - \theta_c < \delta_U$
 where θ_t and θ_c are the medians of the Treatment Group and Control Group.

Input Parameters

Data File: Iris.cyd
 Trial Type: Equivalence
 Population Id: Species_na(Treatment=Verginica, Control=Versicolor)
 Response Variable: Sepal_widt
 Lower Equivalence Limit (δ_L): -5
 Upper Equivalence Limit (δ_U): 5
 Confidence Level (for Equivalence): 0.95

Output

Response variable: (Sepal_widt)
 Total no. of records: 100
 No. of records rejected: 0
 Summary of the Observed Data:

| Treatment Group | Min | Max | Median | Mean | Std.Dev | n |
|-----------------|-----|-----|--------|-------|---------|----|
| Versicolor | 20 | 34 | 28 | 27.64 | 3.141 | 50 |
| Verginica | 22 | 38 | 30 | 29.74 | 3.225 | 50 |

Summary of Test Statistic:

| Statistic | Minimum | Maximum | Mean | Std. dev. | Observed Statistics | Standardized |
|-----------|---------|---------|------|-----------|---------------------|--------------|
| Under H01 | 1278 | 3772 | 2525 | 144.647 | 3656.5 | 7.822 |
| Under H02 | 1283 | 3767 | 2525 | 144.39 | 1905 | -4.294 |

Test of Hypothesis:

| Estimate of Median Difference | Standardized Statistic | | p-value | | 95% Confidence Interval(2-Sided)* | |
|-------------------------------|------------------------|-----------|-----------|-----------|-----------------------------------|-------------|
| | Under H01 | Under H02 | Under H01 | Under H02 | Lower Limit | Upper Limit |
| 2 | 7.822 | -4.294 | 0 | 8.777E-6 | 1 | 3 |

The last section is the **Output**. First part of the output is about the descriptive statistics about the response variable. There are 50 observations in each group. Median sepal lengths are 28 and 30 in I. versicolor and I. verginica groups, respectively. Estimate of $\theta = \theta_t - \theta_c$ is 2. The observed values of test statistic and standardized test statistic are 3656.5 and 7.822, respectively for the H_{0L} and 1905 and -4.294, respectively for the H_{0U} . The p-values associated with H_{0L} and H_{0U} are very close to 0. Therefore, we can reject individually both H_{0L} and H_{0U} . Thus, we reject $H_0: \theta \leq -5$ or $\theta \geq 5$ with very small p-value

81.4 Test for Superiority in Crossover Trial

81.4.1 Example

In a 2×2 crossover design each subject is randomized to one of two sequence groups. Subjects in the sequence group 1 receive the test drug (T) formulation in a first period, have their outcome variable, X recorded, wait out a washout period to ensure that the drug is cleared from their system, then receive the control drug formulation (C) in period 2 and finally have the measurement on X again. In sequence group 2, the order in which the T and C are assigned is reversed. The table below summarizes this type of trial design.

| Group | Period 1 | Washout | Period 2 |
|-------|----------|---------|----------|
| 1(TC) | Test | — | Control |
| 2(CT) | Control | — | Test |

The resulting data are commonly analyzed using a statistical linear model. The response y_{ijk} in period j on subject k in sequence group i , where $i = 1, 2$, $j = 1, 2$, and $k = 1, \dots, n_i$ is modeled as a linear function of an overall mean response μ , formulation effect τ_t and τ_c , period effects π_1 and π_2 , and sequence effects λ_1 and λ_2 . The fixed effects model can be displayed as:

| Group | Period 1 | Washout | Period 2 |
|-------|-----------------------------------|---------|------------------------------------|
| 1(TC) | $\mu + \tau_t + \pi_1 + \gamma_1$ | — | $\mu + \tau_c + \pi_2 + \lambda_1$ |
| 2(CT) | $\mu + \tau_c + \pi_1 + \gamma_2$ | — | $\mu + \tau_t + \pi_2 + \lambda_2$ |

For superiority trial, **East** can test following null hypotheses:

- Test1: $H_0 : \tau_t - \tau_c = 0$. for treatment effect
- Test2: $H_0 : \pi_1 - \pi_2 = 0$. for period effect
- Test1: $H_0 : \lambda_1 - \lambda_2 = 0$. for carryover effect

To test the above hypotheses **East** uses Hodges-Lehmann (HL) implementation of Wilcoxon Mann Whitney test. For example, for test of treatment effect, HL estimate of $\tau_t - \tau_c$ is obtained as

$$\frac{1}{2} \cdot [\text{Median}(Y_{11k_1} - Y_{12k_1}, Y_{22k_2} - Y_{21k_2} : k_1 = 1, \dots, n_1; k_2 = 1, \dots, n_2)]$$

81.4.1 Example

Dataset: CrossOverCaseData.cyd as described in Section 78.3

81 Analysis-Nonparametric Two-Sample

Purpose of the Analysis:

The purpose here is to compare the median morning peak expiratory flow rate (PEFR) between placebo and test drug. Let θ be the median difference between Drug and Placebo groups.

Analysis Steps:

1. Open the Dataset from the **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Two Samples > (Crossover Design) Wilcoxon-Mann-Whitney
3. In the ensuing dialog box choose the variables as shown below:

Analysis: Continuous Endpoint: Two-Sample Test - Crossover Design - Wilcoxon-Mann-Whitney

Data Set: CrossoverCaseData.cyd

Main Advanced

Trial Type: Superiority
 Effect Type: Carryover
 Treatment Specifications
 Treatment 1: Treatment 1
 Treatment 2: Treatment 2

Period ID: PeriodID
 Subject ID: subjectID
 Group ID: GroupID

Response Variable: Response

4. In the **Advanced** tab, leave the fields **By Variable 1** and **By Variable 2** blank and enter **0.95** for **Confidence Level**.
5. Click **OK** to start analysis. The output will be displayed in the main window.

Analysis: Continuous Response: Wilcoxon Mann Whitney Test: 2X2 Crossover

Hypothesis

$H_0 : \lambda_1 - \lambda_2 = 0$ Vs. $H_1 : \lambda_1 - \lambda_2 \neq 0$ for 2-sided test
 Either $H_1 : \lambda_1 - \lambda_2 < 0$
 Or $H_1 : \lambda_1 - \lambda_2 > 0$ for 1-sided test

Input Parameters

Data File: CrossoverCaseData.cyd
 Trial Type: Superiority
 Response Variable: Response
 Group ID: GroupID
 Subject ID: subjectID
 Period ID: PeriodID
 Test Type: t-Test
 Effect Type: Carryover
 Confidence Level: 0.95
 Treatment Assignment:

| Group ID | Period ID | |
|----------|-------------|-------------|
| | P1 | P2 |
| G1 | Treatment 1 | Treatment 2 |
| G2 | Treatment 2 | Treatment 1 |

Output

Response variable: (Response)
 Total no. of records: 112
 No. of records rejected: 0
 Summary of the Observed Data:

81 Analysis-Nonparametric Two-Sample

| Group ID | Statistic | Period ID | | Group Summary |
|----------------|-----------------|-----------|---------|---------------|
| | | 1 | 2 | |
| 1 | No. of Subjects | 27 | 27 | 27 |
| | Mean | 245.839 | 239.203 | 242.521 |
| | Std. Dev. | 82.78 | 81.697 | 81.529 |
| | Median | 235 | 221.842 | 228.421 |
| | Minimum | 67.778 | 70.278 | 67.778 |
| 2 | No. of Subjects | 29 | 29 | 29 |
| | Mean | 215.992 | 230.162 | 223.077 |
| | Std. Dev. | 72.629 | 73.942 | 72.993 |
| | Median | 201.905 | 225 | 212.068 |
| | Minimum | 104.444 | 128.947 | 104.444 |
| Period Summary | No. of Subjects | 56 | 56 | |
| | Mean | 230.382 | 234.521 | |
| | Std. Dev. | 78.43 | 77.197 | |
| | Median | 225.595 | 222.005 | |
| | Maximum | 67.778 | 70.278 | |

Treatment Summary:

| Statistic | Treatment 1 | Treatment 2 |
|-----------------|-------------|-------------|
| No. of Subjects | 56 | 56 |
| Mean | 237.72 | 227.183 |
| Std. Dev. | 78.008 | 77.315 |
| Median | 231.157 | 214.272 |
| Minimum | 67.778 | 70.278 |
| Maximum | 443.25 | 420.5 |

Summary of Test Statistic:

| Minimum | Maximum | Mean | Std. dev. | Observed Statistic | Standardized |
|---------|---------|-------|-----------|--------------------|--------------|
| 378 | 1161 | 769.5 | 60.986 | 835 | 1.074 |

Test of Hypothesis:

| Type | Standardized Statistic | Estimate of Median Difference | (1-Sided) | | (2-Sided) | 95% Confidence Interval(2-Sided)* | | | |
|------------|------------------------|-------------------------------|-----------|---------|-----------|-----------------------------------|-------------|---------|---------|
| | | | Tail | p-value | p-value | Lower Limit | Upper Limit | | |
| Asymptotic | 1.074 | | | 35.476 | G.E. | 0.141 | 0.283 | -34.357 | 114.804 |

*The Nonparametric estimates are based on Hodges-Lehmann's formulation.

In the **Output** section, the first part provides descriptive statistics for the two groups. The second table provides the treatment summary. The third table, labeled as **Test of Hypothesis**, provides results for statistical test of carryover effect. The observed value of statistic and standardized test statistic are 835 and 1.074, respectively. The p-value for two sided test is 0.283. Therefore, the carryover effect is not significant in this case and we can ignore this carryover effect.

Test for Treatment effect

Dataset: CrossOverCaseData.cyd as described in Section 78.3

Purpose of the Analysis:

The purpose here is to compare the median morning peak expiratory flow rate (PEFR) between placebo and test drug.

Analysis Steps:

1. Open the Dataset from the **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Two Samples > (Crossover Design) Wilcoxon-Mann-Whitney
3. In the ensuing dialog box choose the variables as shown below:

Click **OK** to start analysis. Upon completion of analysis, the output will be displayed in the main window. Scroll down to the end of the output. Output for statistical test of treatment effect is displayed in the last two tables.

Treatment Summary:

| Statistic | Treatment 1 | Treatment 2 |
|-----------------|-------------|-------------|
| No. of Subjects | 56 | 56 |
| Mean | 237.72 | 227.183 |
| Std. Dev. | 78.008 | 77.315 |
| Median | 231.157 | 214.272 |
| Minimum | 67.778 | 70.278 |
| Maximum | 443.25 | 420.5 |

Summary of Test Statistic:

| Minimum | Maximum | Mean | Std. dev. | Observed Statistic | Standardized |
|---------|---------|-------|-----------|--------------------|--------------|
| 378 | 1161 | 769.5 | 60.986 | 953 | 3.009 |

Test of Hypothesis:

| Type | Standardized Statistic | Estimate of Median Difference | (1-Sided) | | (2-Sided) | | 95% Confidence Interval(2-Sided)* | |
|------------|------------------------|-------------------------------|-----------|---------|-----------|-------------|-----------------------------------|--|
| | | | Tail | p-value | p-value | Lower Limit | Upper Limit | |
| Asymptotic | 3.009 | 11.494 | G.E. | 0.001 | 0.003 | 4.06 | 18.244 | |

*The Nonparametric estimates are based on Hodges-Lehmann's formulation.

The observed value of statistic and standardized test statistic are 953 and 3.009,

81 Analysis-Nonparametric Two-Sample

respectively. The p-value for two sided test is 0.003. Therefore, the treatment effect is significant in this case. In other words, the test drug significantly increases the median PEFR level over the placebo.

81.5 Test for Non-inferiority in Crossover Trial

81.5.1 Example

Let $\theta = \tau_t - \tau_c$. In a non-inferiority trial, we test the null hypothesis $H_0: \theta \leq \delta_0$ against the alternative hypothesis $H_1: \theta > \delta_0$ if $\delta_0 < 0$ or $H_0: \theta \geq \delta_0$ against the alternative hypothesis $H_1: \theta < \delta_0$ if $\delta_0 > 0$. **East** first subtracts δ_0 from all the observations pertaining to Test drug (T). Then the HL estimator is calculated as discussed in Section 81.4.

81.5.1 Example

Dataset: `pkfood.cyd` as described in Section 79.3.

Purpose of the Analysis:

Here the purpose is to compare the median AUC in regimen A with the regimen B considering the latter as reference and the former as test drug with non-inferiority margin (δ_0) of -5000 and one-sided type I error of 0.025. We will use θ_t and θ_c to denote the median AUC in regimen A and regimen B, respectively. Since $\delta_0 = -5000 < 0$, we are testing $H_0: \theta \leq \delta_0$ against the alternative hypothesis $H_1: \theta > \delta_0$, where $\theta = \theta_t - \theta_c$.

Analysis Steps:

1. Open the Dataset from the **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Two Samples > (Crossover Design) Wilcoxon-Mann-Whitney
3. In the ensuing dialog box choose the variables as shown below:

4. In the **Advanced** tab, leave the fields **By Variable 1** and **By Variable 2** blank and enter 0.975 for **Confidence Level**.
5. Click **OK** to start analysis. Upon completion of analysis, a new node with label **Analysis: Continuous Response: Difference of Means test for Crossover Data1** is added in the **Library** and the output will be displayed in the main window.

Analysis: Continuous Response: Wilcoxon Mann Whitney Test: 2X2 Crossover

Hypothesis

If $\delta_0 < 0$ then $H_0 : \theta_1 - \theta_2 \leq \delta_0$ Vs. $H_1 : \theta_1 - \theta_2 > \delta_0$
 If $\delta_0 > 0$ then $H_0 : \theta_1 - \theta_2 \geq \delta_0$ Vs. $H_1 : \theta_1 - \theta_2 < \delta_0$

where θ_1, θ_c are the medians of the Treatment Group and Control Group.

Input Parameters

Data File: pkfood.cyd
 Trial Type: Noninferiority
 Response Variable: AUC
 Group ID: SEQUENCE
 Subject ID: SUBJECT
 Period ID: PERIOD
 Test Type: t-Test
 Non Inferiority Margin (δ_0): -5000
 Confidence Level: 0.95
 Treatment Assignment:

| Group ID | Period ID | |
|----------|-----------|-----------|
| | P1 | P2 |
| G1 | Test | Reference |
| G2 | Reference | Test |

Output

Response variable: (AUC)
 Total no. of records: 40
 No. of records rejected: 0

81 Analysis-Nonparametric Two-Sample

Summary of the Observed Data:

| Group ID | Statistic | Period ID | | Group Summary |
|----------------|-----------------|-----------|----------|---------------|
| | | 1 | 2 | |
| AB | No. of Subjects | 10 | 10 | 10 |
| | Mean | 8542 | 9582.9 | 9062.45 |
| | Std. Dev. | 2978.945 | 3820.535 | 3376.8 |
| | Median | 8641 | 7859.5 | 8012 |
| | Minimum | 4686 | 5648 | 4686 |
| | Maximum | 14059 | 15765 | 15765 |
| BA | No. of Subjects | 10 | 10 | 10 |
| | Mean | 10620.3 | 8369.4 | 9494.85 |
| | Std. Dev. | 3468.083 | 3313.29 | 3497.233 |
| | Median | 9852.5 | 7607.5 | 9165.5 |
| | Minimum | 5911 | 5118 | 5118 |
| | Maximum | 18864 | 16510 | 18864 |
| Period Summary | No. of Subjects | 20 | 20 | |
| | Mean | 9581.15 | 8976.15 | |
| | Std. Dev. | 3322.27 | 3535.775 | |
| | Median | 9641.5 | 7859.5 | |
| | Minimum | 4686 | 5118 | |
| | Maximum | 18864 | 16510 | |

Treatment Summary:

| Statistic | Test | Reference |
|-----------------|----------|-----------|
| No. of Subjects | 20 | 20 |
| Mean | 8455.7 | 10101.6 |
| Std. Dev. | 3067.804 | 3590.908 |
| Median | 8091.5 | 9604.5 |
| Minimum | 4686 | 5648 |
| Maximum | 16510 | 18864 |

Summary of Test Statistic:

| Minimum | Maximum | Mean | Std. dev. | Observed Statistic | Standardized |
|---------|---------|------|-----------|--------------------|--------------|
| 55 | 155 | 105 | 13.229 | 146 | 3.099 |

Test of Hypothesis:

| Type | Standardized Statistic | Estimate of Median Difference | (1-Sided) | | 95% Confidence Interval(1-Sided)* | |
|------------|------------------------|-------------------------------|-----------|---------|-----------------------------------|-------------|
| | | | Tail | p-value | Lower Limit | Upper Limit |
| Asymptotic | 3.099 | -1427.25 | G.E. | 0.001 | -2432.5 | INF |

*The Nonparametric estimates are based on Hodges-Lehmann's formulation.

In the **Output** section, the first part provides descriptive statistics for the two groups. The second table provides the treatment summary. The table labeled as **Test of Hypothesis** provides results for statistical test of treatment effect. The estimated median difference is -1427.25. The observed value of test statistic and standardized test statistic are 146 and 3.099, respectively. The p-value for one-sided test is 0.001. This is the p-value associated with rejecting $H_0: \theta \leq -5000$ in favor of alternative hypothesis $H_1: \theta > -5000$. The one-sided 97.5% confidence interval is $(-2432, \infty)$. Since the lower limit of the confidence interval is greater than the non-inferiority margin of -5000, we can reject $H_0: \theta \leq -5000$ at one-sided 2.5% level of significance.

81.6 Test for Equivalence in Crossover Trial

Let $\theta = \tau_t - \tau_c$. The null hypothesis $H_0: \theta \leq \delta_L$ or $\theta \geq \delta_U$ is tested against the two-sided alternative hypothesis $H_1: \delta_L < \theta < \delta_U$ at level α , using the following two one-sided tests (TOST).

- Test1: $H_{0L}: \theta \leq \delta_L$ against $H_{1L}: \theta > \delta_L$ at level α
- Test2: $H_{0U}: \theta \geq \delta_U$ against $H_{1U}: \theta < \delta_U$ at level α

East subtracts θ_L and θ_U from all the observations pertaining to Test drug (T) for Test1 and Test2, respectively. Then the HL estimator is calculated as discussed in Section 81.4. To declare equivalence, both H_{0L} and H_{0U} need to be rejected.

81.6.1 Example

Dataset: `pkfood.cyd` as described in Section 79.3.

Purpose of the Analysis:

Here the purpose is to compare the median AUC in regimen A with the regimen B considering the latter as reference and the former as test drug with bioequivalence limits (δ_L, δ_U) as (-5000, 5000) and type I error rate not exceeding 0.05. Let θ_t and θ_c be the median AUC in regimen A and regimen B, respectively, and $\theta = \theta_t - \theta_c$.

We want to test the null hypothesis $H_0: \theta \leq -5000$ or $\theta \geq 5000$ against the alternative hypothesis $H_1: -5000 < \theta < 5000$.

Analysis Steps:

1. Open the Dataset from the **Samples** folder.
2. Choose the menu item:
**Analysis > (Continuous) Two Samples > (Crossover Design)
 Wilcoxon-Mann-Whitney**

81 Analysis-Nonparametric Two-Sample

3. In the ensuing dialog box choose the variables as shown below:

4. Click **OK** to start analysis. Upon completion of analysis, a new node with label **Analysis: Continuous Response: Difference of Means test for Crossover Data1** is added in the **Library** and the output will be displayed in the main window.

Analysis: Continuous Response: Difference of Means test for Crossover Data

$$H_{01} : \mu_1 - \mu_2 \leq \delta_L \quad \text{Or} \quad H_{02} : \mu_1 - \mu_2 \geq \delta_U$$

Vs.

$$H_1 : \delta_L < \mu_1 - \mu_2 < \delta_U$$

Data File: pkfood.cyd
 Trial Type: Equivalence
 Response Variable: AUC
 Group ID: SEQUENCE
 Subject ID: SUBJECT
 Period ID: PERIOD
 Test Type: t-Test
 Lower Equivalence Limit(δ_L): -5000
 Upper Equivalence Limit(δ_U): 5000
 Confidence Level (for Equivalence): 0.9
 Treatment Assignment:

| Group ID | Period ID | |
|----------|-----------|-----------|
| | 1 | 2 |
| AB | Test | Reference |
| BA | Reference | Test |

Output

Response Variable: (AUC)
 Total Number of Records: 40
 Number of Records Rejected: 0

| | | | | |
|--|---------|-------|-------|--|
| | Maximum | 18864 | 16510 | |
|--|---------|-------|-------|--|

Treatment Summary:

| Statistic | Test | Reference |
|-----------------|----------|-----------|
| No. of Subjects | 20 | 20 |
| Mean | 8455.7 | 10101.6 |
| Std. Dev. | 3067.804 | 3590.908 |
| Median | 8091.5 | 9604.5 |
| Minimum | 4686 | 5648 |
| Maximum | 16510 | 18864 |

Summary of Test Statistic:

| Statistic | Minimum | Maximum | Mean | Std. dev. | Observed Statistic | Standardized |
|-----------|---------|---------|------|-----------|--------------------|--------------|
| Under H01 | 55 | 155 | 105 | 13.229 | 146 | 3.099 |
| Under H02 | 55 | 155 | 105 | 13.229 | 55 | -3.78 |

Test of Hypothesis:

| Type | Standardized Statistic | | Estimate of Median Difference | p-value | | 90% Confidence Interval(2-Sided)* | |
|------------|------------------------|-----------|-------------------------------|-----------|-----------|-----------------------------------|-------------|
| | Under H01 | Under H02 | | Under H01 | Under H02 | Lower Limit | Upper Limit |
| Asymptotic | 3.099 | -3.78 | -1427.25 | 9.699E-4 | 7.853E-5 | -2432.5 | -296.5 |

In the **Output** section, the first part provides descriptive statistics for the two groups. The second table provides the treatment summary. The third table labeled as **Test of Hypothesis** provides results for statistical test of treatment effect.

The estimated median difference is -1427.25. The observed values of test statistic and standardized test statistic are 146 and 3.099 respectively for the H_{01} and 55 and -3.78, respectively for the H_{02} .

82

Analysis-ANOVA

Sometimes the goal of a clinical trial is to compare more than two treatment arms. For example, in a phase II dose-finding study multiple doses of an experimental drug may be compared with placebo or some other control. The most popular method applied to this kind of data is Analysis of Variance (ANOVA). Designing of such studies with continuous endpoint is discussed in chapter 14.

In this section, we focus on how to analyze data collected from such studies using ANOVA in **East**. As an alternative to ANOVA, you can analyze these kind of data using multiple comparison procedures as well and this is discussed in chapter 84.

82.1 Example: One Way ANOVA

In a one-way Analysis of Variance (ANOVA) test, we wish to test the equality of means across R independent groups.

Let X_{ij} indicate the response from j^{th} unit of i^{th} group; $i = 1, \dots, R, j = 1, \dots, n_i$. Further assume, $X_{ij} \sim N(\mu_i, \sigma^2)$; $i = 1, \dots, R$. In one-way ANOVA, the goal is to compare the null hypothesis

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_R$$

against the alternative hypothesis

$$H_1 : \text{for at least one pair } (i, i'), \mu_i \neq \mu_{i'}, \text{ where } i, i' = 1, 2, \dots, R.$$

Dataset: leucolyte.cyd.

Data Description

Kontula K et al (1980, 1982) conducted a study to compare the number of glucocorticoid receptor (GR) sites per leukocyte cell in 5 groups of patients:

1. Group 1: normal subjects
2. Group 2: patients with hairy-cell leukemia
3. Group 3: patients with chronic lymphatic leukemia
4. Group 4: patients with chronic myelocytic leukemia
5. Group 5: patients with acute leukemia

Purpose of the Analysis:

The goal is to compare the mean GR sites per leukocyte cell among the 5 groups of patients.

Let μ_i denote the mean number of GR sites per leukocyte cell in i^{th} group of subjects/patients; $i = 1, \dots, R$. To test the null hypothesis

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ with 5% level of significance.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous): Many Samples > (Factorial Design) One-Way ANOVA
3. In the **Main** tab, select **Group** as **Factor** and **GR** as **Response**. Leave the check box for **Contrast** unchecked.

Analysis: Continuous Endpoint Many-Sample Test - Factorial Design - One-Way ANOVA

Data Set: leukocyte.cyd

Main Advanced

Factor: Group Contrast

Response: GR

4. In the **Advanced** tab, you can select up to 2 grouping variables. If only one grouping variable is selected, then a different analysis will be displayed for each level of the selected grouping variable. If two grouping variables are selected then **East** will display different analysis for each combination of levels of two grouping variables. In this analysis, leave the fields **By Variable 1** and **By Variable 2** blank.

Analysis: Continuous Endpoint Many-Sample Test - Factorial Design - One-Way ANOVA

Data Set: leukocyte.cyd

Main Advanced

By Variable 1:

By Variable 2:

82 Analysis-ANOVA

- Click **OK** to start the analysis. After completion of the analysis, the output is displayed in the main window.

Analysis: ANOVA: One Way

Hypothesis

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_r$$

Vs.

H_1 : At least one μ_i differs , for $i=1,2,3 \dots r$ where r is the number of treatment groups

Input Parameters

Data File: leukocyte.cyd

Factor: Group

Response: GR

Output

Total no. of records: 37

No. of records rejected: 0

ANOVA Table:

| Source | DF | Sum of Squares | Mean Square | F Statistic | p-value |
|-----------|----|----------------|--------------|-------------|---------|
| Group | 4 | 353735187.391 | 88433796.848 | 4.308 | 0.007 |
| Residuals | 32 | 656834693.69 | 20526084.178 | | |
| Total | 36 | 1.011E9 | | | |

Summary Table:

| Overall Mean | Root MSE | R-Square | Adjusted R-Square | Coeff. of Variation |
|--------------|----------|----------|-------------------|---------------------|
| 7107.568 | 4530.572 | 0.35 | 0.269 | 63.743 |

The last section is the **Output**. From the **ANOVA Table**, the significance level for Group effect is 0.007. Therefore, the conclusion is to reject

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ at 5% level of significance.

82.2 Example: One Way Contrast

Often one may be interested in testing significance of linear combination of group means instead of just finding the difference in group means. This can be done through the use of contrast. A contrast of the population means is a linear combination of the

μ_i 's.

For the given scalars, $\{c_i : i = 1, \dots, R\}$, $C = \sum c_i \mu_i$ denotes a linear contrast of population mean if $\sum c_i = 0$. For a single contrast test of many means in a one-way ANOVA, the null hypothesis that we wish to test is:

$$H_0 : \sum c_i \mu_i = 0$$

versus a 2-sided alternative

$$H_1 : \sum c_i \mu_i \neq 0$$

Or a 1-sided alternative

$$H_1 : \sum c_i \mu_i < 0 \text{ or } H_1 : \sum c_i \mu_i > 0.$$

Dataset: *leucolyte.cyd* as described in section 82.1

Purpose of the Analysis:

Let μ_i denote the mean number of GR sites per leukocyte cell in i^{th} group of subjects/patients; $i = 1, \dots, R$.

We are interested in comparing the mean number of GR sites in normal subjects (Group 1) with the average of mean number of GR sites in all the remaining groups. That is, we are interested in comparing:

$$\mu_1 \text{ with } \frac{\mu_2 + \mu_3 + \mu_4 + \mu_5}{4}.$$

To do this comparison, test the following null hypothesis:

$$H_0 : \frac{1}{4}\mu_2 + \frac{1}{4}\mu_3 + \frac{1}{4}\mu_4 + \frac{1}{4}\mu_5 - \mu_1 = 0$$

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous): Many Samples > (Factorial Design) One-Way ANOVA

82 Analysis-ANOVA

- In the **Main** tab, select **Group** as **Factor** and **GR** as **Response**. Select the check box for **Contrast**. A table is displayed below it. Enter -1 , 0.25 , 0.25 , 0.25 and 0.25 in **Coefficient** column for the 5 categories.

Analysis: Continuous Endpoint Many-Sample Test - Factorial Design - One-Way ANOVA

Data Set: leukocyte.cyd

Main Advanced

Factor: Group

Response: GR

Contrast

| Category | Coefficient |
|----------|-------------|
| 1 | -1 |
| 2 | 0.25 |
| 3 | 0.25 |
| 4 | 0.25 |
| 5 | 0.25 |

- Click **OK** to start the analysis. After completion of the analysis, the output is

displayed in the main window.

Analysis: One-way Contrast ANOVA

Hypothesis

$H_0 : t_1 \mu_1 + t_2 \mu_2 + \dots + t_m \mu_m = 0$

Vs.

$H_1 : t_1 \mu_1 + t_2 \mu_2 + \dots + t_m \mu_m \neq 0$ where $\sum t = 0$ and m is the levels of the factor

Input Parameters

Data File: leukocyte.cyd

Factor: Group

Response: GR

Contrast:

| Levels | Coefficient |
|--------|-------------|
| 1 | -1 |
| 2 | 0.25 |
| 3 | 0.25 |
| 4 | 0.25 |
| 5 | 0.25 |

Output

Total no. of records: 37

No. of records rejected: 0

ANOVA Table:

| Source | DF | Sum of Squares | Mean Square | F Statistic | p-value |
|-----------|----|----------------|--------------|-------------|---------|
| Group | 4 | 353735187.391 | 88433796.848 | 4.308 | 0.007 |
| Residuals | 32 | 656834693.69 | 20526084.178 | | |
| Total | 36 | 1.011E9 | | | |

Table of contrast:

| Source | DF | Sum of Squares | Mean Square | F Statistic | t Statistic | 1-Sided p-value | | 2-Sided p-value |
|----------|----|----------------|---------------|-------------|-------------|-----------------|-------------|-----------------|
| | | | | | | F Statistic | t Statistic | t Statistic |
| Contrast | 1 | 107544481.594 | 107544481.594 | 5.239 | 2.289 | 0.007 | 0.014 | 0.029 |

Summary Table:

| Overall Mean | Root MSE | R-Square | Adjusted R-Square | Coeff. of Variation |
|--------------|----------|----------|-------------------|---------------------|
| 7107.568 | 4530.572 | 0.35 | 0.269 | 63.743 |

The result of the analysis is divided into three sections. Test for contrast is displayed in the third section labeled as **Output**. The 2-sided p-value for testing the $H_0 : \frac{1}{4}\mu_2 + \frac{1}{4}\mu_3 + \frac{1}{4}\mu_4 + \frac{1}{4}\mu_5 - \mu_1 = 0$ is 0.029. Therefore, we can conclude that mean number of average GR sites in normal subjects (Group 1) is significantly different than the average of mean number of average GR sites in all the remaining groups with observed significance level of 0.029.

82 Analysis-ANOVA

82.3 Example: One Way Repeated Measures (Constant Correlation) ANOVA

As with the one-way ANOVA discussed in section 82.1, the repeated measures ANOVA also tests for equality of population means. However, in a repeated measures setting, the subjects are measured repetitively over time. Therefore, the measurements observed within a same subject are correlated. This correlation between observations from the same subject needs to be accounted for in ANOVA. The constant correlation assumption refers to the equal correlation between any pair of observations from a subject. Denote this constant correlation by ρ . The **Repeated ANOVA** module in **East** allows to test the effects of subject and time as well as test for contrast in subject means.

Dataset: Body_wight.cyd

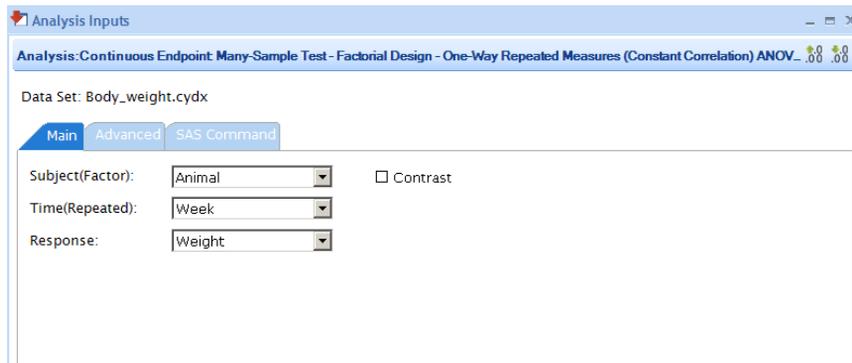
Data Description

Here consider the body weight data of guinea pigs given by Crowder and Hand (1989, p. 27). The data was obtained to investigate the effect of vitamin E diet supplement on the growth of guinea pigs. For each animal the body weight (in gram) were recorded at the end of 1, 3, 4, 5, 6, and 7 weeks. All animals were given a growth-inhibiting substance during week 1 and the vitamin E therapy was started at the beginning of week 5. Three groups of animals, numbering five in each, received respectively zero, low and high doses of vitamin E.

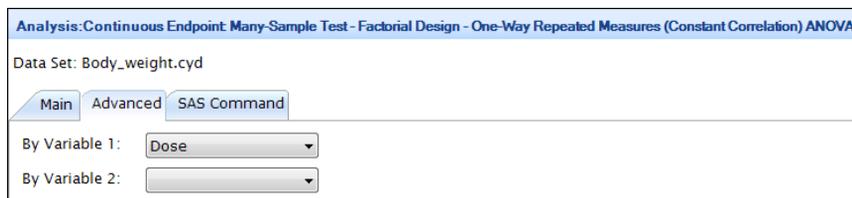
Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous): Many Samples > (Factorial Design) One-Way Repeated Measures
3. In the **Main** tab, select **Animal** as **Subject(Factor)**, **Week** as **Time(Repeated)**

and **Weight** as **Response**. Leave the check box for **Contrast** unchecked.



- In the **Advanced** tab, you can select up to 2 grouping variables. If only one grouping variable is selected, then a different analysis will be displayed for each level of the selected grouping variable. If two grouping variables are selected then **East** will display different analysis for each combination of levels of two grouping variables. Select **Dose** as **By Variable 1**.



- The third tab is **SAS Command** where you can put SAS code for more sophisticated analysis. For this example, do not make any changes in this tab.
- Click **OK** to start the analysis. The output is displayed in the main window.

82 Analysis-ANOVA

ANOVA for all the three dose groups is displayed in the **Output** section.

| By: | | Dose = No | | | |
|---------------------------------|----------|----------------|-------------------|---------------------|-----------|
| Total no. of records: | | 30 | | | |
| No. of records rejected: | | 0 | | | |
| ANOVA Table: | | | | | |
| Source | DF | Sum of Squares | Mean Square | F Statistic | p-value |
| Animal | 4 | 38824.8 | 9706.2 | 12.373 | 3.166E-5 |
| Week | 5 | 41001.867 | 8200.373 | 10.454 | 4.841E-5 |
| Residuals | 20 | 15688.8 | 784.44 | | |
| Total | 29 | 95515.467 | | | |
| Summary Table: | | | | | |
| Overall Mean | Root MSE | R-Square | Adjusted R-Square | Coeff. of Variation | |
| 539.133 | 28.008 | 0.836 | 0.762 | 5.195 | |
| By: | | Dose = Low | | | |
| Total no. of records: | | 30 | | | |
| No. of records rejected: | | 0 | | | |
| ANOVA Table: | | | | | |
| Source | DF | Sum of Squares | Mean Square | F Statistic | p-value |
| Animal | 4 | 42140.533 | 10535.133 | 17.403 | 2.689E-6 |
| Week | 5 | 63185.867 | 12637.173 | 20.875 | 2.587E-7 |
| Residuals | 20 | 12107.467 | 605.373 | | |
| Total | 29 | 117433.867 | | | |
| Summary Table: | | | | | |
| Overall Mean | Root MSE | R-Square | Adjusted R-Square | Coeff. of Variation | |
| 572.267 | 24.604 | 0.897 | 0.851 | 4.299 | |
| By: | | Dose = High | | | |
| Total no. of records: | | 30 | | | |
| No. of records rejected: | | 0 | | | |
| ANOVA Table: | | | | | |
| Source | DF | Sum of Squares | Mean Square | F Statistic | p-value |
| Animal | 4 | 24468.867 | 6117.217 | 25.722 | 1.222E-7 |
| Week | 5 | 48129.5 | 9625.9 | 40.476 | 8.642E-10 |
| Residuals | 20 | 4756.333 | 237.817 | | |
| Total | 29 | 77354.7 | | | |
| Summary Table: | | | | | |
| Overall Mean | Root MSE | R-Square | Adjusted R-Square | Coeff. of Variation | |
| 565.9 | 15.421 | 0.939 | 0.911 | 2.725 | |

The output suggests that the effect of animal and week is highly significant in all the three dose groups.

82.4 Example: Two Way ANOVA

In a two-way ANOVA, there are two factors to consider, say A and B. Let X_{ijk} indicate the response from k^{th} replication of i^{th} level of A and j^{th} level of B; $i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, n$. Further we assume, $X_{ijk} \sim N(\mu_{ij}, \sigma^2)$; $i = 1, \dots, R$.

In two-way ANOVA, the goal is to test the following null hypotheses

- **Test for main effect of factor A.** H_0 : The group means for all the levels of factor A is same.
- **Test for main effect of factor B.** H_0 : The group means for all the levels of factor B is same.
- **Test for interaction effect of A and B.** H_0 : The effect of A remains same for all levels of B or the effect of B remains same for all levels of A.

Dataset: `Body_wight.cyd` as described in Section 82.3

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous): Many Samples > (Factorial Design) Two-Way ANOVA
3. In the **Main** tab, select **Dose** as **Factor1**, **Week** as **Factor2** and **Weight** as **Response**. Leave the **Interaction Effect** check box checked for test of interaction effect between the two factors.

Analysis: Continuous Endpoint Many-Sample Test - Factorial Design - Two-Way ANOVA

Data Set: Body_wight.cyd

Main
Advanced

Factor1:

Factor2:

Response:

Interaction Effect

82 Analysis-ANOVA

4. Click **OK** to start the analysis. The output is displayed in the main window now.

Analysis: ANOVA:Two way

Hypothesis

$H_{01} : \mu_{A1} = \mu_{A2} = \dots = \mu_{Aa}$ Vs. $H_{11} : \text{At least one } \mu_{Ai} \text{ differs. For } i = 1, 2, \dots, a$, where a is the number of levels of the 1st factor 'A'.
 $H_{02} : \mu_{B1} = \mu_{B2} = \dots = \mu_{Bb}$ Vs. $H_{12} : \text{At least one } \mu_{Bj} \text{ differs. For } j = 1, 2, \dots, b$, where b is the number of levels of the 2nd factor 'B'.
 $H_{03} : \text{There is no interaction term in the model. vs } H_{13} : \text{There is an interaction term in the model.}$

Input Parameters

Data File: Body_weight.cydx
 Factor1: Dose
 Factor2: Week
 Response: Weight

Total Number of Records: 90

Number of Records Rejected: 0

ANOVA Table:

| Source | DF | Sum of Squares | Mean Square | F Statistic | p-value |
|-------------|----|----------------|-------------|-------------|----------|
| Dose | 2 | 18548.067 | 9274.033 | 4.839 | 0.011 |
| Week | 5 | 142554.5 | 28510.9 | 14.877 | 5.19E-10 |
| Interaction | 10 | 9762.733 | 976.273 | 0.509 | 0.878 |
| Residuals | 72 | 137986.8 | 1916.483 | | |
| Total | 89 | 308852.1 | | | |

Summary Table:

| Overall Mean | Root MSE | R-Square | Adjusted R-Square | Coeff. of Variation |
|--------------|----------|----------|-------------------|---------------------|
| 559.1 | 43.778 | 0.553 | 0.448 | 7.83 |

The p-values associated with main effect of Dose and Week are 0.011 and 5.19×10^{-10} . These p-values suggest significant main effect for Dose and Week. The interaction between Dose and Week is not significant (p-value = 0.878).

83

Analysis-Regression Procedures

This chapter demonstrates how to run regression analysis in **East**. **East** can perform multiple linear regression, repeated measure regression and fit linear mixed effect (LME) model on data obtained from 2×2 crossover design. The LME model on 2×2 crossover data can be fit either to test for difference of means or ratio of means. These are discussed in sections 83.1, 83.2, 83.2, 83.3 and 83.4. In addition to fitting the regression coefficients, **East** can also be used to:

- perform significance testing of regression coefficients using Wald test
- perform 1st order autocorrelation in residuals using Durbin-Watson test
- compute collinearity diagnostics
- compute different types of residuals
- compute influential statistics
- compute predicted values
- perform variable selection

83.1 Example: Multiple Linear Regression

Dataset: **Werner.cyx** as described in Section 73.4.2.

Purpose of the Analysis:

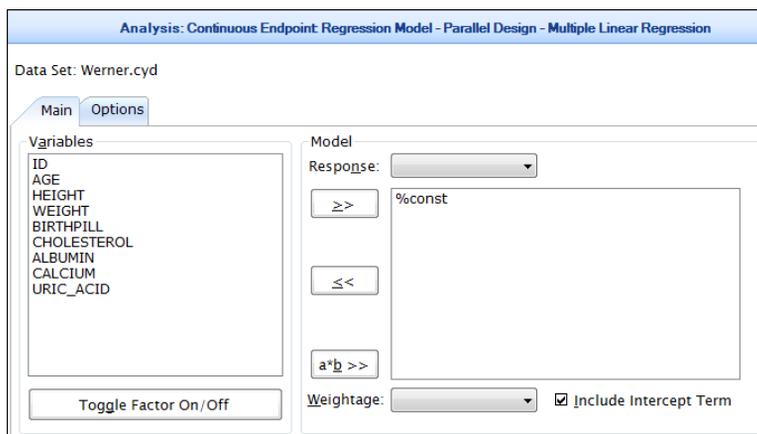
In this example, the multiple regression technique is used to find relationship of the variable Cholesterol with the other variables.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Regression > (Parallel Design) Multiple Linear Regression
This will display several input fields associated with regression analysis in the

83 Analysis-Regression Procedures

main window.



In the **Main** tab, there are two boxes – **Variables** and **Model**. In the **Variable** box, all the numeric variables in the dataset are displayed. The **Toggle Factor On/Off** button can change the status of a variable between numeric and factor variable.

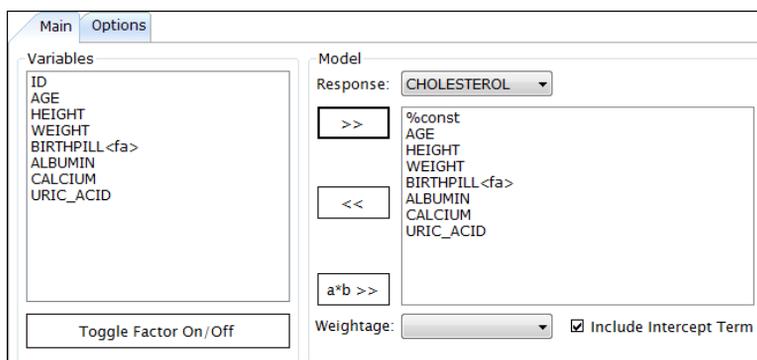
- For example, select the **BIRTHPILL** variable and click **Toggle Factor On/Off** button. This will declare the **BIRTHPILL** variable as factor variable.



The suffix **<fa>** is added to **BIRTHPILL** in the list of variables. The suffix **<fa>** indicates that the **BIRTHPILL** will be treated as factor variable in the multiple linear regression, if included as predictor. We can declare any variable in the **Variables** box as factor variable. For this example, only consider

BIRTHPILL as factor variable.

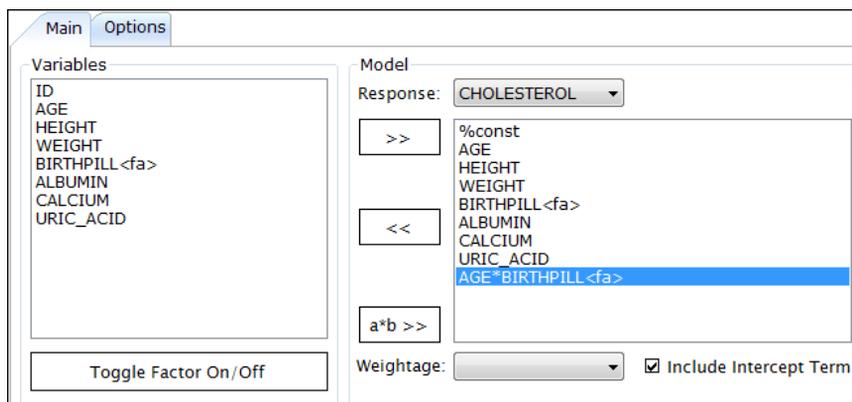
- In the box **Model**, choose **CHOLESTEROL** as **Response** variable. Below this, there is a box with only entry %const. This is where all the predictors in the model has to be included. The term %const refers to the intercept (β_0). To remove this term clear the checkbox **Include Intercept Term**. In the absence of this term, **East** will perform multiple regression analysis without any intercept. For this example, keep this term. Include all the variables except ID in this box. To include a variable in this box, select the variable from the list of variables in the **Variable** box then click button. To de-select a selected term click button.



- Now, we might believe that the effect of birth pill use on cholesterol level varies with age. In other words, there might be interaction between age and birth pill use. To include the interaction effect, select **Age** and **BIRTHPILL < fa >** in the **Variable** box using **Ctrl key**, and click button. This adds the term

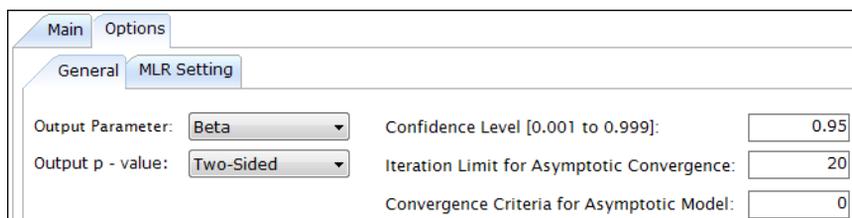
83 Analysis-Regression Procedures

AGE*BIRTHPILL<fa> in the predictor variable list.



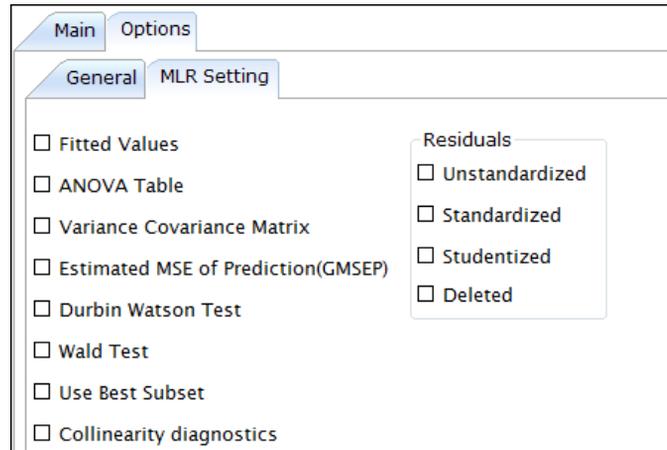
The interaction effect AGE*BIRTHPILL<fa> is an example of first order interaction. In **East**, you can also include interaction effect of higher orders. To include interaction effect, select all the variables that are interacting and click **a*b >>** button.

- Click the **Options** tab. There are two sub-tabs within this tab – **General** and **MLR Setting**.



- In the **General** sub-tab, leave the default choice of **Beta** for **Output Parameter** and **Two-Sided** for **Output p - value**.

In the **MLR Setting** sub-tab, there is a list of checkboxes in two columns.

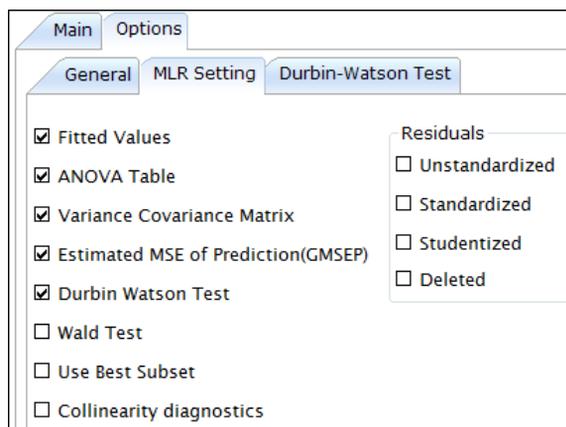


The purpose of the checkboxes is given below:

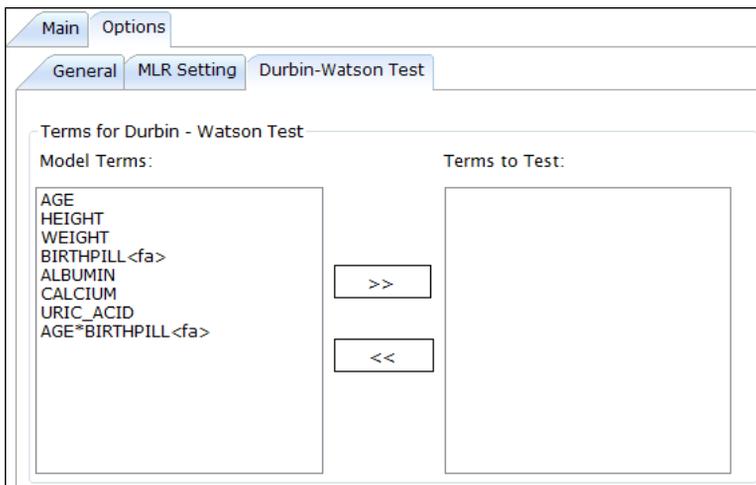
- **Fitted Values:** Calculates the fitted values.
- **ANOVA:** Includes ANOVA table in the regression output.
- **Variance Covariance Matrix:** Includes variance covariance matrix for estimated regression coefficients in the regression output.
- **Estimated MSE of Prediction (GMSEP):** Includes mean squared error (MSE) (or variance of residuals) and mean squared error of prediction (MSEP) in the regression output.
- **Durbin Watson Test:** Performs the test for first order autocorrelation among residuals and the results are displayed in the regression output.
- **Wald Test:** Performs the Wald test for significance of regression coefficients and the results are displayed in the regression output.
- **Use Best Subset:** Performs the subset selection using backward elimination, forward selection, sequential replacement, stepwise selection or exhaustive search technique.
- **Collinearity diagnostics:** Provides collinearity diagnostics such as Eigenvalues of $(X^T X)^{-1}$ and condition numbers. Before calculation of Eigenvalues, $X^T X$ is scaled to have 1's on the diagonal. The condition numbers are the square roots of the ratio of the largest Eigenvalue to each individual Eigenvalue. The largest condition number is the condition number of the scaled X matrix.
- **Unstandardized Residuals:** Calculates the residuals.

83 Analysis-Regression Procedures

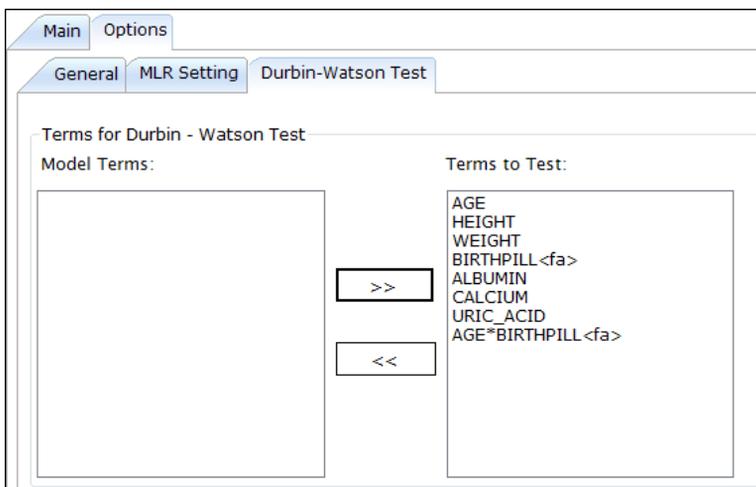
- **Standardized Residuals:** Calculates the standardized residuals.
 - **Studentized Residuals:** Calculates the studentized residuals.
 - **Deleted Residuals:** Calculates the deleted residuals deleting the corresponding observation.
8. Select the first 4 checkboxes – **Fitted Values**, **ANOVA Table**, **Variance Covariance Matrix** and **Estimated MSE of Prediction (MSEP)**. Then select the checkbox for **Durbin Watson Test**. A third sub-tab **Durbin-Watson Test** is added.



9. Click the **Durbin-Watson Test** tab.



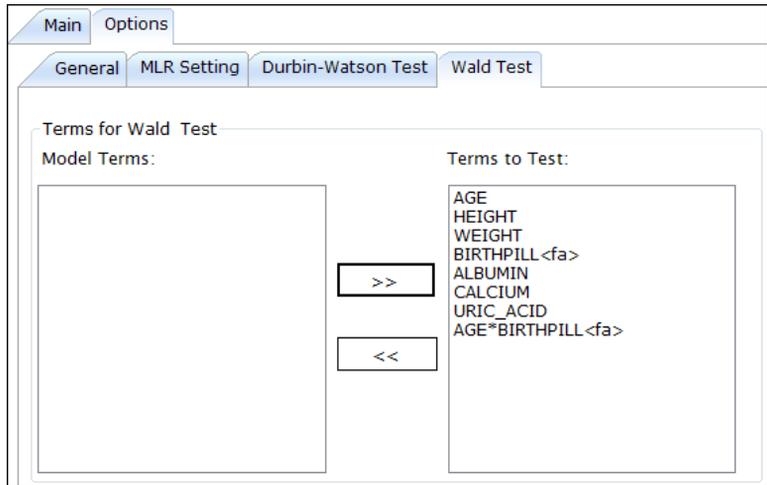
10. In this tab select or de-select the terms for the Durbin-Watson test using and buttons. Select all the variables from **Model Terms** to **Terms to test**.



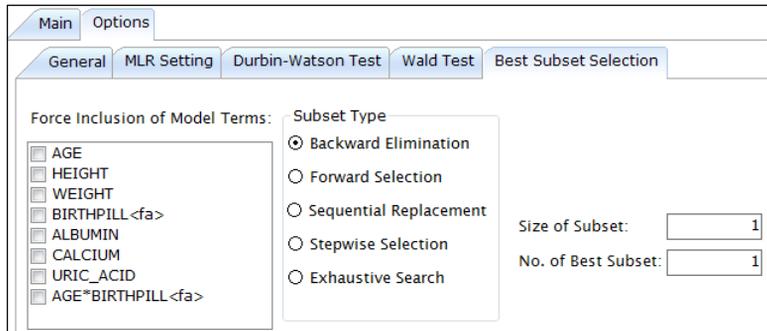
11. Come back to **MLR Setting** sub-tab. Now check the box for Wald Test. This

83 Analysis-Regression Procedures

will add a new sub-tab labeled as **Wald Test**. Click on this sub-tab and select all the variables from **Model Terms** to **Terms to test**.



12. Come back to **MLR Setting** sub-tab and select the **Use Best Subset** check box. This will add a new sub-tab labeled as **Best Subset Selection**. Click on this sub-tab.



13. The first column is a box with label **Force Inclusion of Model Terms** and it includes all the model terms. Here select the variables that needs to be retained forcefully in the model and selection method will be applied on the remaining terms. In this example, use of Birthpill is an important factor that influences cholesterol level. Therefore, select **BIRTHPILL< fa >**. **BIRTHPILL< fa >**

will skip the variable selection procedure and it will always be part of the best subset of variables.

14. In the second column choose the method of subset selection. The choices are **Backward Elimination**, **Forward Selection**, **Sequential Replacement**, **Stepwise Selection** and **Exhaustive Search**.

In *Forward Selection* procedure, the model starts with the constant term (or with the forced terms) and it keeps adding new terms in each step that gives largest reduction in sum of squares of the residuals (SSE). The method stops when inclusion of none of the additional terms results in sufficient amount of reduction in SSE. In *Backward Elimination* procedure, the model starts with all the available terms and then eliminates a variable in each step that provides minimum reduction in SSE. The method stops when the reduction in SSR due to dropping of any variable exceeds some threshold amount. The *Stepwise Selection* procedure is like *Forward Selection* except that at each step dropping of variables is also considered as in *Backward Elimination* procedure. At each step, the F value is calculated for each variable. If S indicates the set of all the variables in the subset in the current step, then for i^{th} variable, F value, F_i is calculated as follows:

$$F_i = \frac{SSR(S \cup \{i\}) - SSR(S)}{MSE(S \cup \{i\})} \quad i \notin S$$

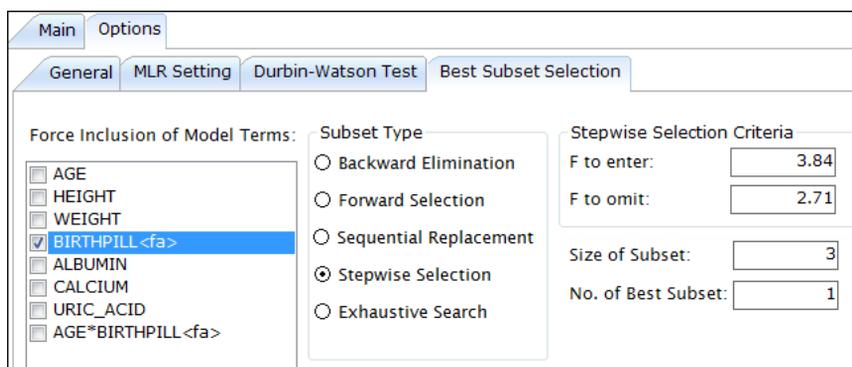
$$F_i = \frac{SSR(S) - SSR(S - \{i\})}{MSE(S)} \quad i \in S$$

For $i \notin S$, the i^{th} variable is entered in subset if $F_i > F_{in}$. For $i \in S$, the i^{th} variable will be dropped from the subset if $F_i < F_{out}$. In the *sequential replacement* procedure, for a given number of variables, variables are sequentially replaced and replacements that improve performance are retained. This approach checks whether any of the variables selected in the current model can be replaced with another variable to give a smaller residual sum of squares. In *exhaustive search* procedure, all possible subset are evaluated and the subset with largest *adjusted R^2* is chosen.

15. Select **Stepwise Selection**. A box labeled as **Stepwise Selection Criteria** appears. Keep the default values 3.84 and 2.71 for **F to enter** and **F to omit**. These two values corresponds to F_{in} and F_{out} , as explained above. There are two fields – **Size of Subset** and **No. of Best Subset**. In **Size of Subset**, enter the maximum allowed size of the subset. For example, the data contains 20

83 Analysis-Regression Procedures

independent variables, but you want to restrict the search to subsets which have a maximum of 7 variables. In this case, specify the size of subset as 7. In the field No. of Best Subset, specify the number of top models of each subset size which will be included in the output. In this example, enter 3 and 1 for these two fields. With this specification, we are looking for subset of variables of size 2 and 3 of which one of the term must be BIRTHPILL<fa>. The subset of size 1 will not be displayed as we have already specified one variable to enter forcefully in all the subset and thus the subset of 1 variable does not require any variable to come from subset selection procedure.



The screenshot shows the 'Best Subset Selection' dialog box with the following settings:

- Force Inclusion of Model Terms:**
 - AGE
 - HEIGHT
 - WEIGHT
 - BIRTHPILL<fa>
 - ALBUMIN
 - CALCIUM
 - URIC_ACID
 - AGE*BIRTHPILL<fa>
- Subset Type:**
 - Backward Elimination
 - Forward Selection
 - Sequential Replacement
 - Stepwise Selection
 - Exhaustive Search
- Stepwise Selection Criteria:**
 - F to enter: 3.84
 - F to omit: 2.71
 - Size of Subset: 3
 - No. of Best Subset: 1

16. Click the **MLR Setting** sub-tab and select the check box **Collinearity**

diagnostics. A new box labeled as **Parameters for Collinearity** appears.

The screenshot shows the 'Options' dialog box in the East 6.4 software. The 'Durbin-Watson Test' tab is selected. In the 'Parameters for Collinearity' section, the 'Multi Collinearity Criterion' is set to 0.05 and the 'No. of Collinearity Component' is set to 2. The 'Residuals' section is also visible, with checkboxes for Unstandardized, Standardized, Studentized, and Deleted.

17. In the **Parameters for Collinearity** box, specify two parameters – **Multi Collinearity Criterion** and **No. of Collinearity Component**.

The **Multi Collinearity Criterion** refers to the value that controls how small the determinant of the matrix (that is inverted to compute the coefficient estimates) is allowed to be. This value must be less than 1 and greater than 0. The latter refers to number of collinearity components we want **East** to display. This number can be between 2 and the number of terms in the model including intercept, if any. In this case choose a number of collinearity components between 2 to 10. **East** specifies default values of 0.05 for **Multi Collinearity Criterion** and 2 for **No. of Collinearity Component**. For this example, keep these two values unchanged.

18. In the **Residual** box, check all the 4 types of residuals - **Unstandardized**, **Standardized**, **Studentized** and **Deleted**. Upon checking any of these residuals,

83 Analysis-Regression Procedures

a new box appears labeled as **Influential statistics**.

Unstandardized residuals (r_i) are obtained simply by subtracting predicted value of response variable (\hat{Y}_i) from the observed value (Y_i) for each observations.

That is,

$$r_i = Y_i - \hat{Y}_i$$

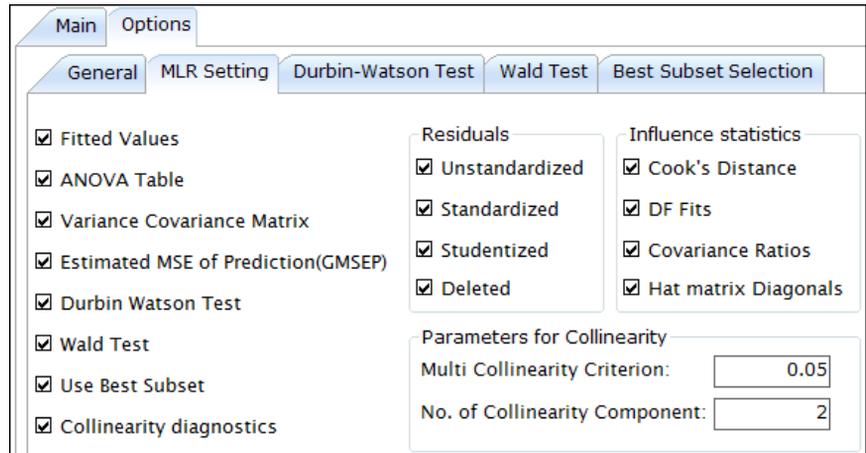
Standardized residuals are the *Unstandardized residuals* divided by the root mean square error (RMSE). Even though this is called as *standardized residuals*, this is not standardized in true sense, because the residuals does not have equal variance (even with constant variance assumption). The variance for i^{th} residual is estimated as $\sigma^2(1 - h_{ii})$, where, h_{ii} is the i^{th} diagonal element of hat matrix, H . It is more appropriate to standardize the residuals as follows:

$$\frac{r_i}{\hat{\sigma} \sqrt{(1 - h_{ii})}}$$

This is known as studentized residuals. Cook and Weisberg refer to this as external studentization. These residuals have t-distributions with $N - K$ degrees of freedom, so any residual with absolute value exceeding 3 usually requires attention. The *deleted residuals* are obtained as $Y_i - \hat{Y}_i^{-i}$, where \hat{Y}_i^{-i} indicates the predicted value of Y_i where prediction is done excluding i^{th} observation.

19. Check all the influential statistics in the **Influence statistics** box next to the

Residuals.



Cook's distance is an overall measure of the impact of i^{th} datapoint on the estimated regression coefficients. This is defined as:

$$D_i = \frac{\sum_{i=1}^N (\hat{Y}_i - \hat{Y}_i^{-i})^2}{K \hat{\sigma}}$$

D_i 's are distributed as $F(K, N - K)$. If $D_i < F(0.2, K, N - K)$ then the i^{th} case has only little apparent influence on the fitted values. On the other hand, if $D_i > F(0.5, K, N - K)$, the i^{th} observation should be considered influential. *DFFITs* are calculated for each observation. For i^{th} observation this is defined as:

$$(DFFIT)_i = \frac{\hat{Y}_i - \hat{Y}_i^{-i}}{\hat{\sigma}^{-i} \sqrt{h_{ii}}}$$

where $\hat{\sigma}^{-i}$ is the RMSE or estimate of σ obtained excluding i^{th} observation. Kutner et al. (2004) suggested to consider a case as influential if the absolute value exceeds 1 for small to medium size data and $2\sqrt{K/N}$ for large datasets. The measure *Covariance Ratios* reflects the change in the variance-covariance matrix of the estimated coefficients when the i^{th} observation is omitted. For i^{th} observation, it is obtained as ratio of determinant of covariance matrix of estimate of β excluding i^{th} observation to the determinant of covariance matrix of estimate of β including all the observations. It is suggested that $|CR_i - 1| \geq 3K/N$ warrants further investigation.

83 Analysis-Regression Procedures

Hat matrix diagonals simply refers to i^{th} diagonal element, h_{ii} , of hat matrix, H , for i^{th} observation. This measure is also known as the leverage of the i^{th} observation. The diagonal elements sum to the number of parameters being fitted. Any value greater than $2K/N$ suggests further investigation.

- Click **OK** to start the analysis. After completion of the analysis a new node with title **Analysis: Multiple Linear Regression1** is added to the **Library**. It has two sub-nodes – **MLR-Residuals1** and **MLR-Best Subset Selection1**.

Library

- Root
 - Birthpill Regression
 - Werner.cyd
 - Analysis: Multiple Linear Regression1
 - MLR-Residuals1
 - MLR-Best Subset Selection1

Analysis: Multiple Linear Regression

Input Parameters

Data File: Werner.cyd
 Model: CHOLESTEROL~%const+AGE+HEIGHT+WEI
 Confidence Level: 0.95

Output

Total no. of records: 188
 No. of records rejected: 7
 Terms dropped due to:

| | |
|-------------------|------|
| No Variation | None |
| Multicollinearity | None |

- The output is displayed in the main window. The first part of the output is as shown below:

Total no. of records: 188
No. of records rejected: 7
Terms dropped due to:

| | |
|-------------------|------|
| No Variation | None |
| Multicollinearity | None |

Summary Statistics:

| Item | Value |
|-------------------|------------|
| Residual df | 172 |
| Multiple R-square | 0.256 |
| Std.Dev.Estimate | 39.49 |
| Residual SS | 268221.407 |

The dataset contains a total of 188 records and out of this 7 are rejected due to missing

observations. The table titled “Terms dropped due to” refers to some essential pre-processing of the data. If a particular independent variable assumes the same value throughout the data set, it is not really a ‘variable’ and has to be dropped. Its presence creates ‘singularity’ in the design matrix X . In the present data set there is no such problem and hence the entry is ‘None’. Multicollinearity is another possible characteristic of the data, which could make the problem unstable. In the present data set, no such difficulty is encountered and hence the entry ‘None’ appears.

The table “Summary Statistics” displays some relevant summary statistics on residuals. In this example, $N = 181$ and $K = 9$. Thus the residual degrees of freedom is $181 - 9 = 172$. The multiple R^2 value is 0.256. This is obtained as:

$$R^2 = \frac{SSR}{SST}$$

The estimate of σ or error variance is 39.49. The *residual sum of squares* or SSE is 268221.407.

| Parameter Estimates: | | | | | | | |
|----------------------|-------------|------------|-------------------------|-------------|-------------|---------|-------------|
| Model term | Coefficient | Std. Error | 95% Confidence Interval | | t-statistic | p-value | SS |
| | | | Lower Limit | Upper Limit | | | |
| %CONST | 70.534 | 96.183 | -119.317 | 260.385 | 0.733 | 0.464 | 9951593.818 |
| AGE | 1.271 | 0.418 | 0.445 | 2.096 | 3.037 | 0.003 | 51320.975 |
| HEIGHT | -2.425 | 1.388 | -5.164 | 0.313 | -1.748 | 0.082 | 1123.665 |
| WEIGHT | 0.061 | 0.182 | -0.298 | 0.42 | 0.335 | 0.738 | 1196.139 |
| BIRTHPILL_0 | -26.842 | 21.168 | -68.625 | 14.941 | -1.268 | 0.207 | 851.006 |
| ALBUMIN | 0.166 | 0.998 | -1.804 | 2.136 | 0.167 | 0.868 | 7263.968 |
| CALCIUM | 2.342 | 0.723 | 0.914 | 3.77 | 3.237 | 0.001 | 19442.367 |
| URIC_ACID | 0.7 | 0.281 | 0.145 | 1.255 | 2.488 | 0.014 | 9811.723 |
| AGE*BIRTHPILL_0 | 0.591 | 0.601 | -0.595 | 1.778 | 0.984 | 0.326 | 1509.933 |

The table with title **Parameter Estimates** provides the estimate of regression coefficients with its standard error. It also provides 95% confidence interval of these estimates, the observed value of t-statistic, the p-value for testing $H_0 : \beta_k = 0$ and sum of squares. It appears that the terms age, calcium and uric acid is significant at 5% level of significance and the term height is significant at 10% level of significance. Notice that the variable BIRTHPILL considered as factor variable now has a suffix “_0”. **East** creates a dummy variable for the level 0 of factor BIRTHPILL. This dummy variable takes 1 for the observations with BIRTHPILL=0; otherwise it takes value 0. Here, the

83 Analysis-Regression Procedures

level 1 for the factor BIRTHPILL is considered as the reference level.

Estimated MSE of Prediction (GMSEP):

| Item | Value |
|-----------------------------|----------|
| Mean Squared Error | 1559.427 |
| Root Mean Squared Error | 39.49 |
| Estimated MSE of Prediction | 1641.401 |

The MSE ($\hat{\sigma}^2$) and RMSE ($\hat{\sigma}$) are 1559.427 and 39.49. The MSE of prediction is

1641.401. The following table displays the estimated covariance matrix of $\hat{\beta}$:

Variance-Covariance Matrix:

| | %CONST | AGE | HEIGHT | WEIGHT | BIRTHPILL_0 | ALBUMIN | CALCIUM | URIC_ACID | AGE*BIRTHPILL_0 |
|-----------------|----------|--------|---------|--------|-------------|---------|---------|-----------|-----------------|
| %CONST | 9251.198 | -5.392 | -97.616 | 3.465 | -229.172 | -3.624 | -30.564 | 0.138 | 4.905 |
| AGE | -5.392 | 0.175 | 0.044 | -0.013 | 5.724 | -0.012 | -0.006 | -0.012 | -0.168 |
| HEIGHT | -97.616 | 0.044 | 1.925 | -0.119 | 3.479 | -0.155 | -0.077 | 0.021 | -0.058 |
| WEIGHT | 3.465 | -0.013 | -0.119 | 0.033 | -0.287 | 0.054 | -0.013 | -0.012 | 0.006 |
| BIRTHPILL_0 | -229.172 | 5.724 | 3.479 | -0.287 | 448.1 | -2.361 | -0.811 | 0.126 | -12.173 |
| ALBUMIN | -3.624 | -0.012 | -0.155 | 0.054 | -2.361 | 0.996 | -0.329 | -0.014 | 0.028 |
| CALCIUM | -30.564 | -0.006 | -0.077 | -0.013 | -0.811 | -0.329 | 0.523 | -0.026 | 0.028 |
| URIC_ACID | 0.138 | -0.012 | 0.021 | -0.012 | 0.126 | -0.014 | -0.026 | 0.079 | -0.003 |
| AGE*BIRTHPILL_0 | 4.905 | -0.168 | -0.058 | 0.006 | -12.173 | 0.028 | 0.028 | -0.003 | 0.361 |

The table below displays the collinearity diagnostics:

Collinearity Diagnostics:

| Components | 1 | 2 |
|-------------------|--------|--------|
| Eigenvalues | 0.001 | 0.001 |
| Condition numbers | 119.39 | 77.177 |
| Intercept | 0.928 | 0.022 |
| AGE | 0.013 | 0.001 |
| HEIGHT | 0.768 | 0.183 |
| WEIGHT | 0.128 | 0.014 |
| BIRTHPILL_0 | 0.014 | 0.005 |
| ALBUMIN | 0.009 | 0.136 |
| CALCIUM | 0.115 | 0.869 |
| URIC_ACID | 0 | 0.016 |
| AGE*BIRTHPILL_0 | 0.006 | 0.005 |

When there is no collinearity at all, the Eigenvalues and condition number will all equal 1. As collinearity increases, Eigenvalues will be both greater and smaller than 1 (Eigenvalues close to zero indicates a multicollinearity problem), and the condition number will increase. Belsey, Kuh, and Welsch (1980) suggest that, when this number is around 10, weak dependencies might be starting to affect the regression estimates.

83 Analysis-Regression Procedures

Montgomery et al recommend use of 100 as indicative of moderate concern while a value of 1000 is an alarm trigger (Montgomery, Peck, and Vining, 2003, page 339). For this model, the condition number of scaled X matrix is 119.39. Thus, it may be pertinent to take corrective step such as centering the data.

| ANOVA: | | | | | |
|------------|-----|------------|-----------|-------------|----------|
| Source | df | SS | MS | F-statistic | p-value |
| Regression | 8 | 92519.776 | 11564.972 | 7.416 | 1.811E-8 |
| Error | 172 | 268221.407 | 1559.427 | | |
| Total | 180 | 360741.182 | | | |

The ANOVA table shows that the total degrees of freedom are 180, 8 independent variables give rise to 8 d.f. for regression and the remaining 172 degrees of freedom are assigned to error. The very low p-value shows that the model fitted is 'significant'. Therefore, we have to reject the null hypothesis that all regression coefficients are zero.

1. Click **MLR-Residuals1** in the **Library**. It displays the predicted values,

residuals and influential observations.

| Predicted Values | Residuals | Standardised Residuals | Studentised Residuals | Deleted Residuals | Cook's Distance |
|------------------|-----------|------------------------|-----------------------|-------------------|-----------------|
| 202.995 | -44.995 | -1.139 | -1.161 | -1.163 | 0.006 |
| 251.545 | -26.545 | -0.672 | -0.716 | -0.715 | 0.008 |
| 189.831 | 20.169 | 0.511 | 0.522 | 0.521 | 0.001 |
| 204.631 | -12.631 | -0.32 | -0.328 | -0.327 | 0.001 |
| 223.673 | 22.327 | 0.565 | 0.579 | 0.578 | 0.002 |
| 210.068 | -10.068 | -0.255 | -0.26 | -0.26 | 0 |
| 209.68 | -1.68 | -0.043 | -0.044 | -0.044 | 1.503E-5 |
| 212.761 | 67.239 | 1.703 | 1.736 | 1.747 | 0.013 |
| 198.321 | 16.679 | 0.422 | 0.437 | 0.436 | 0.001 |
| 217.528 | -10.528 | -0.267 | -0.272 | -0.271 | 0 |
| 213.776 | 16.224 | 0.411 | 0.418 | 0.417 | 0.001 |
| 232.979 | -7.979 | -0.202 | -0.206 | -0.206 | 0 |
| 231.923 | 28.077 | 0.711 | 0.733 | 0.732 | 0.004 |
| 222.406 | 2.594 | 0.066 | 0.068 | 0.068 | 3.643E-5 |
| 235.252 | 9.748 | 0.247 | 0.258 | 0.258 | 0.001 |
| 211.798 | -7.798 | -0.197 | -0.205 | -0.205 | 0 |
| 206.786 | -14.786 | -0.374 | -0.382 | -0.381 | 0.001 |
| 198.929 | -33.929 | -0.859 | -0.894 | -0.893 | 0.007 |
| 226.166 | -6.166 | -0.156 | -0.164 | -0.164 | 0 |

- Click **MLR-Best Subset Selection1** in the **Library**. This displays the output for best subset selection.

Analysis: Multiple Linear Regression

Multiple Linear Regression – Best Subset Selection:

| No. of Variables | RSS | CP | R-Squared | Adj. R-Squared | Probability | Var1 | Var2 |
|------------------|------------|-------|-----------|----------------|-------------|------|---------|
| 3 | 308891.673 | 23.08 | 0.144 | 0.134 | 0 | AGE | * |
| 4 | 284696.132 | 9.565 | 0.211 | 0.197 | 0.066 | AGE | CALCIUM |

For this example, the best subset of model with two terms includes BIRTHPILL and AGE as predictor. The best subset model of size 3 includes the predictors BIRTHPILL, AGE and CALCIUM.

83.2 Repeated Regression

Example: Repeated Regression

In a repeated measures setting, the subjects are measured repetitively over time. Therefore, the measurements observed within a same subject are correlated. In repeated regression analysis we take account of this correlation.

East performs repeated regression analysis using the MIXED procedure of SAS. East

83 Analysis-Regression Procedures

first generates equivalent SAS code and then displays the one obtained from the MIXED procedure in SAS.

Example: Repeated Regression

Dataset: Body_Weight.cyd as described in Section 82.3.

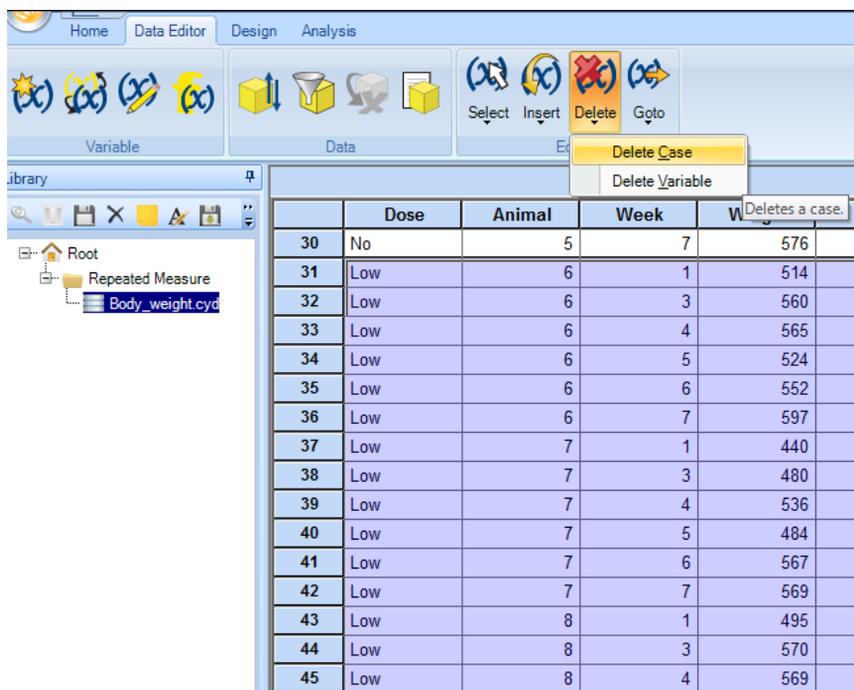
Purpose of the Analysis:

The data was obtained to investigate the effect of vitamin E diet supplement on the growth of guinea pigs. For each animal the body weight (in gram) was recorded at the end of weeks 1, 3, 4, 5, 6 and 7. All animals were given a growth-inhibiting substance during week 1 and the vitamin E therapy was started at the beginning of week 5. Three groups of animals, numbering five in each, received respectively zero, low and high doses of vitamin E. For this example, we will consider only observation from zero and high dose-groups. Here we want to fit the following model:

$$Weight_{ij} = \beta_0 + \beta_1 I(Dose_i = High) + \beta_2 j Week_{ij} + \epsilon_{ij}$$

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Delete the observations pertaining to “Low” dose (row 31 to 60). To delete the observations, select these observations, click  under the **Data Editor** menu and click **Delete Case**.



Now there are 60 observations left from “No” and “High” dose groups.

3. Choose the menu item:
Analysis > (Continuous) Regression > (Parallel Design) Repeated Measures Regression
 This will display several input fields associated with repeated regression analysis

83 Analysis-Regression Procedures

in the main window.

Analysis: Continuous Endpoint Regression Model - Parallel Design - Repeated Measures Regression

Data Set: Body_weight.cyd

Main Advanced SAS Command

Response: = Treatment: + Time(Repeated):

Subject:

Method of Estimation: REML Covariate Structure: UN DF: Contain

Covariates: Dose Animal Week Weight

Frequency Variable:

- There are three tabs in this window – **Main**, **Advanced** and **SAS Command**. In the **Main** tab, select **Weight** as **Response**, **Dose** as **Treatment**, **Animal** as **Subject** and **Week** as **Time(Repeated)**. All the remaining variables are displayed in **Covariates** field and you can select all or some of them as covariates. In the last row, there are 3 fields. First one is the **Method of Estimation** with choices of restricted maximum likelihood estimation (**REML**) and maximum likelihood estimation (**MLE**). Second one is the field **Covariance Structure** with choices of first order auto-regressive correlation (**AR(1)**), compound symmetry (**CS**), unstructured (**UN**), unstructured using correlations (**UNR**) and variance components (**VC**). Since all the animals were measured at fixed and equal times points, we can choose any reasonable covariance structure from **AR(1)**, **CS**, **UN** and **UNR**. The last one is the **DF** where you have to specify the method for computing the denominator degrees of freedom for the tests of significance of coefficients. Keep the default selections of **REML**, **UN** and **Contain** for **Method of Estimation**, **Covariance Structure** and **DF**.

Analysis: Continuous Endpoint Regression Model - Parallel Design - Repeated Measures Regression

Data Set: Body_weight.cyd

Main Advanced SAS Command

Response: Weight = Treatment: Dose + Time(Repeated): Week

Subject: Animal

Method of Estimation: REML Covariate Structure: UN DF: Contain

Covariates:

Frequency Variable:

- In the **Advanced** tab, leave the fields **By Variable 1** and **By Variable 2** blank. Enter 0.95 for **Confidence Level**.

| Analysis: Continuous Endpoint Regression Model - Parallel Design - Repeated Measures Regression | | | |
|---|----------------------|-------------------|-----------------------------------|
| Data Set: Body_weight.cyd | | | |
| Main Advanced SAS Command | | | |
| By Variable 1: | <input type="text"/> | Confidence Level: | <input type="text" value="0.95"/> |
| By Variable 2: | <input type="text"/> | Sing Res.: | <input type="text"/> |

- The third tab is **SAS Command** where you can put SAS code for more sophisticated analysis. For this example, do not make any changes in this tab.
- Click **OK** to start analysis. The output will be displayed in the main window now. ANOVA for all the three dose groups is displayed in the **Output** section.
- The output for estimated covariance structure is displayed in following

83 Analysis-Regression Procedures

screenshot.

| Covariance Parameter Estimates | | | | | |
|--------------------------------|---------|----------|----------------|---------|--------|
| Cov Parm | Subject | Estimate | Standard Error | Z Value | Pr Z |
| UN(1,1) | Animal | 588.80 | 299.63 | 1.97 | 0.0247 |
| UN(2,1) | Animal | 563.65 | 396.86 | 1.42 | 0.1555 |
| UN(2,2) | Animal | 1461.05 | 721.09 | 2.03 | 0.0214 |
| UN(3,1) | Animal | 407.72 | 385.36 | 1.06 | 0.2901 |
| UN(3,2) | Animal | 1406.26 | 718.83 | 1.96 | 0.0504 |
| UN(3,3) | Animal | 1513.03 | 752.46 | 2.01 | 0.0222 |
| UN(4,1) | Animal | 505.86 | 432.38 | 1.17 | 0.2420 |
| UN(4,2) | Animal | 1574.66 | 801.32 | 1.97 | 0.0494 |
| UN(4,3) | Animal | 1588.52 | 815.07 | 1.95 | 0.0513 |
| UN(4,4) | Animal | 1934.95 | 944.40 | 2.05 | 0.0202 |
| UN(5,1) | Animal | 673.91 | 494.38 | 1.36 | 0.1728 |
| UN(5,2) | Animal | 1484.58 | 852.63 | 1.74 | 0.0817 |
| UN(5,3) | Animal | 1435.99 | 855.19 | 1.68 | 0.0931 |
| UN(5,4) | Animal | 1762.99 | 991.23 | 1.78 | 0.0753 |
| UN(5,5) | Animal | 2863.26 | 1350.69 | 2.12 | 0.0170 |
| UN(6,1) | Animal | 393.59 | 410.05 | 0.96 | 0.3371 |
| UN(6,2) | Animal | 1338.72 | 759.75 | 1.76 | 0.0781 |
| UN(6,3) | Animal | 1320.28 | 766.60 | 1.72 | 0.0850 |
| UN(6,4) | Animal | 1605.75 | 887.28 | 1.81 | 0.0703 |
| UN(6,5) | Animal | 2442.42 | 1176.46 | 2.08 | 0.0379 |
| UN(6,6) | Animal | 2265.32 | 1068.05 | 2.12 | 0.0170 |

9. The estimated coefficients are given in the following screenshot:

| Solution for Fixed Effects | | | | | | | | | | |
|----------------------------|------|------|----------|----------------|----|---------|---------|-------|----------|----------|
| Effect | Dose | Week | Estimate | Standard Error | DF | t Value | Pr > t | Alpha | Lower | Upper |
| Intercept | | | 572.94 | 15.3413 | 53 | 37.35 | <.0001 | 0.05 | 542.17 | 603.71 |
| Dose | High | | 49.3225 | 5.9412 | 53 | 8.30 | <.0001 | 0.05 | 37.4060 | 61.2390 |
| Dose | No | | 0 | . | . | . | . | . | . | . |
| Week | | 1 | -115.50 | 14.3769 | 53 | -8.03 | <.0001 | 0.05 | -144.34 | -86.6637 |
| Week | | 3 | -70.6000 | 10.2417 | 53 | -6.89 | <.0001 | 0.05 | -91.1423 | -50.0577 |
| Week | | 4 | -23.3000 | 10.6667 | 53 | -2.18 | 0.0334 | 0.05 | -44.6947 | -1.9053 |
| Week | | 5 | -30.9000 | 9.9436 | 53 | -3.11 | 0.0030 | 0.05 | -50.8444 | -10.9556 |
| Week | | 6 | -30.2000 | 4.9369 | 53 | -6.12 | <.0001 | 0.05 | -40.1022 | -20.2978 |
| Week | | 7 | 0 | . | . | . | . | . | . | . |

Therefore, the fitted model in this case is:

$$Weight_{ij} = 572.94 + 49.32I(Dose_i = Height) - 115.5I(Week = 1)$$

$$-70.6I(Week = 3) - 23.3I(Week = 4) - 30.9I(Week = 5) - 30.2I(Week = 6)$$

for $i = 1, \dots, 60$ with covariance structure as

$$\begin{bmatrix} 588.80 & 563.65 & 407.72 & 505.86 \\ 563.65 & 1461.05 & 1406.26 & 1574.66 \\ 407.72 & 1406.26 & 1513.03 & 1588.52 \\ 505.86 & 1574.66 & 1588.52 & 1934.95 \end{bmatrix}$$

83.3 Linear Mixed Effects Model: Difference of Means (Crossover Data)

In linear mixed effects model a linear model is fitted to explain variability in the response variable with the help of factors with levels, which have fixed effects and more than one random effect. Mixed Effects model is used for hierarchical or dependent data.

You will need to specify **Response** variable for which you want to fit the model. In this particular design, **Response** variable is often difference of means in test and control group. You will need to specify factors (variables) with fixed effects namely **Period**

83 Analysis-Regression Procedures

ID, **Group ID** and **Treatment ID**. You will also need to specify **Subject ID** which identifies the source of the response variable.

You have an option of checking the box **Run using SAS** on **Advanced** tab. By doing this **East** will invoke **Mixed** procedure of SAS. You can also choose not to use SAS. If you use SAS, you will have the option of including covariates in our model. Without SAS, your model will not include covariates. You can also invoke SAS command option from the dialogue box for this test.

East will display among other things estimates, t-statistics and ANOVA table for fixed effects. In this section, we will illustrate repeated regression analysis of 2x2 crossover data using all the three options.

Dataset: CrossoverCaseData.cyd

Analysis Using East:

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

Analysis > (Continuous) Regression > (Crossover Design) Linear Mixed Effects Model: Difference of Means

This will display several input fields associated with linear mixed effects model analysis in the main window.

Analysis: Continuous Endpoint Regression Model - Crossover Design - Linear Mixed Effects Model: Difference of Means

Data Set: CrossoverCaseData.cyd

Main Advanced SAS Command

Response: = Period ID: + Group ID: + Treatment ID:

Subject ID (Random Effect):

3. There are three tabs in this window – **Main**, **Advanced** and **SAS Command**. In the **Main** tab, select **Response** as **Response**, **PeriodID** as **Period ID**, **GroupID** as **Group ID** and **SubjectID** as **Subject ID (Random Effect)**. Once you select a random effect, options of **Method of Estimation** with choices of restricted maximum likelihood estimation (**REML**) and maximum likelihood estimation (**MLE**) will become available. Keep the default selection of **REML**.

Analysis: Continuous Endpoint Regression Model - Crossover Design - Linear Mixed Effects Model: Difference of Means

Data Set: CrossoverCaseData.cyd

Main **Advanced** SAS Command

Response: = Period ID: + Group ID: + Treatment ID: Method of estimation: REML MLE

Subject ID (Random Effect):

- In the **Advanced** tab, leave the fields **By Variable 1** and **By Variable 2** blank. Keep the default value **0.95** for **Confidence Level** as well as all the statistics to be computed. Don't check the **Run Using SAS** checkbox.

Analysis: Continuous Endpoint Regression Model - Crossover Design - Linear Mixed Effects Model: Difference of Means

Data Set: CrossoverCaseData.cyd

Main **Advanced** SAS Command

By Variable 1: Confidence Level: Select All

By Variable 2: Run Using SAS Statistics: Fit Statistics
 Fixed Estimates
 Fixed Effects Tests
 Covariance Matrix

- The third tab is **SAS Command** where you can put SAS code for more sophisticated analysis. For this example, do not make any changes in this tab.

83 Analysis-Regression Procedures

6. Click **OK** to start analysis. The output will be displayed in the main window.

Analysis: Linear Mixed Effects Model: Difference of Means

Data File: CrossoverCaseData.cyd
 Model: Response = Intercept + PeriodID + GroupID + subjectID(GroupID)
 Estimation Method: REML
 Degrees of Freedom method: Residual
 Confidence Level: 0.95

Output

Response Variable: Response
 Total Number of Records: 112
 Number of Records Rejected: 0
 Fit Statistics:
 Residual Variance Estimate: 375.332
 Subject Variance Estimate: 5690.719
 -2 (Log Likelihood) : 1152.838
 Fixed Effects Estimates :

| Effect | Estimate | Std. Error | 95% Confidence Interval | |
|----------|----------|------------|-------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| %Const | 225.146 | 14.355 | 196.696 | 253.596 |
| PeriodID | -4.139 | 3.661 | -11.395 | 3.118 |
| GroupID | 19.444 | 20.504 | -21.194 | 60.083 |

Fixed Effects Tests:

t tests:

| Effect | DF | t | p-value |
|----------|-----|-------|---------|
| PeriodID | 109 | -1.13 | 0.261 |
| GroupID | 109 | 0.948 | 0.345 |

ANOVA:

| Effect | Num DF | Den DF | F | p-value |
|----------|--------|--------|-------|---------|
| PeriodID | 1 | 109 | 1.278 | 0.261 |
| GroupID | 1 | 109 | 0.899 | 0.345 |

Covariance Matrix Estimate:

| | %Const | PeriodID | GroupID |
|----------|----------|-----------|-----------|
| %Const | 206.054 | -6.702 | -202.703 |
| PeriodID | -6.702 | 13.405 | 2.472E-14 |
| GroupID | -202.703 | 2.472E-14 | 420.421 |

Analysis Using SAS:

Analysis Steps:

1. Choose the menu item:

Analysis > (Continuous) Regression > (Crossover Design) Linear Mixed Effects Model: Difference of Means

As explained earlier, in the **Main** tab, select **Response** as **Response**, **PeriodID** as **Period ID**, **GroupID** as **Group ID** and **SubjectID** as **Subject ID (Random Effect)**. Once you select **Random Effect**, options of **Method of Estimation** with choices of restricted maximum likelihood estimation (**REML**) and maximum likelihood estimation (**MLE**) will become available. Keep the default selection of **REML**.

- In the **Advanced** tab, leave the fields **By Variable 1** and **By Variable 2** blank. Keep the default value **0.95** for **Confidence Level** as well as all the statistics to be computed. Now check the **Run Using SAS** checkbox.

Note that you can use covariates while running SAS. Don't check any of the covariates for this example.

- Click **OK** to start analysis. East will invoke SAS and the SAS output will be

83 Analysis-Regression Procedures

displayed in the main window.

The SAS System
The Mixed Procedure

| Model Information | |
|---------------------------|------------------------|
| Data Set | WORK.CROSSOVERCASEDATA |
| Dependent Variable | Response |
| Covariance Structure | Variance Components |
| Estimation Method | REML |
| Residual Variance Method | Profile |
| Fixed Effects SE Method | Model-Based |
| Degrees of Freedom Method | Residual |

| Class Level Information | | |
|-------------------------|--------|---|
| Class | Levels | Values |
| subjectID | 56 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 |
| PeriodID | 2 | 1 2 |
| GroupID | 2 | 1 2 |

| Dimensions | |
|-----------------------|-----|
| Covariance Parameters | 2 |
| Columns in X | 5 |
| Columns in Z | 56 |
| Subjects | 1 |
| Max Obs Per Subject | 112 |

| Number of Observations | |
|---------------------------------|-----|
| Number of Observations Read | 112 |
| Number of Observations Used | 112 |
| Number of Observations Not Used | 0 |

| Iteration History | | | |
|-------------------|-------------|-----------------|------------|
| Iteration | Evaluations | -2 Res Log Like | Criterion |
| 0 | 1 | 1269.20852181 | |
| 1 | 1 | 1152.83778206 | 0.00000000 |

Convergence criteria met.

| Covariance Parameter Estimates | |
|--------------------------------|----------|
| Cov Parm | Estimate |
| subjectID | 5690.72 |
| Residual | 375.33 |

| Fit Statistics | |
|--------------------------|--------|
| -2 Res Log Likelihood | 1152.8 |
| AIC (smaller is better) | 1156.8 |
| AICC (smaller is better) | 1157.0 |
| BIC (smaller is better) | 1160.9 |

| Solution for Fixed Effects | | | | | | | | | | |
|----------------------------|----------|---------|----------|----------------|-----|---------|---------|-------|----------|---------|
| Effect | PeriodID | GroupID | Estimate | Standard Error | DF | t Value | Pr > t | Alpha | Lower | Upper |
| Intercept | | | 225.15 | 14.3546 | 109 | 15.68 | <.0001 | 0.05 | 196.70 | 253.60 |
| PeriodID | 1 | | -4.1387 | 3.6612 | 109 | -1.13 | 0.2608 | 0.05 | -11.3952 | 3.1178 |
| PeriodID | 2 | | 0 | . | . | . | . | . | . | . |
| GroupID | | 1 | 19.4442 | 20.5042 | 109 | 0.95 | 0.3451 | 0.05 | -21.1944 | 60.0828 |
| GroupID | | 2 | 0 | . | . | . | . | . | . | . |

| Covariance Matrix for Fixed Effects | | | | | | | | |
|-------------------------------------|------------|----------|---------|---------|---------|------|---------|------|
| Row | Effect | PeriodID | GroupID | Col1 | Col2 | Col3 | Col4 | Col5 |
| 1 | Intercept | | | 206.05 | -6.7024 | | -202.70 | |
| 2 | PeriodID 1 | | | -6.7024 | 13.4047 | | | |
| 3 | PeriodID 2 | | | | | | | |

83 Analysis-Regression Procedures

| Type 3 Tests of Fixed Effects | | | | |
|-------------------------------|--------|--------|---------|--------|
| Effect | Num DF | Den DF | F Value | Pr > F |
| PeriodID | 1 | 109 | 1.28 | 0.2608 |
| GroupID | 1 | 109 | 0.90 | 0.3451 |

| Estimates | | | | | |
|---------------------|----------|----------------|-----|---------|---------|
| Label | Estimate | Standard Error | DF | t Value | Pr > t |
| PeriodID Estimate : | 4.1387 | 3.6612 | 109 | 1.13 | 0.2608 |
| GroupID Estimate : | -19.4442 | 20.5042 | 109 | -0.95 | 0.3451 |

| Least Squares Means | | | | | | | | | | |
|---------------------|----------|---------|----------|----------------|-----|---------|---------|-------|--------|--------|
| Effect | PeriodID | GroupID | Estimate | Standard Error | DF | t Value | Pr > t | Alpha | Lower | Upper |
| PeriodID | 1 | | 230.73 | 10.4142 | 109 | 22.16 | <.0001 | 0.05 | 210.09 | 251.37 |
| PeriodID | 2 | | 234.87 | 10.4142 | 109 | 22.55 | <.0001 | 0.05 | 214.23 | 255.51 |
| GroupID | | 1 | 242.52 | 14.7553 | 109 | 16.44 | <.0001 | 0.05 | 213.28 | 271.77 |
| GroupID | | 2 | 223.08 | 14.2374 | 109 | 15.67 | <.0001 | 0.05 | 194.86 | 251.29 |

| Differences of Least Squares Means | | | | | | | | | | | |
|------------------------------------|----------|---------|-----------|----------|----------|----------------|-----|---------|---------|-------|----------|
| Effect | PeriodID | GroupID | _PeriodID | _GroupID | Estimate | Standard Error | DF | t Value | Pr > t | Alpha | Lower |
| PeriodID | 1 | | 2 | | -4.1387 | 3.6612 | 109 | -1.13 | 0.2608 | 0.05 | -11.3952 |
| GroupID | | 1 | | 2 | 19.4442 | 20.5042 | 109 | 0.95 | 0.3451 | 0.05 | -21.1944 |

Analysis Using SAS Command: Analysis Steps:

1. Choose the menu item: **Analysis > (Continuous) Regression > (Crossover Design) Linear Mixed Effects Model: Difference of Means**
As explained earlier, in the **Main** tab, select **Response** as **Response**, **PeriodID** as **Period ID**, **GroupID** as **Group ID** and **SubjectID** as **Subject ID (Random Effect)**. Once you select a random effect, options of **Method of Estimation** with choices of restricted maximum likelihood estimation (**REML**) and maximum likelihood estimation (**MLE**) will become available. Keep the default selection of **REML**.

Analysis: Continuous Endpoint Regression Model - Crossover Design - Linear Mixed Effects Model: Difference of Means

Data Set: CrossoverCaseData.cyd

Main **Advanced** SAS Command

Response: = Period ID: + Group ID: + Treatment ID: Method of estimation: REML MLE

Subject ID (Random Effect):

- In the **Advanced** tab, leave the fields **By Variable 1** and **By Variable 2** blank. Keep the default value **0.95** for **Confidence Level** as well as all the statistics to be computed. Don't check the **Run Using SAS** checkbox.

Analysis: Continuous Endpoint Regression Model - Crossover Design - Linear Mixed Effects Model: Difference of Means

Data Set: CrossoverCaseData.cyd

Main **Advanced** SAS Command

By Variable 1: Confidence Level: Select All

By Variable 2: Run Using SAS Statistics: Fit Statistics
 Fixed Estimates
 Fixed Effects Tests
 Covariance Matrix

- Go to the **SAS Command** tab. You will see a SAS code already written in the main window. A partial view of the same is shown below:

83 Analysis-Regression Procedures

```
input
GroupID
PeriodID
subjectID
Response
LnResp
GrpStr$
bin
;
run;
options nodate pageno=1 linesize=70 pagesize=60 nobyline;
ods listing select all;
ods listing File = "C:\Users\vidya.prayag\AppData\Local\Temp\GeneralisedMixed.txt";
ods html body='C:\Users\vidya.prayag\AppData\Local\Temp\GeneralisedMixed.html';
DATA CrossoverCaseData ;
set CrossoverCaseData ;
<Response> = log(<Response>);
run;
proc sort data = CrossoverCaseData out = SASortMixed;
by <By Variable 1> <By Variable 2> ... <By Variable N> ;
run;
/* Write your code here */
ods html close;
```

The first few commands are meant for reading the data in SAS. You will also see the statement

```
/* Write your code here */
```

We will replace this part by our code.

```
DATA CrossoverCaseData ;
set CrossoverCaseData ;
proc mixed method = REML;
class GroupID PeriodID ;
model Response = GroupID PeriodID ;
repeated PeriodID ;
random subjectID;
run;
```

Please remove the following lines from the existing code.

```
<Response> = log(<Response>);
run;
proc sort data = CrossoverCaseData out =
SASortMixed;
by <By Variable 1> <By Variable 2> ... <By Variable
N> ;
run;
```

This is required as we don't want to log transform the **Response** and also don't want to sort the data on any by variable.

- Click **OK**. East will invoke SAS and the SAS output will be displayed in the main window.

The SAS System
The Mixed Procedure

| Model Information | |
|----------------------------------|------------------------|
| Data Set | WORK.CROSSOVERCASEDATA |
| Dependent Variable | Response |
| Covariance Structure | Variance Components |
| Estimation Method | REML |
| Residual Variance Method | Parameter |
| Fixed Effects SE Method | Model-Based |
| Degrees of Freedom Method | Containment |

| Class Level Information | | |
|-------------------------|--------|--------|
| Class | Levels | Values |
| GroupID | 2 | 12 |
| PeriodID | 2 | 12 |

| Dimensions | |
|------------------------------|-----|
| Covariance Parameters | 2 |
| Columns in X | 5 |
| Columns in Z | 1 |
| Subjects | 1 |
| Max Obs Per Subject | 112 |

83 Analysis-Regression Procedures

| Number of Observations | |
|---------------------------------|-----|
| Number of Observations Read | 112 |
| Number of Observations Used | 112 |
| Number of Observations Not Used | 0 |

| Iteration History | | | |
|-------------------|-------------|-----------------|------------|
| Iteration | Evaluations | -2 Res Log Like | Criterion |
| 0 | 1 | 1269.20852181 | |
| 1 | 1 | 1267.98423989 | 0.00000000 |

Convergence criteria met.

| Covariance Parameter Estimates | |
|--------------------------------|----------|
| Cov Parm | Estimate |
| subjectID | 2.0095 |
| PeriodID | 5878.60 |

| Fit Statistics | |
|--------------------------|--------|
| -2 Res Log Likelihood | 1268.0 |
| AIC (smaller is better) | 1272.0 |
| AICC (smaller is better) | 1272.1 |
| BIC (smaller is better) | 1268.0 |

| Type 3 Tests of Fixed Effects | | | | |
|-------------------------------|--------|--------|---------|--------|
| Effect | Num DF | Den DF | F Value | Pr > F |
| GroupID | 1 | 108 | 4.26 | 0.0414 |
| PeriodID | 1 | 108 | 0.08 | 0.7757 |

83.4 Linear Mixed Effects Model: Ratio of Means (Crossover Data)

This test is very similar to the test described in previous subsection except here the **Response** variable is often the ratio of means of treatment and control group. The previous test is applied to logarithm of the response variable.

Both the options of SAS link (Run using SAS) and SAS commands are available for this test.

84 *Analysis-Multiple Comparison Procedures for Continuous Data*

It is often the case that multiple objectives are to be addressed in one single trial. These objectives are formulated into a family of hypotheses. Type I error rate is inflated when one considers the inferences together as a family. Failure to compensate for multiplicities can have adverse consequences. For example, a drug could be approved when actually it is not better than Placebo. Multiple comparison (MC) procedures provide a guard against inflation of type I error due to multiple testing. Probability of making at least one type I error is known as family wise error rate (FWER). **East** supports several parametric and p-value based MC procedures.

We have seen how to simulate data under different MC procedures with specified group means and variance in chapter 15. In this chapter we explain how to analyze data with different MC procedures available in **East**. For MC procedures in **East**, we can either provide the dataset containing the observations under each arm or the raw p-values to obtain the adjusted p-values.

84.1 Available Procedures

The probability of making at least one type I error is known as family wise error rate (FWER). All the MC procedures available in **East** strongly control FWER. Strong control of FWER refers to preserving the probability of incorrectly claiming at least one null hypothesis. To contrast strong control with weak control of FWER, the latter controls the FWER under the assumption that all hypotheses are true. **East** supports following MC procedures based on continuous endpoint.

| Category | Procedure | Reference |
|---------------|-----------------------|------------------------------------|
| Parametric | Dunnett's Single Step | Dunnett CW (1955) |
| | Dunnett's Step Down | Dunnett CW and Tamhane AC (1991) |
| | Dunnett's Step Up | Dunnett CW and Tamhane AC (1992) |
| P-value Based | Bonferroni | Bonferroni CE (1935, 1936) |
| | Sidak | Sidak Z (1967) |
| | Weighted Bonferroni | Benjamini Y and Hochberg Y (1997) |
| | Holm's Step Down | Holm S (1979) |
| | Hochberg's Step Up | Hochberg Y (1988) |
| | Hommel's Step Up | Hommel G (1988) |
| | Fixed Sequence | Westfall PH, Krishen A (2001) |
| | Fallback | Wiens B, Dmitrienko A (2005) |

Dose-Finding Hypertension Trial

Throughout this chapter we consider the data from a dose-finding hypertension trial (Dmitreinko and Offen, 2005) to illustrate different MC procedures. The trial was conducted to compare four doses of a new antihypertensive drug to a Placebo. The primary outcome is reduction in diastolic blood pressure. Doses with significant mean reduction in mean diastolic blood pressure will be declared efficacious. The data from this trial are available in **East** through the dataset **Hypertension-trial.cyd**.

Let $\mu_0, \mu_1, \mu_2, \mu_3$ and μ_4 indicate the group means in Placebo, Dose1, Dose2, Dose3 and Dose4 treatment groups. We are interested in testing following right tailed tests:

$$H_i : \mu_i - \mu_0 \leq 0 \text{ vs } K_i : \mu_i - \mu_0 > 0 \quad i = 1, 2, 3, 4$$

and the global null hypothesis

$$H_0 : \mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4$$

We want to control the FWER at 5% level of significance.

84.2 Example: Dunnett's single step

Dataset: Hypertension-trial.cyd

Data Description:

The trial was conducted to compare four doses of a new antihypertensive drug to a Placebo. The primary outcome is reduction in diastolic blood pressure. Doses with significant mean reduction in mean diastolic blood pressure will be declared efficacious.

The dataset has 130 observations and 2 columns. The first column **Dose** contains the information on the dose level. There are 5 dose levels including Placebo. In this column, **P** represents Placebo where as “D1” through “D4” represent 4 dose levels of the drug. The second column, **Response**, contains the reduction in diastolic blood pressure (expressed in mmHg). Each line in the data set represents a subject in the study.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

**Analysis > (Continuous) Many Samples > (Multiple Comparisons)
Pairwise Comparisons to Controls - Difference of Means**

84 Analysis-Multiple Comparison Procedures for Continuous Data

- In the ensuing dialog box, under the **Main** tab choose the variables as shown below.

Analysis Inputs

Analysis: Continuous Endpoint Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Differences of Means

Data Set: Hypertension-trial.cyd

Main Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
 Select Treatment Variable:
 Dose

D1 P
 D2
 D3
 D4

Response Variable
 Select Response Variable: Response

Equal Variance Unequal Variance

Multiple Comparison Procedure
 Select MCP: Dunnett's single step

- Now click on **Advanced** tab. Leave the fields **By Variable 1** and **By Variable 2** blank. On the left, enter **0.95** for **Confidence Level** and select **Right-Tail** for **Rejection Region**.

Data Set: Hypertension-trial.cyd

Main Advanced

By Variable 1: [] Confidence Level: 0.95

By Variable 2: [] Rejection Region: Right-Tail

- Click **OK** to start the analysis. Once the analysis is over, the output will be

displayed in the main window now.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis
 $H_1 : \mu_i - \mu_0 \leq 0$ Vs. $K_1 : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: Equal
 Multiple Comparison Procedure: Dunnett's single step
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 1.963 | -0.354 | 0.638 | 0.896 | -3.95 | INF |
| D2 | 26 | 5.254 | 1.924 | 2.364 | 0.01 | 0.033 | 1.36 | INF |
| D3 | 24 | 5.629 | 1.963 | 2.508 | 0.007 | 0.023 | 1.671 | INF |
| D4 | 26 | 7.331 | 1.924 | 3.443 | 0 | 0.001 | 3.437 | INF |

Adjusted Global p-value: 0.001
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The last section is the **Output**. For each treatment group (referred as **Arm**), sample size, mean and standard error of difference of mean are given in a table. The **Ctrl** arm in this table indicates “Placebo”. Mean responses for Placebo, Dose1, Dose2, Dose3 and Dose4 are 0.704, 0.008, 5.254, 5.629 and 7.331 mmHg, respectively.

The table in the **Output** section also includes the observed value of test statistic and p-values for comparison with control group along with 95% one-sided confidence interval for the difference with Placebo. There are two types of p-values in this table. The **Naive** p-values are referred to raw or un-adjusted p-values. The p-values in the **Adjusted** column are obtained after multiplicity adjustment according to Dunnett’s single step procedure so that FWER is maintained at 5% level of significance. The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 0.896, 0.033, 0.023 and 0.001, respectively. Therefore, after multiplicity adjustment according to Dunnett’s single step procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

84 Analysis-Multiple Comparison Procedures for Continuous Data

Under this table, adjusted global p-value is given which is 0.001 in this case. This is the p-value to reject the following global null hypothesis:

$$H_0 : \mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4$$

One can verify that global p-value is the minimum of all the 4 adjusted p-values given in the table above.

84.3 Example: Dunnett's step-down and step-up procedures

Dataset: *Hypertension-trial.cyd* as described in Section 84.2

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below.

The screenshot shows the 'Analysis Inputs' dialog box for a 'Continuous Endpoint Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Differences of Means'. The 'Data Set' is 'Hypertension-trial.cyd'. The 'Main' tab is selected. Under 'Treatment Variable and its Control Arm', 'Dose' is selected as the 'Treatment Variable' and 'P' is selected as the 'Control Arm'. Under 'Response Variable', 'Response' is selected. The 'Multiple Comparison Procedure' is set to 'Dunnett's step-down'. The 'Equal Variance' option is selected.

4. Now click on **Advanced** tab. Leave the fields **By Variable 1** and **By Variable 2** blank. On the left, enter 0.95 for **Confidence Level** and select **Right-Tail** for **Rejection Region**.

- Click **OK** to analyze the data. The output will be displayed in the main window once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis
 $H_1: \mu_i - \mu_0 \leq 0$ Vs. $K_1: \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: Equal
 Multiple Comparison Procedure: Dunnett's step-down
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 1.963 | -0.354 | 0.638 | 0.638 | -3.95 | INF |
| D2 | 26 | 5.254 | 1.924 | 2.364 | 0.01 | 0.018 | 1.36 | INF |
| D3 | 24 | 5.629 | 1.963 | 2.508 | 0.007 | 0.018 | 1.671 | INF |
| D4 | 26 | 7.331 | 1.924 | 3.443 | 0 | 0.001 | 3.437 | INF |

Adjusted Global p-value: 0.001
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 0.638, 0.018, 0.018 and 0.001, respectively. Therefore, after multiplicity adjustment according to Dunnett’s step-down procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

We can perform **Dunnett’s step-up** test by selecting **Dunnett’s step-up** from the drop-down menu in **Select MCP** in the **Main** tab of input window. However, Dunnett’s step-up test cannot be performed with **Hypertension-trial.cyd** dataset as the number of observations for all the 5 treatment are not equal. In other words, the treatment groups are not balanced in this data. Number of observations in Placebo, Dose1, Dose2, Dose3 and Dose4 groups are 25, 24, 26, 24 and 26 respectively.

Comparison of Dunnett’s single step and step-down procedures results

The table below compares the p-values for comparison with Placebo for the two

84 Analysis-Multiple Comparison Procedures for Continuous Data

different methods (Dunnett's single step and step-down) along with the raw p-values.

| Arm | Raw | Single step | Step-down |
|-----|-------|-------------|-----------|
| D1 | 0.638 | 0.896 | 0.638 |
| D2 | 0.010 | 0.033 | 0.018 |
| D3 | 0.007 | 0.023 | 0.018 |
| D4 | 0.000 | 0.001 | 0.001 |

Notice that the p-values for the step-down procedure are all smaller than the p-values for the single-step procedure except for the Dose4.

84.4 p-value based Procedures

The p-value based procedures strongly control the FWER regardless of the joint distribution of the raw p-values as long as the individual raw p-values are legitimate p-values. Assume that there are k arms including the Placebo arm. Let n_i be the number of subjects for i -th treatment arm ($i = 0, 2, \dots, k - 1$). Let $N = \sum_{i=0}^{k-1} n_i$ be the total sample size and the arm 0 refers to Placebo. Let Y_{ij} be the response from subject j in treatment arm i and y_{ij} be the observed value of Y_{ij} ($i = 0, 2, \dots, k - 1, j = 1, 2, \dots, n_i$).

Suppose that

$$Y_{ij} = \mu_i + e_{ij} \tag{84.1}$$

where $e_{ij} \sim N(0, \sigma_i^2)$.

We are interested in the following hypotheses:

- For the right tailed test: $H_i : \mu_i - \mu_0 \leq 0$ vs $K_i : \mu_i - \mu_0 > 0$
- For the left tailed test: $H_i : \mu_i - \mu_0 \geq 0$ vs $K_i : \mu_i - \mu_0 < 0$

For the global null hypothesis at least one of the H_i is rejected in favor of K_i after controlling for FWER. Here H_i and K_i refer to null and alternative hypotheses, respectively, for comparison of i -th arm with the Placebo arm.

Let \bar{y}_i be the sample mean for treatment arm i , s_i^2 be the sample variance from i -th arm and s^2 be the pooled sample variance for all arms. For the equal variance case, one

need to replace s_i^2 and s_0^2 by the pooled sample variance s^2 . For both the case, T_i is distributed as Student's t distribution. However, the degrees of freedom varies for equal variance and unequal variance case. For equal variance case the degrees of freedom would be $N - k$. For the unequal variance case, the degrees of freedom is subject to Satterthwaite correction.

Let t_i be the observed value of T_i and these observed values for $K - 1$ treatment arms can be ordered as $t_{(1)} \geq t_{(2)} \geq \dots \geq t_{(k-1)}$. For the right tailed test the marginal p-value for comparing the i -th arm with Placebo is calculated as $p_i = P(T > t_i)$ and for left tailed test $p_i = P(T < t_i)$, where T is distributed as Student's t distribution. Let $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k-1)}$ be the ordered p-values.

For the unequal variance case, the test statistic for comparing treatment effect of arm i with Placebo can be defined as

$$T_i = \frac{\bar{y}_i - \bar{y}_0}{\sqrt{\frac{1}{n_i} s_i^2 + \frac{1}{n_0} s_0^2}} \tag{84.2}$$

84.5 Single step MC procedures

East supports three p-value based single step MC procedures:

- Bonferroni procedure
- Sidak procedure and
- Weighted Bonferroni procedure.

For the Bonferroni procedure, H_i is rejected if $p_i < \frac{\alpha}{k-1}$ and the adjusted p-value is given as $\min(1, (k - 1)p_i)$.

For the Sidak procedure, H_i is rejected if $p_i < 1 - (1 - \alpha)^{\frac{1}{k-1}}$ and the adjusted p-value is given as $1 - (1 - p_i)^{k-1}$.

For the weighted Bonferroni procedure, H_i is rejected if $p_i < w_i \alpha$ and the adjusted p-value is given as $\min(1, \frac{p_i}{w_i})$. Here w_i denotes the proportion of α allocated to the H_i such that $\sum_{i=1}^{k-1} w_i = 1$. Note that, if $w_i = \frac{1}{k-1}$, then the Bonferroni procedure is reduced to the regular Bonferroni procedure.

Example: Bonferroni procedure

Dataset: Hypertension-trial.cyd as described in Section 84.2

84 Analysis-Multiple Comparison Procedures for Continuous Data

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below.

Analysis Inputs

Analysis: Continuous Endpoint Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Differences of Mean

Data Set: Hypertension-trial.cyd

Main Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
Select Treatment Variable:
Dose

D1 P
 D2
 D3
 D4

Response Variable
Select Response Variable: Response

Equal Variance Unequal Variance

Multiple Comparison Procedure
Select MCP: Bonferroni

4. Click **OK** to analyze the data. The output will be displayed in the main window.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis
 $H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: UnEqual
 Multiple Comparison Procedure: Bonferroni
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 1 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.031 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.045 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 0 | 0.001 | 3.585 | INF |

Adjusted Global p-value: 0.001
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 1, 0.031, 0.045 and 0.001, respectively. Therefore, after multiplicity adjustment according to Bonferroni procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

Example: Sidak procedure

Dataset: Hypertension-trial.cyd as described in Section 84.2

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown

84 Analysis-Multiple Comparison Procedures for Continuous Data

below.

Analysis: Continuous Endpoint Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Differences of Means

Data Set: Hypertension-trial.cvd

Main Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
Select Treatment Variable:
Dose

D1 P
 D2
 D3
 D4

Response Variable
Select Response Variable: Response

Equal Variance Unequal Variance

Multiple Comparison Procedure
Select MCP: Sidak

4. Click **OK** to analyze the data. The output will be displayed in the main window once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis

$H_1: \mu_i - \mu_0 \leq 0$ Vs. $K_1: \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: UnEqual
 Multiple Comparison Procedure: Sidak
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One side) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|---|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 0.982 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.031 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.044 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 0 | 0.001 | 3.585 | INF |

Adjusted Global p-value: 0.001
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 0.982, 0.031, 0.044 and 0.001, respectively. Therefore, after multiplicity adjustment according to Sidak procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

Example: Weighted Bonferroni procedure

Dataset: Hypertension-trial.cyd as described in Section 84.2

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown

84 Analysis-Multiple Comparison Procedures for Continuous Data

below.

Analysis: Continuous Endpoint Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Differences... .00 .00

Data Set: Hypertension-trial.cyd

Main Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
 Select Treatment Variable: Dose

D1 P
 D2
 D3
 D4

Response Variable
 Select Response Variable: Response

Equal Variance Unequal Variance

Multiple Comparison Procedure
 Select MCP: Weighted Bonferroni

No. of Treatment Arms: 4

Table of Treatment specific parameters:

| Arm | Proportion of... |
|-----|------------------|
| D1 | 0.25 |
| D2 | 0.25 |
| D3 | 0.25 |
| D4 | 0.25 |

4. Upon selection of weighted Bonferroni procedure, a table will appear under the drop-down box. The table has two columns - **Arm** and **Proportion of Alpha**. In the column **Proportion of Alpha**, you have to specify the proportion of total alpha you want to spend in each test. Ideally, the values in this column should add up to 1; if not, then **East** will normalize it to add them up to 1. By default, **East** distributes the total alpha equally among all tests. Here we have 4 tests in total, therefore each of the tests have proportion of alpha as $1/4$ or 0.25 . You can specify other proportions as well. For this example, keep the equal proportion of alpha for each test.
5. Click **OK** to analyze the data. The output will be displayed in the main window

once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis

$H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: UnEqual
 Multiple Comparison Procedure: Weighted Bonferroni
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail
 Table For Proportion of Alpha:

| Arms | Proportion of Alpha |
|------|---------------------|
| D1 | 0.25 |
| D2 | 0.25 |
| D3 | 0.25 |
| D4 | 0.25 |

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 1 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.031 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.045 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 0 | 0.001 | 3.585 | INF |

Adjusted Global p-value: 0.001
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 0.982, 0.031, 0.044 and 0.001, respectively. Therefore, after multiplicity adjustment according to Sidak procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

Notice that the adjusted p-values in weighted Bonferroni MC procedure and the simple Bonferroni procedures are identical. This is because the weighted Bonferroni procedure with equal proportion reduces to the simple Bonferroni procedure.

84 Analysis-Multiple Comparison Procedures for Continuous Data

84.6 Step down MC procedure

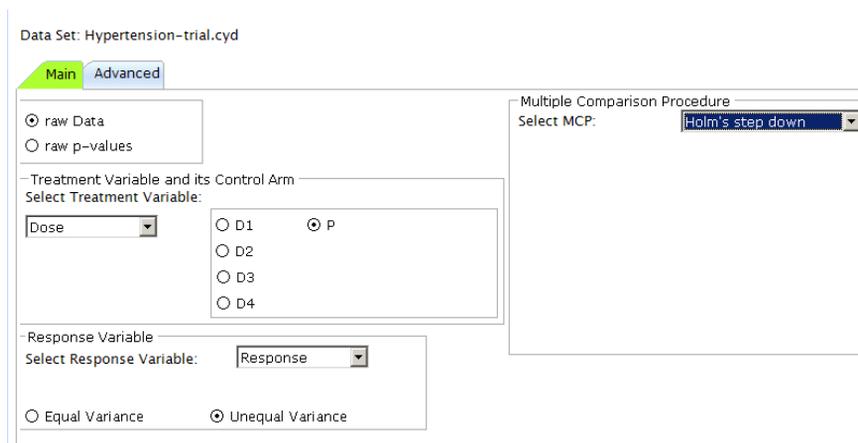
In the single step MC procedures, the decision to reject any hypothesis does not depend on the decision to reject other hypotheses. On the other hand, in the stepwise procedures decision of one hypothesis test can influence the decisions on the other tests of hypotheses. There are two types of stepwise procedures. One type of procedures proceeds in data-driven order. The other type proceeds in a fixed order set a priori. Stepwise tests in a data-driven order can proceed in step-down or step-up manner. **East** supports Holm step-down MC procedure which start with the most significant comparison and continue as long as tests are significant until the test for certain hypothesis fails. The testing procedure stops at the first time a non-significant comparison occurs and all remaining hypotheses will be retained. In i -th step, $H_{(k-i)}$ is rejected if $p_{(k-i)} \leq \frac{\alpha}{i}$ and go to the next step.

Example: Holm's step-down

Dataset: Hypertension-trial.cyd as described in Section 84.2

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below.



4. Click **OK** to analyze the data. The output will be displayed in the main window

once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis

$H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: UnEqual
 Multiple Comparison Procedure: Holm's step down
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 0.634 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.023 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.023 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 0 | 0.001 | 3.585 | INF |

Adjusted Global p-value: 0.001
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 0.634, 0.023, 0.023 and 0.001, respectively. Therefore, after multiplicity adjustment according to Holm's step-down procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

84.7 Data-driven step-up MC procedures

Step-up tests start with the least significant comparison and continue as long as tests are not significant until the first time when a significant comparison occurs and all remaining hypotheses will be rejected. East supports two such MC procedures - Hochberg step-up and Hommel step-up procedures. In the Hochberg step-up procedure, in i -th step $H_{(k-i)}$ is retained if $p_{(k-i)} > \frac{\alpha}{i}$. In the Hommel step-up procedure, in i -th step $H_{(k-i)}$ is retained if $p_{(k-j)} > \frac{i-j+1}{i} \alpha$ for $j = 1, \dots, i$. Fixed sequence test and fallback test are the types of tests which proceed in a prespecified order.

84 Analysis-Multiple Comparison Procedures for Continuous Data

Example: Hochberg's step-up procedure

Dataset: Hypertension-trial.cyd as described in Section 84.2

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below.

The screenshot shows the 'Multiple Comparison Procedure' dialog box for the dataset 'Hypertension-trial.cyd'. The 'Main' tab is selected. The 'raw Data' radio button is chosen. Under 'Treatment Variable and its Control Arm', 'Dose' is selected in the dropdown, and the 'P' radio button is selected. Under 'Response Variable', 'Response' is selected in the dropdown. The 'Unequal Variance' radio button is selected. On the right side, 'Hochberg's step up' is selected in the 'Select MCP:' dropdown.

4. Click **OK** to analyze the data. The output will be displayed in the main window

once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis
 $H_1: \mu_i - \mu_0 \leq 0$ Vs. $K_i: \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: Unequal
 Multiple Comparison Procedure: Hochberg's step Up
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 0.634 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.022 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.022 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 0 | 0.001 | 3.585 | INF |

Adjusted Global p-value: 0.001
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 0.634, 0.022, 0.022 and 0.001, respectively. Therefore, after multiplicity adjustment according to Hochberg's step-up procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

Example: Hommel's step-up procedure

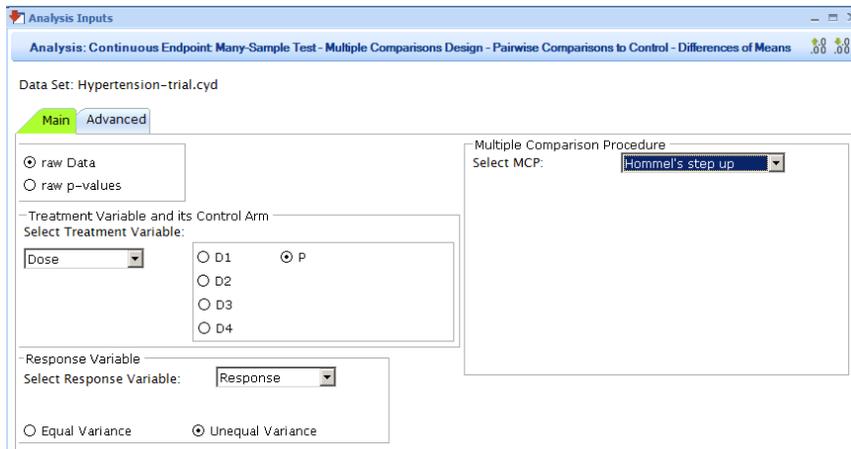
Dataset: Hypertension-trial.cyd as described in Section 84.2

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown

84 Analysis-Multiple Comparison Procedures for Continuous Data

below.



4. Click **OK** to analyze the data. The output will be displayed in the main window

once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

$H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: UnEqual
 Multiple Comparison Procedure: Hommel's step up
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|----------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 0.634 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.017 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.022 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 3.196E-4 | 0.001 | 3.585 | INF |

Adjusted Global p-value: 0.001
 Total Number of Records: 130
 Number of Records Rejected: 5
 Total Number of Arms (k): 5

The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 0.634, 0.017, 0.022 and 0.001, respectively. Therefore, after multiplicity adjustment according to Hommel's step-up procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

84.8 Fixed-sequence stepwise MC procedures

In data-driven stepwise procedures, we don't have any control on the order of the hypotheses to be tested. However, sometimes based on our preference or prior knowledge we might want to fix the order of tests a priori. Fixed sequence test and fallback test are the types of tests which proceed in a pre-specified order. East supports both these procedures.

Assume that H_1, H_2, \dots, H_{k-1} are ordered hypotheses and the order is pre-specified so that H_1 is tested first followed by H_2 and so on. Let p_1, p_2, \dots, p_{k-1} be the associated raw marginal p-values. In the fixed sequence testing procedure, for $i = 1, \dots, k - 1$, in i -th step, if $p_i < \alpha$, reject H_i and go to the next step; otherwise

84 Analysis-Multiple Comparison Procedures for Continuous Data

retain H_i, \dots, H_{k-1} and stop.

Fixed sequence testing strategy is optimal when early tests in the sequence have largest treatment effect and performs poorly when early hypotheses have small treatment effect or are nearly true (Westfall and Krishen (2001)). The drawback of fixed sequence test is that once a hypothesis is not rejected no further testing is permitted. This will lead to lower power to reject hypotheses tested later in the sequence.

Fallback test alleviates the above undesirable feature for fixed sequence test. Let w_i be the proportion of α for testing H_i such that $\sum_{i=1}^{k-1} w_i = 1$. In the fixed sequence testing procedure, in i -th step ($i = 1, \dots, k-1$), test H_i at $\alpha_i = \alpha_{i-1} + \alpha w_i$ if H_{i-1} is rejected and at $\alpha_i = \alpha w_i$ if H_{i-1} is retained. If $p_i < \alpha_i$, reject H_i ; otherwise retain it. Unlike the fixed sequence testing approach, the fallback procedure can continue testing even if a non-significant outcome is encountered by utilizing the fallback strategy. If a hypothesis in the sequence is retained, the next hypothesis in the sequence is tested at the level that would have been used by the weighted Bonferroni procedure. With $w_1 = 1$ and $w_2 = \dots = w_{k-1} = 0$, the fallback procedure simplifies to fixed sequence procedure.

Example: Fixed sequence testing procedure

Dataset: Hypertension-trial.cyd as described in Section 84.2

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

**Analysis > (Continuous) Many Samples > (Multiple Comparisons)
Pairwise Comparisons to Controls - Difference of Means**

3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below. Upon selection of **Fixed Sequence** procedure, a table will appear under the drop-down box. The table has two columns - **Arm** and **Test Sequence**. In the column **Test Sequence**, you have to specify the order in which the hypotheses will be tested. Specify 1 for the arm that will be compared first with Placebo, 2 for the arm that will be compared next and so on. By default **East** specifies 1 to the first arm, 2 to the second arm and so on. This default order implies that Dose1 will be compared first with Placebo, then Dose2 will be compared followed by comparison of Dose3 vs. Placebo and finally Dose 4 will be compared with Placebo. However, if we believe that efficacy of drug increases with dose, then the dose groups should be compared in descending order of dose. Therefore, specify 4, 3, 2 and 1 in column **Test Sequence** for D1,

D2, D3 and D4, respectively. This order implies that Dose4 will be compared first with Placebo, then Dose3 will be compared followed by comparison of Dose2 vs. Placebo and finally Dose 1 will be compared with Placebo.

Data Set: Hypertension-trial.cyd

Main
Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
 Select Treatment Variable:

D1 P
 D2
 D3
 D4

Response Variable
 Select Response Variable:

Equal Variance Unequal Variance

Multiple Comparison Procedure
 Select MCP:

No. of Treatment Arms:

Table of Treatment specific parameters:

| Arm | Test Sequence |
|-----|---------------|
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 4 |

Click **OK** to analyze the data. The output will be displayed in the main window

84 Analysis-Multiple Comparison Procedures for Continuous Data

once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis
 $H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: UnEqual
 Multiple Comparison Procedure: Fixed Sequence
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail
 Table For Proportion of Alpha:

| Arms | Test Sequence |
|------|---------------|
| D1 | 4 |
| D2 | 3 |
| D3 | 2 |
| D4 | 1 |

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 0.634 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.011 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.011 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 0 | 0 | 3.585 | INF |

Adjusted Global p-value: 0
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The input section of the output displays the tests sequence along with the other input values we have provided. The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 0.634, 0.011, 0.011 and 0.000, respectively. Therefore, after multiplicity adjustment according to fixed sequence procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

Example; Fallback procedure

Dataset: [Hypertension-trial.cyd](#) as described in Section [84.2](#)

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below. Upon selection of **Fallback** procedure, a table will appear under the drop-down box. The table has three columns - **Arm**, **Proportion of Alpha** and **Test Sequence**. Specify **4, 3, 2** and **1** in column **Test Sequence** for D1, D2, D3 and D4, respectively. For this example, keep the equal proportion of alpha for each test in the column **Proportion of Alpha**.

Data Set: Hypertension-trial.cyd

Main Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
 Select Treatment Variable: Dose

D1 P
 D2
 D3
 D4

Response Variable
 Select Response Variable: Response

Equal Variance Unequal Variance

Multiple Comparison Procedure
 Select MCP: Fallback

No. of Treatment Arms: 4

Table of Treatment specific parameters:

| Arm | Proportion of... | Test Sequence |
|-----|------------------|---------------|
| 1 | 0.25 | 4 |
| 2 | 0.25 | 3 |
| 3 | 0.25 | 2 |
| 4 | 0.25 | 1 |

4. Click **OK** to analyze the data. The output will be displayed in the main window

84 Analysis-Multiple Comparison Procedures for Continuous Data

once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis
 $H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: UnEqual
 Multiple Comparison Procedure: Fallback
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Table For Proportion of Alpha:

| Arms | Proportion of Alpha | Test Sequence |
|------|---------------------|---------------|
| D1 | 0.25 | 4 |
| D2 | 0.25 | 3 |
| D3 | 0.25 | 2 |
| D4 | 0.25 | 1 |

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 0.634 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.022 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.022 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 0 | 0.001 | 3.585 | INF |

Adjusted Global p-value: 0.001
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The input section of the output displays the tests sequence along with the other input values we have provided. The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs. Placebo and Dose4 vs. Placebo are 0.634, 0.022, 0.022 and 0.001, respectively. Therefore, after multiplicity adjustment according to fallback procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

84.9 Example: Raw p-values as input

Suppose we don't have the dataset containing all the observations, rather we have the raw p-values and we want to adjust these using Bonferroni procedure. Here we will consider the 4 raw p-values returned by **East** using **Hypertension-trial.cyd** in all the above output. These p-values are 0.634, 0.008, 0.011 and 0.000. We will use these raw

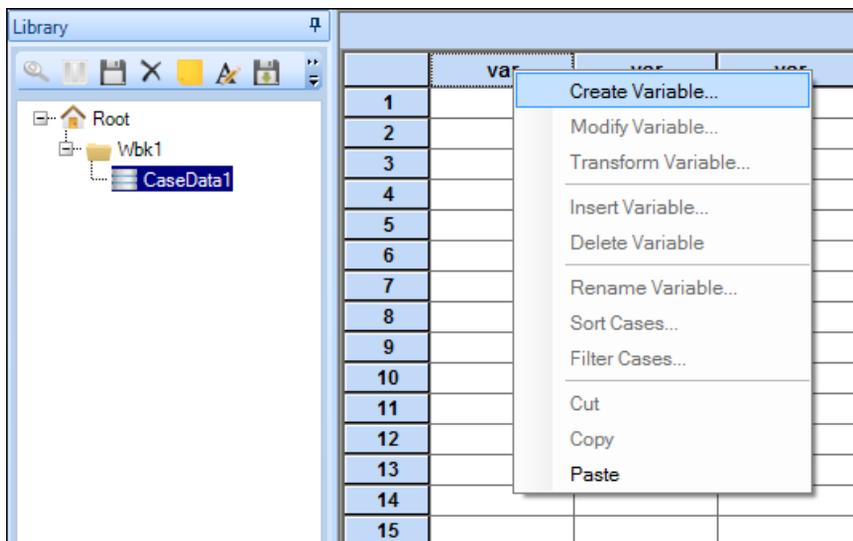
p-values to obtain adjusted p-values. In order to do this, first, we need to create a dataset containing these p-values.

Dataset: New Dataset to be created.

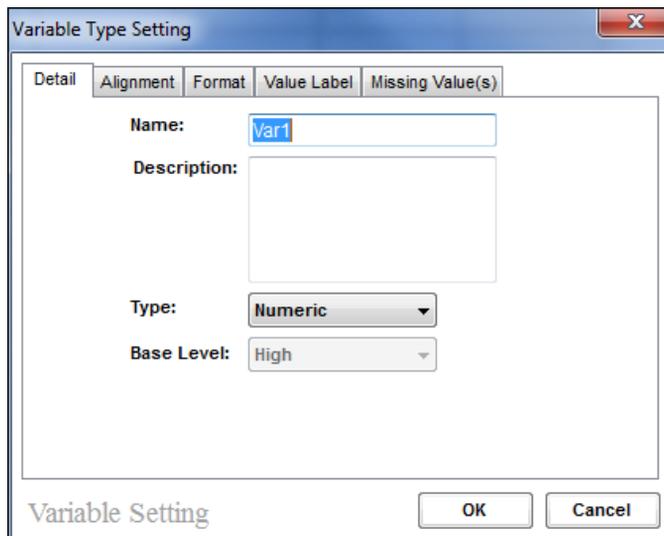
84 Analysis-Multiple Comparison Procedures for Continuous Data

Analysis Steps:

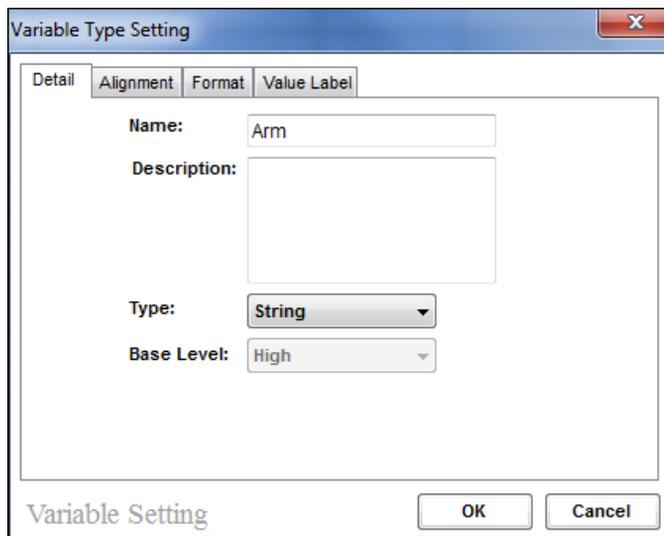
1. Choose textbf New > Case Data. This will open a black dataset in the main window. Now right click on the column header and click **Create Variable** as shown below.



2. This will bring up the following **Variable Type Setting** dialog box.

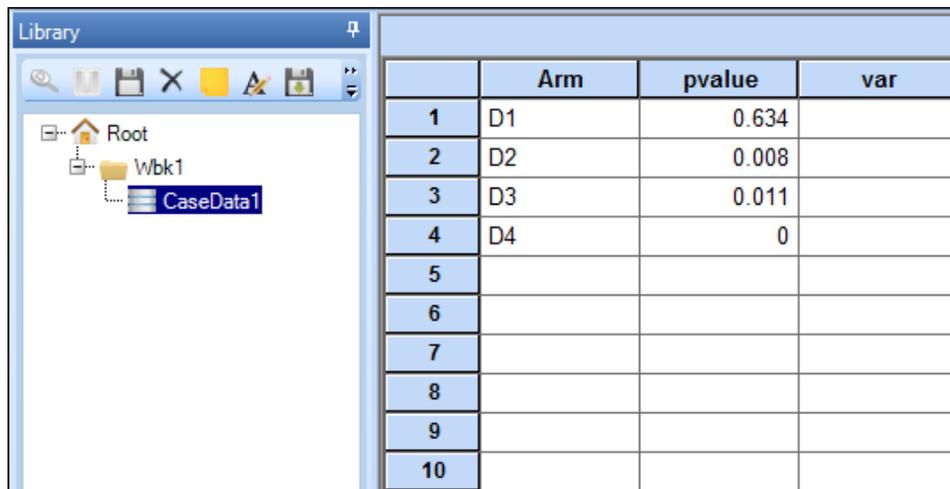


3. Type in **Arm** for **Name** and choose the type of variable as **String**.



84 Analysis-Multiple Comparison Procedures for Continuous Data

- Click **OK** and this will add a column with name **Arm** in the dataset. Similarly, create a numeric column with label **pvalue**. Now, enter the values in the table as follows:



| | Arm | pvalue | var |
|----|-----|--------|-----|
| 1 | D1 | 0.634 | |
| 2 | D2 | 0.008 | |
| 3 | D3 | 0.011 | |
| 4 | D4 | 0 | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |

5. East assigns a default name **CaseData1** to this dataset.
6. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
7. This will display several input fields associated with multiple comparison test in the main window. In the **Main tab**, select the radio-button corresponding to **raw p-values**. In the ensuing two boxes, select **Arm** as **Treatment variable** and select **pvalue** for **Select raw p-values**. Choose **Bonferroni** from the drop-down list in **Select MCP**.

Analysis: Continuous Endpoint: Many-Sample Test - Multi-Arm Design - Pairwise Comparisons to Control - Differences of Means

Data Set: CaseData1

The screenshot displays the software's configuration window for a multiple comparison test. It features two tabs: 'Main' and 'Advanced'. Under the 'Main' tab, there are two radio buttons: 'raw Data' (unselected) and 'raw p-values' (selected). Below these, there are two sections for variable selection. The first section, 'Treatment Variable', has a dropdown menu set to 'Arm' and a list of options: D1, D2, D3, and D4. The second section, 'Raw p-value Variable', has a dropdown menu set to 'pvalue'. To the right of these sections is a separate box titled 'Multiple Comparisons Procedures' containing a dropdown menu labeled 'Select MCP:' which is set to 'Bonferroni'.

84 Analysis-Multiple Comparison Procedures for Continuous Data

8. Click **OK**. The output will be displayed in the main window.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis
 $H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: CaseData1
 Treatment Variable: Arm
 Raw p-values variable: pvalue
 Multiple Comparison Procedure: Bonferroni
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | p-value | |
|-----|---------|----------|
| | Raw | Adjusted |
| D1 | 0.634 | 1 |
| D2 | 0.008 | 0.032 |
| D3 | 0.011 | 0.044 |
| D4 | 0 | 0 |

Adjusted Global p-value: 0
 Total no. of records: 4
 No. of records rejected: 0
 Total Number of Arms (k): 4

The adjusted p-values for D1, D2, D3 and D4 are 1, 0.032, 0.044 and 0.000, respectively. Note that these adjusted p-values are very close to what we have obtained with Bonferroni procedure using the dataset [Hypertension-trial.cyd](#). Ideally, both set of p-values should exactly match. The difference in p-values is only due to rounding error.

85 *Analysis-Multiple Endpoints for Continuous Data*

In Chapter 16, we have seen how to evaluate different gatekeeping procedures through intensive simulations. In this chapter, we will illustrate how to analyze a trial with gatekeeping multiple comparison procedures. Consider the Alzheimer's disease example reported in Reisberg et al. 2003. This study is designed to investigate memantine, an N-methyl-D-aspartate (NMDA) antagonist, for the treatment of Alzheimer's disease in which patients with moderate-to-severe Alzheimer's disease were randomly assigned to receive placebo or 20 mg of memantine daily for 28 weeks. The two primary efficacy variables were: (1) the Clinician's Interview-Based Impression of Change Plus Caregiver Input (CIBIC-Plus) global score at 28 weeks, (2) the change from base line to week 28 in the Alzheimer's Disease Cooperative Study Activities of Daily Living Inventory modified for severe dementia (ADCS-ADLsev). The CIBIC-Plus measures overall global change relative to base line and is scored on a seven-point scale ranging from 1 (markedly improved) to 7 (markedly worse). The secondary efficacy endpoints included the Severe Impairment Battery and other measures of cognition, function, and behavior. Suppose that the trial is declared successful only if the treatment effect is demonstrated on both endpoints. If the trial is successful, it is of interest to assess the two secondary endpoints: (1) Severe Impairment Battery (SIB), (2) Mini-Mental State Examination (MMSE). The data set is saved in the installation folder of EAST as Alzheimer.csv. To analyze this data set, we need to import the data into EAST by clicking on the Import icon as seen in the following screen.



Select the Alzheimer.csv file and click OK to see the data set displayed in EAST. The

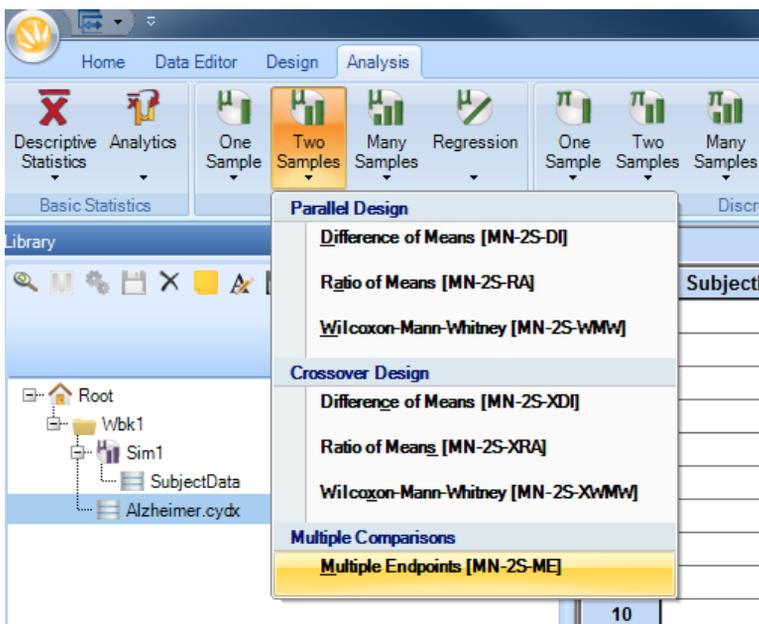
85 Analysis-Multiple Endpoints for Continuous Data

following screen shows a snapshot of the data set.

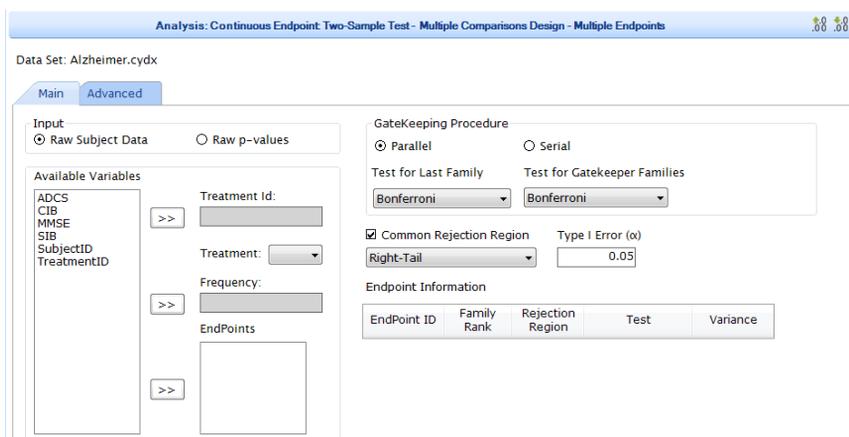
| SubjectID | TreatmentID | CIB | ADCS | SIB | MMSE |
|-----------|-------------|------------|-------------|-------------|--------------|
| 1 | 0 | 5.66492653 | 4.36394922 | -6.32570582 | -3.32797968 |
| 2 | 0 | 3.95074349 | -3.3850377 | 7.06432058 | -0.483275732 |
| 3 | 0 | 4.36113329 | -7.09915667 | -12.5260734 | -2.08573408 |
| 4 | 0 | 3.19733356 | -12.2679343 | -27.2717159 | -4.24276931 |
| 5 | 0 | 3.59878247 | -6.02594982 | -27.1493254 | -3.27394314 |
| 6 | 0 | 4.60578509 | -10.071646 | -20.9660857 | -3.32429613 |
| 7 | 0 | 5.86933156 | 0.525622886 | -8.14458525 | 2.68263096 |
| 8 | 0 | 4.73326277 | -5.41094781 | -6.79993922 | -1.27626971 |
| 9 | 0 | 4.74888428 | -9.43789172 | -33.5473889 | -0.853782244 |
| 10 | 0 | 3.48522851 | -5.75471209 | -7.64045638 | -5.57948689 |
| 11 | 0 | 3.2053906 | -14.8363446 | -17.7253033 | -5.44609824 |
| 12 | 0 | 6.75611159 | 4.52154357 | 23.8728393 | 1.47453882 |
| 13 | 0 | 3.929523 | -2.8980647 | -19.0296683 | -2.55714285 |
| 14 | 0 | 5.37510233 | -8.38883153 | -15.1115339 | 3.31528606 |
| 15 | 0 | 3.95012484 | -9.23173693 | -1.39008734 | 4.34674624 |
| 16 | 0 | 5.84104094 | 3.18117957 | -11.1542966 | 6.16735651 |
| 17 | 0 | 5.27755294 | -4.42540448 | -25.2515189 | -4.67474173 |

Now click on the Analysis menu on the top of EAST window, select Two Samples and

then select Multiple Comparisons-Multiple Endpoints from the dropdown list.



The main input dialog window pops up as seen in the following screen.



85 Analysis-Multiple Endpoints for Continuous Data

the Alzheimer’s disease example, the data is raw subject level data so we select the left radio button. The left bottom panel of the screen displays all the variables contained in the data set. We need to specify which variable contains the information on treatment group ID for each subject and further specify which one is active treatment group. The next input is to identify all the endpoints to be analyzed. For this example, CIBIC-Plus and ADCS-ADLsev constitute the primary family of endpoints. SIB and MMSE constitutes the secondary family of endpoints. Suppose we need to analyze the data using serial gatekeeping procedure and using Bonferroni to adjust the multiplicity for the two endpoints from the secondary family. After filling in all inputs, the screen looks as follows

Analysis: Continuous Endpoint Two-Sample Test - Multiple Comparisons Design - Multiple Endpoints

Data Set: Alzheimer.cyx

Main
Advanced

Input

Raw Subject Data Raw p-values

Available Variables

SubjectID

Treatment Id:

TreatmentID

Treatment: 1

Frequency:

EndPoints

ADCS
CIB
MMSE
SIB

GateKeeping Procedure

Parallel Serial

Test for Last Family

Bonferroni

Common Rejection Region Type I Error (α)

0.05

Endpoint Information

| EndPoint ID | Family Rank | Rejection Region | Test | Variance |
|-------------|-------------|------------------|---------------|----------|
| ADCS | 1 | Right-Tail | Diff of Means | Equal |
| CIB | 1 | Left-Tail | Diff of Means | Equal |
| MMSE | 2 | Right-Tail | Diff of Means | Equal |
| SIB | 2 | Right-Tail | Diff of Means | Equal |

Now click on OK button on the right bottom of the screen to run the analysis. The

following screen displays the detailed output of this analysis.

Analysis: Continuous Endpoint: Two-Sample Test - Multiple Comparisons Design - Multiple Endp

Data File: Alzheimer.cydx
 Data Input Method: Raw Data
 Treatment Variable: TreatmentID
 Total Number of Endpoints: 4
 Type I Error (α): 0.05
 Gatekeeping Procedure: Serial
 Test for Last Family: Bonferroni

Output

Total Number of Records: 100
 Number of Records Rejected: 0
 Summary of each Endpoint:

| EndPoint ID | Test | Variance Type | Sample Size | Control Mean | Treatment Mean | Est. of Delta | Std. Err. of Est. of Delta | Test Statistic | 95% Confidence Interval of Delta (Two-Sided) | |
|-------------|---------------|---------------|-------------|--------------|----------------|---------------|----------------------------|----------------|--|-------------|
| | | | | | | | | | Lower Limit | Upper Limit |
| ADCS | Diff of Means | Equal | 100 | -3.836 | -1.994 | 1.843 | 1.275 | 1.445 | -0.687 | 4.373 |
| CIB | Diff of Means | Equal | 100 | 4.755 | 4.349 | -0.406 | 0.222 | -1.832 | -0.847 | 0.034 |
| MMSE | Diff of Means | Equal | 100 | -0.639 | -0.734 | -0.095 | 0.595 | -0.16 | -1.277 | 1.086 |
| SIB | Diff of Means | Equal | 100 | -9.519 | -8.983 | 0.536 | 2.265 | 0.236 | -3.96 | 5.031 |

Inference of Gatekeeping Procedures:

| Family Rank | EndPoint ID | Rejection Region | Test Statistic | Degrees of Freedom(DF) | p-Value | | Status |
|-------------------|-------------|------------------|----------------|------------------------|---------|----------|------------------|
| | | | | | Raw | Adjusted | |
| 1 | ADCS | Right-Tail | 1.445 | 98 | 0.076 | 0.076 | Unable to Reject |
| | CIB | Left-Tail | -1.832 | 98 | 0.035 | 0.076 | Unable to Reject |
| Not Passed | | | | | | | |
| 2 | MMSE | Right-Tail | -0.16 | 98 | 0.563 | 1 | Not Tested |
| | SIB | Right-Tail | 0.236 | 98 | 0.407 | 0.814 | Not Tested |

*Using Bonferroni test for the last family in the Serial gatekeeping procedure, the treatment arm is not significantly different from the control arm on the co-primary endpoints. As a result, the other non-primary endpoints are not tested.

The first table shows the summary statistics for each endpoint including mean for each treatment group, estimate of treatment effect, standard error of the effect estimate, test statistic and marginal two-sided confidence interval. The second table shows the inference summary including raw p-values, multiplicity adjusted p-values with the gatekeeping procedure and significance status. It also shows whether the primary family is passed as the serial gatekeeper for the secondary family of endpoints.

86 Analysis-Binomial Superiority One-Sample

This chapter demonstrates how **East** can be used to perform inferences on data collected from a single-sample superiority study when the observations on a binary variable have an unknown probability of success. You need to either test a null hypothesis about the probability, or compute an exact confidence interval for the probability of success. The section also discusses the analysis of paired data on a binary random variable.

Chapter 22 deals with the design, simulation and interim monitoring of these types of trials with reference to a single sample test for proportion.

East supports both the asymptotic and exact analysis of these tests. These are accessible from the **Analysis** menu and allow the validation of whether the data supports the null or alternative hypothesis of the study. Analysis of a single mean superiority test is discussed in section 86.1, while McNemar's test for paired observations is discussed in section 86.2.

86.1 Example: Single Proportion

Dataset: Pilot.cyx

Data Description

In a pilot study of a new drug, 20 patients were treated. The column **Response** displays the successes and failures after administering the drug. There were 4 responders (successes) and 16 non-responders (failures).

Purpose of the Analysis:

Consider the null hypothesis: $H_0: \pi = \pi_0$ to be tested against a two-sided alternative hypothesis $H_1: \pi \neq \pi_0$ or a one-sided alternative hypothesis $H_1: \pi < \pi_0$ or $H_1: \pi > \pi_0$.

In this analysis, the hypothesis is tested asymptotically as well as using Exact Inference. We will obtain a 95% confidence interval for the underlying success rate and test the null hypothesis that $\pi = 0.05$. We would also like to compute the power of the test for the alternative hypothesis that $\pi = 0.30$.

Analysis Steps: Asymptotic Test

1. Open the dataset from **Samples** folder.

2. Choose the menu item
Analysis > (Discrete) One Sample > (Single Arm Design) Single Proportion
3. In the ensuing dialog box (under the **Main** tab) choose the variables as shown below. To run the **Asymptotic** test, do not check the **Perform Exact Computation** checkbox.

Analysis: Discrete Endpoint One-Sample Test - Single Arm Design - Single Proportion

Data Set: Pilot.cyx

Main Advanced

Response Variable: Proportion Specified under H0:

Response Value:

Perform Exact Computations

86 Analysis-Binomial Superiority One-Sample

- Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: Single Proportion

$H_0 : \pi = \pi_0$ Vs. $H_1 : \pi \neq \pi_0$ for 2-Sided test
 Either $H_1 : \pi > \pi_0$
 Or $H_1 : \pi < \pi_0$ for 1-Sided test

Input Parameters

Data File: Pilot.cydx
 Response Variable: Response(1)
 Exact p-values for testing π : 0.05
 Confidence Level: 0.95

Output

Response Variable: (Response)
 Total Number of Records: 20
 Number of Records Rejected: 0

Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|---|-------|
| 1 | 16 | 4 | 20 |
| Total | 16 | 4 | 20 |

Test of Hypothesis:

| Type | Proportion | Std. Error | Std. Error (H0) | Test Statistic | (1-Sided) | | (2-Sided) | 95% Confidence Interval(2-Sided) | |
|------------|------------|------------|-----------------|----------------|-----------|------|-----------|----------------------------------|-------------|
| | | | | | p-value | Tail | p-value | Lower Limit | Upper Limit |
| Asymptotic | 0.2 | 0.089 | 0.049 | 3.078 | 0.001 | G.E. | 0.002 | 0.025 | 0.375 |

Note that the test statistic is 3.078 with a 1-sided p-value of 0.001. Since the hypothesized proportion under the null hypothesis is 0.05 which is less than the observed proportion of responders in the data, namely 0.2, the tail type considered for one sided alternative hypothesis is G.E. meaning greater than or equal to. The null hypothesis that $\pi = 0.05$ is rejected at the 5% significance level.

Analysis Steps: Exact Test

- Click the **Analysis Inputs/Outputs** tab on the status bar below.
- Under the **Main** tab, select variables as shown below. Make sure to check the

Perform Exact Computation checkbox.

Analysis: Discrete Endpoint One-Sample Test - Single Arm Design - Single Proportion

Data Set: Pilot.cyx

Main **Advanced**

Compute Blyth-Still-Casella Confidence Interval

Response Variable: Proportion Specified under H0:

Response Value:

Perform Exact Computations

- Under the **Advanced** tab leave the fields **By Variable 1** and **By Variable 2** blank. Select the **Compute Power** checkbox, enter the value 0.05 for **Alpha** and 0.3 for **Probability under H1**. Keep the default value 0.95 for **Confidence Level**.

Analysis: Discrete Endpoint One-Sample Test - Single Arm Design - Single Proportion

Data Set: Pilot.cyx

Main **Advanced**

Compute Power

Alpha: Probability under H₁:

By Variable 1:

By Variable 2:

Confidence Level:

86 Analysis-Binomial Superiority One-Sample

4. Click **OK** to start the analysis. The result is displayed in the main window.

Analysis: Binomial Response: Single Proportion

$H_0 : \pi = \pi_0$ Vs. $H_1 : \pi \neq \pi_0$ for 2-Sided test
 Either $H_1 : \pi > \pi_0$
 Or $H_1 : \pi < \pi_0$ for 1-Sided test

Input Parameters

Data File: Pilot.cyx
 Response Variable: Response(1)
 Exact p-values for testing π : 0.05
 Confidence Level: 0.95

Output

Response Variable: (Response)
 Total Number of Records: 20
 Number of Records Rejected: 0

Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|---|-------|
| 1 | 16 | 4 | 20 |
| Total | 16 | 4 | 20 |

Test of Hypothesis:

| Type | Statistic | p-value | | | | |
|-------|-----------|---------|---------|-----------|------|---------|
| | | Tail | 1-Sided | 2*1 Sided | Tail | 2-Sided |
| Exact | 4 | G.E. | 0.016 | 0.032 | G.E. | 0.032 |

Confidence Interval:

| Type | 95% Confidence Interval | | |
|---------------------|-------------------------|-------------|-------------|
| | MLE of π | Lower Limit | Upper Limit |
| Clopper-Pearson | 0.2 | 0.057 | 0.437 |
| Blyth-Still-Casella | 0.2 | 0.071 | 0.411 |

Exact Conditional Power:

| | |
|--------------------------------|-------|
| Null Hypothesis (H0) | 0.05 |
| Alternative Hypothesis (H1) | 0.3 |
| Desired Significance Level | 0.05 |
| Cut-off value for rejecting H0 | 4 |
| Power | 0.893 |

The exact 95% confidence interval using the Clopper-Pearson method is (0.057, 0.437). Notice that the Blyth-Still-Casella confidence interval is (0.071, 0.411), which is thus about 10% narrower than the Clopper-Pearson confidence interval. The exact 1-sided p-value is 0.016, and so the null hypothesis that $\pi = 0.05$ is rejected at the 5% significance level. The power of the test for the Type-I error $\alpha = .05$, where testing $H_0 : \pi = 0.05$ against $H_1 : \pi = 0.30$ at $\alpha = 0.05$, is 0.893.

86.2 Example: McNemar's Test for Matched Pairs . Dataset: Vote.cyx

Data Description

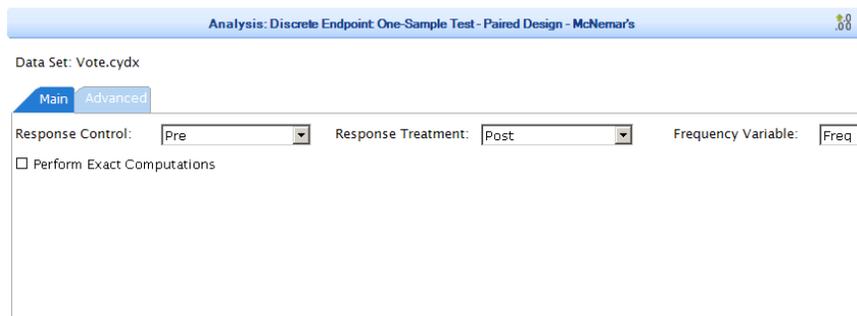
This data are taken from Siegel and Castellan (1988, page 77). It shows changes in preference for Presidential candidates before and after a television debate.

Table 86.1: Preference for Presidential Candidates

| Preference Before TV Debate | Preference After TV Debate | |
|-----------------------------|----------------------------|--------|
| | Carter | Reagan |
| Carter | 28 | 13 |
| Reagan | 7 | 27 |

Analysis Steps: Asymptotic Test

1. Open the dataset from **Samples** folder.
2. Choose the menu item **Analysis > (Discrete) One Sample > (Paired Design) McNemar's**
3. In the ensuing dialog box (under the **Main** tab) choose the variables as shown below. To run the **Asymptotic** test, do not check the **Perform Exact Computation** checkbox.



4. Under the **Advanced** tab, leave the fields **By Variable 1** and **By Variable 2** blank, keep the **Confidence Level** as 0.95.

86 Analysis-Binomial Superiority One-Sample

5. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: McNemar's test for Matched Pairs

Hypothesis

$H_0 : \pi_t - \pi_c = 0$ Vs. $H_1 : \pi_t - \pi_c \neq 0$ for 2-sided Test

Either $H_1 : \pi_t - \pi_c < 0$

Or $H_1 : \pi_t - \pi_c > 0$ for 1-sided Test

Input Parameters

Data File: Vote.cydx
 Response Control: Pre
 Response Treatment: Post
 Frequency Variable: Freq
 Confidence Level: 0.95

Output

Total no. of records: 4

No. of records rejected: 0

Table of observed frequencies:

| | Carter | Reagan | Total |
|--------|--------|--------|-------|
| Carter | 28 | 7 | 35 |
| Reagan | 13 | 27 | 40 |
| Total | 41 | 34 | 75 |

Summary of the Observed Data:

| Response Variable | n | Responses | Proportion |
|-------------------|----|-----------|------------|
| Pre | 75 | 41 | 0.547 |
| Post | 75 | 35 | 0.467 |

Test of Hypothesis:

| n | Diff. of Proportions | Std. Error | Test Statistic | (1-Sided) | | (2-Sided) | | 95% Confidence Interval(2-Sided) | |
|----|----------------------|------------|----------------|-----------|------|-----------|-------------|----------------------------------|--|
| | | | | p-value | Tail | p-value | Lower Limit | Upper Limit | |
| 75 | -0.08 | 0.06 | -1.342 | 0.09 | L.E. | 0.18 | -0.197 | 0.037 | |

The negative sign of the test statistic indicates that of the 20 discordant pairs, more switched preference from Carter to Reagan (13) than those switched preference from Reagan to Carter (7). The 2-sided p-value is 0.18 indicating not a significant change in preference for Presidential candidates before and after the television debate. The 95% confidence interval for difference of proportions based on the data is $(-0.197, 0.037)$. The fact that this interval includes 0 indicate that we are unable to reject the null hypothesis of no difference on the basis of the data.

Analysis Steps: Exact Test

1. Click the **Analysis Inputs/Outputs** tab on the status bar below.
2. In the ensuing dialog box (under the **Main**) tab select the **Perform Exact Computation** checkbox.

Analysis: Discrete Endpoint: One-Sample Test - Paired Design - McNemar's

Data Set: Vote.cydx

Main Advanced

Response Control: Pre Response Treatment: Post Frequency Variable: Freq

Perform Exact Computation

3. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: McNemar's test for Matched Pairs

Hypothesis

$H_0 : \pi_t - \pi_c = 0$ Vs. $H_1 : \pi_t - \pi_c \neq 0$ for 2-Sided Test

Either $H_1 : \pi_t - \pi_c < 0$

Or $H_1 : \pi_t - \pi_c > 0$ for 1-Sided Test

Input Parameters

Data File: Vote.cydx
 Response Control: Pre
 Response Treatment: Post
 Frequency Variable: Freq
 Confidence Level: 0.95

Output

Total Number of Records: 4

Number of Records Rejected: 0

Table of Observed Frequencies:

| | Carter | Reagan | Total |
|--------|--------|--------|-------|
| Carter | 28 | 7 | 35 |
| Reagan | 13 | 27 | 40 |
| Total | 41 | 34 | 75 |

86 Analysis-Binomial Superiority One-Sample

Summary of the Observed Data:

| Response Variable | n | Responses | Proportion |
|-------------------|----|-----------|------------|
| Pre | 75 | 41 | 0.547 |
| Post | 75 | 35 | 0.467 |

Test of Hypothesis:

| Type | n | Diff. of Proportions | Std. Error | Test Statistic | (1-Sided) | | (2-Sided) |
|-------|----|----------------------|------------|----------------|-----------|------|-----------|
| | | | | | p-value | Tail | p-value |
| Exact | 75 | -0.08 | 0.06 | -1.342 | 0.132 | L.E. | 0.263 |

The exact p-value is 0.263 indicating not a significant change in preference for Presidential candidates before and after the television debate.

87 *Analysis-Binomial Superiority Two-Sample*

In clinical trials involving binomial endpoint data, the interest lies in investigating if the subjects on treatment arm possess significantly different proportion of some characteristic, such as proportion of patients developing tumor, showing some side effect, requiring special attention etc as against the same on the control arm.

Chapter 23 deals with designing of such clinical trials considering difference of proportions, ratio of proportions or odds ratio of proportions of the two populations.

This chapter explores how **East** is used to analyze data from two independent binomial samples generated while conducting a superiority trial. Assume that the data are sampled independently from two binomial populations with response probabilities π_t and π_c for treatment and control. This comparison is based on difference of response probabilities, ratio of proportions or odds ratio of the two populations.

87.1 *Example: Difference of Proportions-Asymptotic*

Dataset: Clntrt.cydx

Data Description:

The following 2×2 table is obtained from a clinical trial of two treatments with a binary end-point:

| Outcome | Drug A | Drug B |
|-------------|--------|--------|
| Response | 5 | 9 |
| No Response | 5 | 1 |

The Drug B is the treatment whereas Drug A is control.

Purpose of the Analysis:

The following 2×2 table is obtained from a clinical trial of two treatments with a binary end-point:

| Outcome | Drug A | Drug B |
|-------------|--------|--------|
| Response | 5 | 9 |
| No Response | 5 | 1 |

87 Analysis-Binomial Superiority Two-Sample

To test the hypothesis $H_0 : \delta = 0$ against a 1-sided alternative hypothesis $H_1 : \delta > 0$. For this analysis, consider 1-sided type I error of 0.05.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item **Analysis > (Discrete) Two Samples > (Parallel Design) Difference of Proportions**
3. In the ensuing dialog box (under the **Main**) tab select **Superiority** as the **Trial Type**. Choose other variables as shown below. Do not check **Perform Exact Computation** checkbox.

Analysis: Discrete Endpoint Two-Sample Test - Parallel Design - Difference of Proportions

Data Set: Clntrt.cydX

Main Advanced

Trial Type: Superiority Response Variable: Outcome Frequency Variable: Freq

Population Id: Drug Response Value: 1

Control: A

Treatment: B

Perform Exact Computation

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: Difference of Proportions

Hypothesis

$H_0 : \pi_t - \pi_c = 0$ Vs. $H_1 : \pi_t - \pi_c \neq 0$ for 2-Sided test

Either $H_1 : \pi_t - \pi_c < 0$

Or $H_1 : \pi_t - \pi_c > 0$ for 1-Sided test

Input Parameters

Data File: Clntrt.cydx
 Trial Type: Superiority
 Population Variable: Drug(Treatment=B, Control=A)
 Response Variable: Outcome(1)
 Frequency Variable: Freq
 Confidence Level: 0.95

Output

Response Variable: (Outcome)

Total Number of Records: 4

Number of Records Rejected: 0

Table of Observed Frequencies:

| | A | B | Total |
|-------|----|----|-------|
| 1 | 5 | 9 | 14 |
| 0 | 5 | 1 | 6 |
| Total | 10 | 10 | 20 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| A | 10 | 5 | 0.5 |
| B | 10 | 9 | 0.9 |

Test of Hypothesis:

| Type | n | Difference of Proportions | Std. Error | Test Statistic | 1-Sided | | 2-Sided | 95% Confidence Interval | |
|------------|----|---------------------------|------------|----------------|---------|---------|---------|-------------------------|-------------|
| | | | | | Tail | p-value | p-value | Lower Limit | Upper Limit |
| Asymptotic | 20 | 0.4 | 0.205 | 1.952 | G.E. | 0.025 | 0.051 | -0.002 | 0.699 |

The observed value of test statistic is 1.952. The p-value for 2-sided test is 0.051. The p-values for 2-sided test and for the right tailed test are 0.051 and 0.025 respectively. This p-value is associated with the rejection of $H_0: \delta = 0$ in favor of the alternative hypothesis $H_1: \delta > 0$. East displays the p-value associated with right tailed test on this occasion because $\hat{\delta} > 0$. The 2-sided 95% confidence interval is (-0.002, 0.699).

87 Analysis-Binomial Superiority Two-Sample

The p-value as well as the confidence interval indicates the rejection of null hypothesis and superiority of the drug over Control.

87.2 Example: Difference of Proportions-Exact

Dataset: `Clntrt.cydx` as described in Section 87.1

Purpose of the Analysis:

The following 2×2 table is obtained from a clinical trial of two treatments with a binary end-point:

| Outcome | Drug A | Drug B |
|-------------|--------|--------|
| Response | 5 | 9 |
| No Response | 5 | 1 |

The drug B is the treatment where as drug A is control. To test the hypothesis $H_0 : \delta = 0$ against a 1-sided alternative hypothesis $H_1 : \delta > 0$. For this analysis, consider 1-sided type I error of 0.05.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item
Analysis > (Discrete) Two Samples > (Parallel Design) Difference of Proportions
3. In the ensuing dialog box (under the **Main**) tab select **Superiority** as the **Trial Type**. Choose other variables as shown below. Check **Perform Exact**

Computation checkbox.

Analysis: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions .00 .00

Data Set: Clntrt.cydx

Main Advanced

Trial Type: Superiority Response Variable: Outcome Frequency Variable: Freq
 Population Id: Drug Response Value: 1
 Control: A
 Treatment: B
 Perform Exact Computation

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: Difference of Proportions

Hypothesis

$H_0 : \pi_t - \pi_c = 0$ Vs. $H_1 : \pi_t - \pi_c \neq 0$ for 2-Sided test
 Either $H_1 : \pi_t - \pi_c < 0$
 Or $H_1 : \pi_t - \pi_c > 0$ for 1-Sided test

Input Parameters

Data File: Clntrt.cydx
 Trial Type: Superiority
 Test Type: Wald
 Population Variable: Drug(Treatment=B, Control=A)
 Response Variable: Outcome(1)
 Frequency Variable: Freq
 Confidence Level: 0.95

Output

Response Variable: (Outcome)
 Total Number of Records: 4
 Number of Records Rejected: 0

87 Analysis-Binomial Superiority Two-Sample

Table of Observed Frequencies:

| | A | B | Total |
|-------|----|----|-------|
| 1 | 5 | 9 | 14 |
| 0 | 5 | 1 | 6 |
| Total | 10 | 10 | 20 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| A | 10 | 5 | 0.5 |
| B | 10 | 9 | 0.9 |

Test of Hypothesis:

| Type | n | Difference of Proportions | Std. Error | Test Statistic | 1-Sided | | 2-Sided | 95% Confidence Interval | |
|-------|----|---------------------------|------------|----------------|---------|---------|---------|-------------------------|-------------|
| | | | | | Tail | p-value | p-value | Lower Limit | Upper Limit |
| Exact | 20 | 0.4 | 0.205 | 1.952 | G.E. | 0.031 | 0.062 | -0.02 | 0.741 |

The one-sided p-value as well as the confidence interval indicates the rejection of null hypothesis and superiority of the Treatment over Control.

87.3 Example: Ratio of Proportions-Asymptotic

Dataset: Clntrt.cyx as described in Section 87.1.

Purpose of the Analysis:

In the **Ratio of Proportions** test, let π_t and π_c denote the proportions of the successes from the experimental treatment (T) and the control treatment (C), respectively.

To test the null hypothesis $H_0: \pi_t/\pi_c = 1$ against the 2-sided alternative hypothesis $H_1: \pi_t/\pi_c \neq 1$ or a 1-sided alternative hypothesis $H_1: \pi_t/\pi_c < 1$ or $H_1: \pi_t/\pi_c > 1$.

Analysis Steps

1. Open the dataset from **Samples** folder.
2. Choose the menu item
Analysis > (Discrete) Two Samples > (Parallel Design) Ratio of Proportions
3. In the ensuing dialog box (under the **Main**) tab select **Superiority** as the **Trial Type**. Choose other variables as shown below. Do not check **Perform Exact**

Computation checkbox.

Analysis: Discrete Endpoint: Two-Sample Test - Parallel Design - Ratio of Proportions

Data Set: Clntrt.cydx

Main Advanced

Trial Type: Superiority Response Variable: Outcome Frequency Variable: Freq
 Population Id: Drug Response Value: 1
 Control: A
 Treatment: B
 Perform Exact Computations

- Click **OK** to start the analysis. Upon completion of the analysis, the output is displayed in the main window.

Analysis: Binomial Response: Ratio of Proportions

$H_0 : \pi_t / \pi_c = 1$ Vs. $H_1 : \pi_t / \pi_c \neq 1$ for 2-Sided test
 Either $H_1 : \pi_t / \pi_c < 1$
 Or $H_1 : \pi_t / \pi_c > 1$ for 1-Sided test

Input Parameters

Data File: Clntrt.cydx
 Trial Type: Superiority
 Population Variable: Drug(Treatment=B, Control=A)
 Response Variable: Outcome(1)
 Frequency Variable: Freq
 Confidence Level: 0.95

Output

Response Variable: (Outcome)
 Total Number of Records: 4
 Number of Records Rejected: 0
 Table of Observed Frequencies:

| | A | B | Total |
|-------|----|----|-------|
| 1 | 5 | 9 | 14 |
| 0 | 5 | 1 | 6 |
| Total | 10 | 10 | 20 |

87 Analysis-Binomial Superiority Two-Sample

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| A | 10 | 5 | 0.5 |
| B | 10 | 9 | 0.9 |

Test of Hypothesis:

| Type | n | Ratio of Proportions | Std. Error | Test Statistic | (1-Sided) | | (2-Sided) | 95% Confidence Interval | |
|------------|----|----------------------|------------|----------------|-----------|------|-----------|-------------------------|-------------|
| | | | | | p-value | Tail | p-value | Lower Limit | Upper Limit |
| Asymptotic | 20 | 1.8 | 0.205 | 1.952 | 0.025 | G E | 0.051 | 0.997 | 3.873 |

The observed value of test statistic is 1.952 with a 1-sided p-value equal to 0.025. The 2-sided 95% confidence interval for π_t/π_c is (0.997, 3.873). The null hypothesis is rejected establishing the superiority of the Treatment over Control.

87.4 Example: Ratio of Proportions-Exact

Dataset: Clntrt.cydx as described in Section 87.1.

Purpose of the Analysis:

In the **Ratio of Proportions** test, let π_t and π_c denote the proportions of the successes from the experimental treatment (T) and the control treatment (C), respectively.

To test the null hypothesis $H_0: \pi_t/\pi_c = 1$ against the 2-sided alternative hypothesis $H_1: \pi_t/\pi_c \neq 1$ or a 1-sided alternative hypothesis $H_1: \pi_t/\pi_c < 1$ or $H_1: \pi_t/\pi_c > 1$.

Analysis Steps

1. Open the dataset from **Samples** folder.
2. Choose the menu item
Analysis > (Discrete) Two Samples > (Parallel Design) Ratio of Proportions
3. In the ensuing dialog box (under the **Main**) tab select **Superiority** as the **Trial Type**. Choose other variables as shown below. Check **Perform Exact**

Computation checkbox.

Analysis: Discrete Endpoint: Two-Sample Test - Parallel Design - Ratio of Proportions

Data Set: Clntrt.cyx

Main Advanced

Trial Type: Superiority Response Variable: Outcome Frequency Variable: Freq
 Population Id: Drug Response Value: 1
 Control: A
 Treatment: B
 Perform Exact Computation

- Click **OK** to start the analysis. Upon completion of the analysis, the output is displayed in the main window.

Analysis: Binomial Response: Ratio of Proportions

Hypothesis

$H_0 : \pi_t / \pi_c = 1$ Vs. $H_1 : \pi_t / \pi_c \neq 1$ for 2-Sided test
 Either $H_1 : \pi_t / \pi_c < 1$
 Or $H_1 : \pi_t / \pi_c > 1$ for 1-Sided test

Input Parameters

Data File: Clntrt.cyx
 Trial Type: Superiority
 Population Variable: Drug(Treatment=B, Control=A)
 Response Variable: Outcome(1)
 Frequency Variable: Freq
 Confidence Level: 0.95

Output

Response Variable: (Outcome)
 Total Number of Records: 4
 Number of Records Rejected: 0

87 Analysis-Binomial Superiority Two-Sample

Table of Observed Frequencies:

| | A | B | Total |
|-------|----|----|-------|
| 1 | 5 | 9 | 14 |
| 0 | 5 | 1 | 6 |
| Total | 10 | 10 | 20 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| A | 10 | 5 | 0.5 |
| B | 10 | 9 | 0.9 |

Test of Hypothesis:

| Type | n | Ratio of Proportions | Std. Error | Test Statistic | (1-Sided) | | (2-Sided) | 95% Confidence Interval | |
|-------|----|----------------------|------------|----------------|-----------|------|-----------|-------------------------|-------------|
| | | | | | p-value | Tail | p-value | Lower Limit | Upper Limit |
| Exact | 20 | 1.8 | 0.205 | 1.952 | 0.031 | G.E. | 0.062 | 0.97 | 4.805 |

The one-sided p-value indicates the rejection of null hypothesis and establishes the superiority of drug B over A.

87.5 Example: Odds Ratio of Proportions

Dataset: Clntrt.cydx as described in Section 87.1.

Purpose of the Analysis:

Let π_t and π_c denote proportion of responses under treatment and control arm respectively. The odds ratio of proportions denoted by Ψ is defined as

$$\Psi = \frac{\pi_t (1 - \pi_c)}{\pi_c (1 - \pi_t)}$$

The null hypothesis $H_0 : \Psi = 1$ is to be tested against the 2-sided alternative hypothesis $H_1 : \Psi \neq 1$ or against 1-sided alternative hypotheses $H_1 : \Psi < 1$ or $H_1 : \Psi > 1$.

Analysis Steps

1. Open the dataset from **Samples** folder.
2. Choose the menu item
Analysis > (Discrete) Two Samples > (Parallel Design) Odds Ratio of Proportions
3. In the ensuing dialog box (under the **Main**) tab select **Superiority** as the **Trial**

Type. Choose other variables as shown below.

Analysis: Discrete Endpoint Two-Sample Test - Parallel Design - Odds Ratio of Proportions

Data Set: Clntrt.cyx

Main Advanced

Trial Type: Superiority Response Variable: Outcome Frequency Variable: Freq

Population Id: Drug Response Value: 1

Control: A

Treatment: B

4. In the **Advanced** tab, leave the fields **By Variable 1** and **By Variable 2** blank and keep default value of 0.95 in **Confidence Level**.

Analysis: Discrete Endpoint Two-Sample Test - Parallel Design - Odds Ratio of Proportions

Data Set: Clntrt.cyx

Main Advanced

By Variable 1: Confidence Level: 0.95

By Variable 2:

5. Click **OK** to start the analysis. Upon completion of the analysis, the output is

87 Analysis-Binomial Superiority Two-Sample

displayed in the main window.

Analysis: Binomial Response: Odds Ratio of Proportions

Hypothesis

$H_0 : \pi_t (1 - \pi_c) / \pi_c (1 - \pi_t) = 1$ Vs. $H_1 : \pi_t (1 - \pi_c) / \pi_c (1 - \pi_t) \neq 1$ for 2-sided test

Either $H_1 : \pi_t (1 - \pi_c) / \pi_c (1 - \pi_t) < 1$

Or $H_1 : \pi_t (1 - \pi_c) / \pi_c (1 - \pi_t) > 1$ for 1-sided test

Input Parameters

Data File: C:\ntrt.cyx
 Trial Type: Superiority
 Population Variable: Drug(Treatment=B, Control=A)
 Response Variable: Outcome(1)
 Frequency Variable: Freq
 Confidence Level: 0.95

Output

Response Variable: (Outcome)

Total no. of records: 4

No. of records rejected: 0

Table of observed frequencies:

| | A | B | Total |
|-------|----|----|-------|
| 1 | 5 | 9 | 14 |
| 0 | 5 | 1 | 6 |
| Total | 10 | 10 | 20 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| A | 10 | 5 | 0.5 |
| B | 10 | 9 | 0.9 |

Test of Hypothesis:

| n | Odds Ratio | M-H Estimate | 2-Sided p-value | | 95% Confidence Interval(2-Sided) | |
|----|------------|--------------|-----------------|--------------|----------------------------------|-------------|
| | | | RBG variance | M-H variance | Lower Limit | Upper Limit |
| 20 | 9 | 9 | 0.074 | 0.057 | 0.809 | 100.139 |

The output gives estimate of odds ratio and 2-sided p-value using RBG variance and M-H variance. The two sided p values indicate failing to reject the null hypothesis.

87.6 Example: Common Odds Ratio of Proportions for stratifies 2X2 tables

Dataset: BD.cydx

Data Description

The data below for six age groups, relating alcohol to oesophageal cancer, are taken from Breslow and Day (1980).

| Age Group | Alcohol Exposure | | No Exposure | |
|-----------|------------------|---------|-------------|---------|
| | Case | Control | Case | Control |
| 25-34 | 1 | 9 | 0 | 106 |
| 35-44 | 4 | 26 | 5 | 164 |
| 45-54 | 25 | 29 | 21 | 138 |
| 55-64 | 42 | 27 | 34 | 139 |
| 65-74 | 19 | 18 | 36 | 88 |
| 75+ | 5 | 0 | 8 | 31 |

Purpose of the Analysis:

The Homogeneity test is executed on these data to determine if the Odds-Ratios across the six age groups are constant.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item
Analysis > (Discrete) Two Samples > (Parallel Design) Common Odds Ratio for Stratifies 2X2 Tables
3. In the ensuing dialog box (under the **Main**) tab choose the variables as shown

87 Analysis-Binomial Superiority Two-Sample

below.

The screenshot shows the SAS software interface for a binomial superiority two-sample test. The title bar reads "Analysis: Discrete Endpoint Two-Sample Test - Parallel Design - Common Odds Ratio for Stratified". Below the title bar, the data set is identified as "BD.cydx". There are two tabs, "Main" and "Advanced", with "Advanced" selected. The "Row:" dropdown is set to "Row", the "Column:" dropdown is set to "Col", and the "Stratum:" dropdown is set to "Stratum". The "Frequency Variable:" field is set to "Freq".

Analysis: Discrete Endpoint Two-Sample Test - Parallel Design - Common Odds Ratio for Stratified

Data Set: BD.cydx

Main Advanced

Row: Row Column: Col Frequency Variable: Freq

Stratum: Stratum

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Common Odds Ratio For Stratified 2X2 table

Hypothesis

Define $\Psi_i = \pi_{1i} (1 - \pi_{ci}) / \pi_{ci} (1 - \pi_{1i})$ $H_0 : \Psi_i = \Psi$ for all $i = 1, \dots, s$ where s is the number of strata Vs. (i) $H_1 : \Psi_i \neq \Psi$ for 2-sid

Input Parameters

Data File: BD.cydx
 Row Variable: Row
 Column Variable: Col
 Stratum Variable: Stratum
 Frequency Variable: Freq
 Confidence Level: 0.95

Output

Total no. of records: 22
 No. of records rejected: 0
 Summary of the Observed Data:

| Strata | n | $\pi_{.0}$ | $\pi_{.1}$ | Odds Ratio |
|--------|-----|------------|------------|------------|
| 1 | 116 | 0.1 | 0 | 0 |
| 2 | 199 | 0.133 | 0.03 | 0.198 |
| 3 | 213 | 0.463 | 0.132 | 0.177 |
| 4 | 242 | 0.609 | 0.197 | 0.157 |
| 5 | 161 | 0.514 | 0.29 | 0.388 |
| 6 | 44 | 1 | 0.205 | 0 |

Breslow and Day Statistic: 9
 (with Tarone's correction): 9
 Test of Hypothesis:

| Type | | p-value | | 99% CI for exact P-Value | | | |
|------------|--|-----------|----|--------------------------|---------|-------------|-------------|
| | | Statistic | DF | Tail | 2-Sided | Lower Limit | Upper Limit |
| Asymptotic | Breslow and Day | 9.323 | 5 | G.E. | 0.097 | 0.085 | 0.086 |
| | Breslow and Day with Tarone's Correction | 9.299 | 5 | G.E. | 0.098 | | |

Warning Message: Since the asymptotic p value is not contained in the 99% CI, the Asymptotic

Estimation and Testing of Common Odds Ratio:

Mantel-Haenszel Inference:

| Type | 2-Sided P-Value | | | 95% CI with RBG Variance | |
|------------|-----------------|--------------|--------------|--------------------------|-------------|
| | Point Estimate | RBG variance | M-H variance | Lower Limit | Upper Limit |
| Asymptotic | 5.158 | 0 | 4.799E-10 | 3.562 | 7.468 |

The output gives observed odds ratios across strata, Breslow and Day statistic with and without Tarone's correction and 2-sided p-value using RBG variance and M-H variance. Note that the two sided p values for both Breslow and Day (1980) statistic and with Tarone's correction are greater than 0.05 thereby accepting the null

87 Analysis-Binomial Superiority Two-Sample

hypothesis of common odds ratio across the strata. However, you see a warning in the output.

East computes 95% confidence intervals for the exact p value and checks if the asymptotic p value lies in the interval. In case, it doesn't, **East** gives the warning message that the asymptotic inference would be unreliable.

Having accepted the hypothesis of common odds ratio across all strata, the Mantel-Haenszel inference tests the hypothesis that this common odds ratio is equal to 1. Both the p values using RBG variance and MH variance are very close to zero indicating rejection of the null hypothesis that the common odds ratio is equal to 1.

87.7 Example: Fisher's Exact Test

Dataset: `Clntrt.cydx` as described in Section 87.1.

Purpose of the Analysis:

As in the **Difference of Proportions** test, suppose π_t and π_c denote the proportions of the successes from the experimental treatment (T) and the control treatment (C). To test the null hypothesis:

$$H_0: \pi_t = \pi_c, \quad (87.1)$$

against 1-sided alternatives of the form,

$$H_1: \pi_t > \pi_c, \quad (87.2)$$

or

$$H'_1: \pi_t < \pi_c, \quad (87.3)$$

and against 2-sided alternatives of the form

$$H_2: \pi_t \neq \pi_c. \quad (87.4)$$

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item
Analysis > (Discrete) Two Samples > (Parallel Design) Fisher's Exact

3. In the ensuing dialog box (under the **Main**) tab choose the variables as shown below.

Analysis: Discrete Endpoint Two-Sample Test - Parallel Design

Data Set: Clntrt.cydx

Main Advanced

Row: Drug Column: Outcome Frequency Variable: Freq

Control: 0
Treatment: 1

87 Analysis-Binomial Superiority Two-Sample

- Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: Fisher's Exact Test

Hypothesis
 $H_0 : \pi_t = \pi_c$ Vs. $H_1 : \pi_t \neq \pi_c$ for 2-Sided test
 Either $H_1 : \pi_t > \pi_c$ Or $H_1 : \pi_t < \pi_c$ for 1-Sided test

Input Parameters
 Data File: C:\ntrt.cyx
 Column Variable: Outcome
 Row Variable: Drug
 Frequency Variable: Freq

Output
 Total no. of records: 4
 No. of records rejected: 0
 Observed Frequency Table:

| | 0 | 1 | Total |
|-------|---|----|-------|
| A | 5 | 5 | 10 |
| B | 1 | 9 | 10 |
| Total | 6 | 14 | 20 |

Summary of Test Statistic:
 Y: Observed value in Row 1 and column 1

| Y | Minimum | Maximum | Mean | Std.Deviation | Observed |
|---|---------|---------|------|---------------|----------|
| 5 | 0 | 6 | 3 | 1.051 | 3.58 |

Hypergeometric Prob. of Table: 0.065
 Test of Hypothesis:

| Type | Statistic | DF | (1-Sided) | | (2-Sided) |
|------------|-----------|----|-----------|---------|-----------|
| | | | Tail | p-value | p-value |
| Asymptotic | 3.58 | 1 | G.E. | 0.029 | 0.058 |
| Exact | 3.58 | | G.E. | | 0.141 |
| Exact | 5 | | G.E. | 0.07 | |

The above output provides Fisher statistic, and the 2-sided asymptotic p-value as the tail area to the right of the observed Fisher statistic from a chi-square distribution with 1 df as shown in the equation. It is 0.058. The asymptotic 1-sided p-value is defined to be half the corresponding 2-sided p-value, or 0.0292. The bottom portion of the screen provides exact 1 and 2-sided p-values. The exact 2-sided p-value, 0.141 with $D(Y)$ as the Fisher statistic. This is considerably larger than the asymptotic p-value,

highlighting the unreliability of asymptotic inference for small datasets. The output screen shows that the 1-sided p-value is obtained as the tail area to the left of 5 from the distribution of y_{11} . The magnitude of the p-value is 0.07.

The one sided exact p value can be obtained from the exact distribution of y_{11} , the entry in row 1 and column 1 of the 2×2 table.

88 *Analysis-Binomial Noninferiority Two-Sample*

In a binomial noninferiority trial the goal is to establish that the response rate of an experimental treatment is no worse than that of an active control, rather than attempting to establish that it is superior. A therapy that is demonstrated to be noninferior to the current standard therapy for a particular indication might be an acceptable alternative if, for instance, it is easier to administer, cheaper, or less toxic.

Such noninferiority trials are designed by specifying a noninferiority margin. The amount by which the response rate on the experimental arm is worse than the response rate on the control arm must fall within this margin in order for the claim of noninferiority to be sustained. Chapter 18 deals with the designing of such clinical trials considering difference of proportions, ratio of proportions or odds ratio of proportions of the two populations.

This chapter demonstrates how **East** is used to analyze the data from two independent binomial samples generated while conducting a noninferiority trial. We shall assume that the data is sampled independently from two binomial populations with response probabilities π_t and π_c for treatment and control. This comparison is based on difference of proportions, ratio of proportions or odds ratio of the two populations. For difference and ratio of proportions, we follow two formulations, namely Wald's (1940) and Farrington and Manning's (1990) score.

88.1 *Example: Noninferiority -Diff. of Proportions - Asymptotic*

Dataset: Nephrodash.cyd.

Data Description

The data is for childhood nephroblastoma. Details of the data are as given below:

| Response | Chemo (New) | Radio (Standard) | Total |
|-----------------|------------------------|-----------------------------|--------------|
| Rupture-free | 83 | 80 | 163 |
| Ruptured tumor | 5 | 7 | 12 |
| Total | 88 | 87 | 175 |

The dataset has three variables **Resp**, **PopID** and **Freq**. A value of 1 in **Resp** represents response and 0 as non-response. In **PopID**, 0 is control and 1 is treatment.

Purpose of the Analysis:

The standard treatment for this disease is nephrectomy followed by post-operative radiotherapy. Whereas the experimental treatment is pre-operative chemotherapy to reduce the tumor mass, followed by nephrectomy.

First perform superiority test to see if the experimental treatment is superior to the standard therapy. For this analysis, consider 1-sided type I error of 0.05.

This will be followed by a noninferiority test with a noninferiority margin of 0.1

Analysis Steps: For Superiority Test

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Discrete) Two Samples > (Parallel Design) Difference of Proportions
3. In the **Main** tab, select variables as shown below. Do not check **Perform Exact Computation** checkbox.

Analysis: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

Data Set: Nephrodash.cyx

Main Advanced

Trial Type: Superiority Response Variable: Resp Frequency Variable: Freq

Population Id: PopID Response Value: 1

Control: 0

Treatment: 1

Perform Exact Computations

88 Analysis-Binomial Noninferiority Two-Sample

- Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: Difference of Proportions

Hypothesis

$H_0: \pi_t - \pi_c = 0$ Vs. $H_1: \pi_t - \pi_c \neq 0$ for 2-Sided test

Either $H_1: \pi_t - \pi_c < 0$

Or $H_1: \pi_t - \pi_c > 0$ for 1-Sided test

Input Parameters

Data File: Nephrodash.cyx
 Trial Type: Superiority
 Population Variable: PopD(Treatment=1, Control=0)
 Response Variable: Resp(1)
 Frequency Variable: Freq
 Confidence Level: 0.95

Output

Response Variable: (Resp)

Total Number of Records: 4

Number of Records Rejected: 0

Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|----|-------|
| 1 | 80 | 83 | 163 |
| 0 | 7 | 5 | 12 |
| Total | 87 | 88 | 175 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| 0 | 87 | 80 | 0.92 |
| 1 | 88 | 83 | 0.943 |

Test of Hypothesis:

| Type | n | Difference of Proportions | Std. Error | Test Statistic | 1-Sided | | 2-Sided | | 95% Confidence Interval | |
|------------|-----|---------------------------|------------|----------------|---------|---------|---------|-------------|-------------------------|--|
| | | | | | Tail | p-value | p-value | Lower Limit | Upper Limit | |
| Asymptotic | 175 | 0.024 | 0.038 | 0.619 | G E | 0.268 | 0.536 | -0.057 | 0.108 | |

Note that the p-value for one-sided test is 0.268. Clearly there is no evidence that the chemotherapy arm is superior to radiotherapy. However, the goal of this study was different. The investigators only wished to establish the noninferiority of chemotherapy relative to radiotherapy at a noninferiority margin of 10%. In other words, the chemotherapy arm is considered to be non-inferior to the radiotherapy arm if the probability of being rupture free following the surgery is at most 10% lower for the chemotherapy arm than for the radiotherapy arm.

Analysis Steps: For Noninferiority Test

- Click **Analysis Inputs** tab on the status bar below. This will open recent inputs you gave for superiority in the main window. In the **Main** tab, change the trial

type to **Noninferiority**. Input the value of Noninferiority margin as 0.1. Click **Wald** in **Test Type**. Here also, do not check **Perform Exact Computation** checkbox.

Analysis: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

Data Set: Nephrodash.cydx

Main | **Advanced**

Trial Type: Response Variable: Frequency Variable:

Population Id: Response Value:

Control: Test Type: Wald Score
 Noninferiority Margin:

Treatment:

Perform Exact Computations

2. Click **OK** to display following output in the main window.

Analysis: Binomial Response: Difference of Proportions

Hypothesis
 If $\delta_0 > 0$ then $H_0 : \pi_t - \pi_c \geq \delta_0$ Vs. $H_1 : \pi_t - \pi_c < \delta_0$
 If $\delta_0 < 0$ then $H_0 : \pi_t - \pi_c \leq \delta_0$ Vs. $H_1 : \pi_t - \pi_c > \delta_0$

Input Parameters
 Data File: Nephrodash.cydx
 Trial Type: Noninferiority
 Test Type: Wald
 Population Variable: PopID(Treatment=1, Control=0)
 Response Variable: Resp(1)
 Frequency Variable: Freq
 Noninferiority Margin(δ_0): 0.1
 Confidence Level: 0.95

Output
 Response Variable: (Resp)
 Total Number of Records: 4
 Number of Records Rejected: 0

88 Analysis-Binomial Noninferiority Two-Sample

Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|----|-------|
| 1 | 80 | 83 | 163 |
| 0 | 7 | 5 | 12 |
| Total | 87 | 88 | 175 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| 0 | 87 | 80 | 0.92 |
| 1 | 88 | 83 | 0.943 |

Test of Hypothesis:

| Type | n | Difference of Proportions | Std. Error | Test Statistic | 1-Sided | | 95% Confidence Interval | |
|------------|-----|---------------------------|------------|----------------|---------|---------|-------------------------|-------------|
| | | | | | Tail | p-value | Lower Limit | Upper Limit |
| Asymptotic | 175 | 0.024 | 0.038 | -1.999 | L.E. | 0.023 | -1 | 0.086 |

Note the 1-sided p-value is now 0.023. This p-value is associated with the rejection of $H_0: \delta \leq 0$ in favor of the alternative hypothesis $H_1: \delta > 0$. East displays the p-value associated with right tailed test on this occasion because $\hat{\delta} > 0$. The 2-sided 95% confidence interval is (-1, 0.086). The p-value as well as the confidence interval indicate the rejection of null hypothesis and Noninferiority of chemotherapy over radiotherapy.

In the **Main** tab, if you select **Score** in Test Type, the following output is displayed:

Analysis: Binomial Response: Difference of Proportions

Hypothesis

If $\delta_0 > 0$ then $H_0: \pi_t - \pi_c \geq \delta_0$ Vs. $H_1: \pi_t - \pi_c < \delta_0$

If $\delta_0 < 0$ then $H_0: \pi_t - \pi_c \leq \delta_0$ Vs. $H_1: \pi_t - \pi_c > \delta_0$

Input Parameters

Data File: Nephrodash.cydx
 Trial Type: Noninferiority
 Test Type: Score
 Population Variable: PopID(Treatment=1, Control=0)
 Response Variable: Resp(1)
 Frequency Variable: Freq
 Noninferiority Margin(δ_0): 0.1
 Confidence Level: 0.95

Output

Response Variable: (Resp)
 Total Number of Records: 4
 Number of Records Rejected: 0

Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|----|-------|
| 1 | 80 | 83 | 163 |
| 0 | 7 | 5 | 12 |
| Total | 87 | 88 | 175 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| 0 | 87 | 80 | 0.92 |
| 1 | 88 | 83 | 0.943 |

Test of Hypothesis:

| Type | n | Difference of Proportions | Std. Error | Test Statistic | 1-Sided | | 95% Confidence Interval | |
|------------|-----|---------------------------|------------|----------------|---------|---------|-------------------------|-------------|
| | | | | | Tail | p-value | Lower Limit | Upper Limit |
| Asymptotic | 175 | 0.024 | 0.042 | -1.801 | L.E. | 0.036 | -Infinity | 0.093 |

In this case, the 1-sided p-value is 0.036 establishing Noninferiority.

88.2 Example: Diff. of Proportions - Exact

Dataset: Nephrodash.cyd as described in Section 88.1.

Purpose of the Analysis:

The standard treatment for this disease is nephrectomy followed by post-operative radiotherapy. Whereas the experimental treatment is pre-operative chemotherapy to reduce the tumor mass, followed by nephrectomy.

First perform superiority test to see if the experimental treatment is superior to the standard therapy. For this analysis, consider 1-sided type I error of 0.05.

This will be followed by a noninferiority test type with a noninferiority margin of 0.1

Analysis Steps: For Superiority Test

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Discrete) Two Samples > (Parallel Design) Difference of Proportions
3. In the **Main** tab, select variables as shown below. Check **Perform Exact**

88 Analysis-Binomial Noninferiority Two-Sample

Computation checkbox.

Analysis: Discrete Endpoint: Two-Sample Test - Parallel Design - Difference of Proportions

Data Set: Nephrodash.cydx

Main Advanced

Trial Type: Superiority Response Variable: Resp Frequency Variable: Freq

Population ID: PopID Response Value: 1

Control: 0 Treatment: 1

Perform Exact Computation

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: Difference of Proportions

Hypothesis

$H_0: \pi_t - \pi_c = 0$ Vs. $H_1: \pi_t - \pi_c \neq 0$ for 2-Sided test

Either $H_1: \pi_t - \pi_c < 0$

Or $H_1: \pi_t - \pi_c > 0$ for 1-Sided test

Input Parameters

Data File: Nephrodash.cydx
 Trial Type: Superiority
 Test Type: Wald
 Population Variable: PopID(Treatment=1, Control=0)
 Response Variable: Resp(1)
 Frequency Variable: Freq
 Confidence Level: 0.95

Output

Response Variable: (Resp)
 Total Number of Records: 4
 Number of Records Rejected: 0

Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|----|-------|
| 1 | 80 | 83 | 163 |
| 0 | 7 | 5 | 12 |
| Total | 87 | 88 | 175 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| 0 | 87 | 80 | 0.92 |
| 1 | 88 | 83 | 0.943 |

Test of Hypothesis:

| Type | n | Difference of Proportions | Std. Error | Test Statistic | 1-Sided | | 2-Sided | 95% Confidence Interval | |
|-------|-----|---------------------------|------------|----------------|---------|---------|---------|-------------------------|-------------|
| | | | | | Tail | p-value | p-value | Lower Limit | Upper Limit |
| Exact | 175 | 0.024 | 0.038 | 0.619 | G.E. | 0.289 | 0.579 | -0.057 | 0.108 |

Note that the p-value for one-sided test is 0.289. Clearly there is no evidence that the chemotherapy arm is superior to radiotherapy. However, the goal of this study was different. The investigators only wished to establish the noninferiority of chemotherapy relative to radiotherapy at a noninferiority margin of 10%. In other words, the chemotherapy arm is considered to be non-inferior to the radiotherapy arm if the probability of being rupture free following the surgery is at most 10% lower for the chemotherapy arm than for the radiotherapy arm.

88 Analysis-Binomial Noninferiority Two-Sample

Analysis Steps: For Noninferiority Test

1. Click **Analysis Inputs** tab on the status bar below. This will open recent inputs you gave for superiority in the main window. In the **Main** tab, change the trial type to **Noninferiority**. Input the value of Noninferiority margin as 0.1. Click **Score** in **Test Type**. Check **Perform Exact Computation** checkbox.
2. Click **OK** to display following output in the main window.

Analysis: Binomial Response: Difference of Proportions

Hypothesis

If $\delta_0 > 0$ then $H_0 : \pi_t - \pi_c \geq \delta_0$ Vs. $H_1 : \pi_t - \pi_c < \delta_0$

If $\delta_0 < 0$ then $H_0 : \pi_t - \pi_c \leq \delta_0$ Vs. $H_1 : \pi_t - \pi_c > \delta_0$

Input Parameters

Data File: Nephrodash.cydx
 Trial Type: Noninferiority
 Test Type: Score
 Population Variable: PopID(Treatment=1, Control=0)
 Response Variable: Resp(1)
 Frequency Variable: Freq
 Noninferiority Margin(δ_0): 0.1
 Confidence Level: 0.95

Output

Response Variable: (Resp)
 Total Number of Records: 4
 Number of Records Rejected: 0

Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|----|-------|
| 1 | 80 | 83 | 163 |
| 0 | 7 | 5 | 12 |
| Total | 87 | 88 | 175 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| 0 | 87 | 80 | 0.92 |
| 1 | 88 | 83 | 0.943 |

Test of Hypothesis:

| Type | n | Difference of Proportions | Std. Error | Test Statistic | 1-Sided | | 95% Confidence Interval | |
|-------|-----|---------------------------|------------|----------------|---------|---------|-------------------------|-------------|
| | | | | | Tail | p-value | Lower Limit | Upper Limit |
| Exact | 175 | 0.024 | 0.042 | -1.801 | L.E. | 0.037 | -Infinity | 0.093 |

In exact computations, the p-value is 0.037 indicating the significance. This concludes that Chemotherapy is noninferior to Radiotherapy.

88.3 Ratio of Proportions

*Example: Ratio of
Proportions -
Asymptotic*

*Example: Ratio of
Proportions - Exact*

As before, let π_t and π_c denote the proportions of the successes from the experimental treatment (T) and the control treatment (C), respectively. To test the null hypothesis, we transform the original data using log and perform difference of proportions test.

*Example: Ratio of Proportions - Asymptotic
Dataset: Vaccine.cydx.*

Data Description

Chan (1998) discusses a vaccine efficacy study of a recombinant DNA Influenza A vaccine against wild-type H1N1 virus challenge. The study compares the infection rates in the vaccinated and placebo groups. There were 15 individuals in each group. The following data was obtained.

| Disease Status | Treatment Group | | Total |
|----------------|-----------------|---------|-------|
| | Placebo | Vaccine | |
| Infected | 12 (80%) | 7 (47%) | 19 |
| Not Infected | 3 (20%) | 8 (53%) | 11 |
| Total | 15 | 15 | 30 |

Purpose of the Analysis:

Let π_t be the infection rate in the vaccinated group and π_c be the infection rate in the placebo group. Define $\rho = \pi_t/\pi_c$, and define $\lambda = 1 - \rho$. The parameter λ is known as the vaccine efficacy. Assume that $\pi_t \leq \pi_c$. Therefore the new vaccine has 100% efficacy if $\pi_t = 0$ and no efficacy if $\pi_t = \pi_c$. From a public health standpoint, the benefits from vaccination must exceed a given threshold in order to justify the risk of vaccinating healthy subjects. Therefore, in designing vaccine trials, one typically chooses a non-zero efficacy lower bound. Suppose we choose $\lambda_0 = 0.1$ as the non-zero efficacy lower bound. This implies that if $\lambda \leq 0.1$, the virus does not offer sufficient benefit relative to placebo to justify using it on a large scale for the prevention of infection. Thus we wish to test the null hypothesis of insufficient vaccine efficacy (i.e., inferiority) $\lambda \leq 0.1$ against the 1-sided alternative hypothesis of sufficient vaccine efficacy (i.e., noninferiority), $\lambda > 0.1$.

Equivalently, we wish to test the null hypothesis of inferiority,

$$H_0 : \rho \geq 0.9, \quad (88.1)$$

88 Analysis-Binomial Noninferiority Two-Sample

against the alternative hypothesis of noninferiority.

$$H_1 : \rho < 0.9. \tag{88.2}$$

Analysis Steps

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Discrete) Two Samples > (Parallel Design) Ratio of Proportions
3. In the Main tab, select **Noninferiority** as **Trial Type**. Select all other variables as shown below. Do not check **Perform Exact Computation** checkbox.

Analysis: Discrete Endpoint: Two-Sample Test - Parallel Design - Ratio of Proportions

Data Set: Vaccine.cydx

Main Advanced

Trial Type: Noninferiority Response Variable: Resp Frequency Variable: freq

Population Id: Popn Response Value: 1

Control: 0

Treatment: 1

Test Type

Noninferiority Margin: 0.9

Wald Score(Farrington Manning)

Perform Exact Computation

4. In the **Advanced** tab, leave the **By Variable 1** and **By Variable 2** blank and keep default value of 0.95 in **Confidence level**.

5. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: Ratio of Proportions

Hypothesis

If $\rho_0 > 1$ then $H_0: \pi_t / \pi_c \geq \rho_0$ Vs. $H_1: \pi_t / \pi_c < \rho_0$

If $\rho_0 < 1$ then $H_0: \pi_t / \pi_c \leq \rho_0$ Vs. $H_1: \pi_t / \pi_c > \rho_0$

Input Parameters

Data File: Vaccine.cyx
 Trial Type: Noninferiority
 Test Type: Wald
 Population Variable: Popn(Treatment=1, Control=0)
 Response Variable: Resp(1)
 Frequency Variable: freq
 Noninferiority Margin (ρ_0): 0.9
 Confidence Level: 0.95

Output

Response Variable: (Resp)

Total Number of Records: 4

Number of Records Rejected: 0

Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|----|-------|
| 1 | 12 | 7 | 19 |
| 0 | 3 | 8 | 11 |
| Total | 15 | 15 | 30 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| 0 | 15 | 12 | 0.8 |
| 1 | 15 | 7 | 0.467 |

Test of Hypothesis:

| Type | n | Ratio of Proportions | Std. Error | Test Statistic | (1-Sided) | | 95% Confidence Interval | |
|------------|----|----------------------|------------|----------------|-----------|------|-------------------------|-------------|
| | | | | | p-value | Tail | Lower Limit | Upper Limit |
| Asymptotic | 30 | 0.583 | 0.305 | -1.423 | 0.923 | G.E. | 0.353 | Infinity |

The observed value of test statistic is -1.423 with a 1-sided p-value equal to 0.923. The 1-sided 95% confidence interval for π_t / π_c is (0.353, *Infinity*). The p-value indicates that the null hypothesis of insufficient vaccine efficacy cannot be rejected. The corresponding 95% lower confidence bound for ρ is 0.353, which confirms that we cannot rule out the possibility that $\rho \leq 0.9$.

88 Analysis-Binomial Noninferiority Two-Sample

If you select **Score (Farrington Manning)** in Test Type, the following output is displayed:

Analysis: Binomial Response: Ratio of Proportions

Hypothesis

If $\rho_0 > 1$ then $H_0: \pi_t / \pi_c \geq \rho_0$ Vs. $H_1: \pi_t / \pi_c < \rho_0$

If $\rho_0 < 1$ then $H_0: \pi_t / \pi_c \leq \rho_0$ Vs. $H_1: \pi_t / \pi_c > \rho_0$

Input Parameters

Data File: Vaccine.cydx
 Trial Type: Noninferiority
 Test Type: Score(Farrington Manning)
 Population Variable: Popn(Treatment=1, Control=0)
 Response Variable: Resp(1)
 Frequency Variable: freq
 Noninferiority Margin (ρ_0): 0.9
 Confidence Level: 0.95

Output

Response Variable: (Resp)

Total Number of Records: 4

Number of Records Rejected: 0

Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|----|-------|
| 1 | 12 | 7 | 19 |
| 0 | 3 | 8 | 11 |
| Total | 15 | 15 | 30 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| 0 | 15 | 12 | 0.8 |
| 1 | 15 | 7 | 0.467 |

Test of Hypothesis:

| Type | n | Ratio of Proportions | Std. Error | Test Statistic | (1-Sided) | | 95% Confidence Interval | |
|------------|----|----------------------|------------|----------------|-----------|------|-------------------------|-------------|
| | | | | | p-value | Tail | Lower Limit | Upper Limit |
| Asymptotic | 30 | 0.583 | 0.159 | -1.525 | 0.936 | G.E. | 0.257 | Infinity |

Since the p-value is 0.936, the noninferiority can not be established.

Example: Ratio of Proportions - Exact

Dataset: Vaccine.cydx as described in Section 88.3.

Purpose of the Analysis:

To test the null hypothesis of inferiority,

$$H_0 : \rho \geq 0.9, \tag{88.3}$$

against the alternative hypothesis of noninferiority.

$$H_1 : \rho < 0.9. \tag{88.4}$$

Analysis Steps

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Discrete) Two Samples > (Parallel Design) Ratio of Proportions
3. In the Main tab, select **Noninferiority** as **Trial Type**. Also make sure to check **Perform Exact Computation** checkbox.

Analysis: Discrete Endpoint: Two-Sample Test - Parallel Design - Ratio of Proportions .00 .00

Data Set: Vaccine.cyx

Main **Advanced**

| | | | | | |
|----------------|----------------|---|------|---------------------|------|
| Trial Type: | Noninferiority | Response Variable: | Resp | Frequency Variable: | freq |
| Population Id: | Popn | Response Value: | 1 | | |
| Control: | 0 | Test Type | | | |
| Treatment: | 1 | Noninferiority Margin: 0.9 | | | |
| | | <input type="radio"/> Wald <input checked="" type="radio"/> Score(Farrington Manning) | | | |

Perform Exact Computation

4. In the **Advanced** tab, leave the **By Variable 1** and **By Variable 2** blank and keep default value of 0.95 in **Confidence level**.

88 Analysis-Binomial Noninferiority Two-Sample

5. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: Ratio of Proportions

Hypothesis

If $\rho_0 > 1$ then $H_0 : \pi_t / \pi_c \geq \rho_0$ Vs. $H_1 : \pi_t / \pi_c < \rho_0$

If $\rho_0 < 1$ then $H_0 : \pi_t / \pi_c \leq \rho_0$ Vs. $H_1 : \pi_t / \pi_c > \rho_0$

Input Parameters

Data File: Vaccine.cyx
 Trial Type: Noninferiority
 Test Type: Score(Farrington Manning)
 Population Variable: Popn(Treatment=1, Control=0)
 Response Variable: Resp(1)
 Frequency Variable: freq
 Noninferiority Margin (ρ_0): 0.9
 Confidence Level: 0.95

Output

Response Variable: (Resp)

Total Number of Records: 4

Number of Records Rejected: 0

Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|----|-------|
| 1 | 12 | 7 | 19 |
| 0 | 3 | 8 | 11 |
| Total | 15 | 15 | 30 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| 0 | 15 | 12 | 0.8 |
| 1 | 15 | 7 | 0.467 |

Test of Hypothesis:

| Type | n | Ratio of Proportions | Std. Error | Test Statistic | (1-Sided) | | 95% Confidence Interval | |
|-------|----|----------------------|------------|----------------|-----------|------|-------------------------|-------------|
| | | | | | p-value | Tail | Lower Limit | Upper Limit |
| Exact | 30 | 0.583 | 0.166 | -1.525 | 0.086 | L.E. | -Infinity | 0.967 |

The p-value is 0.086 suggesting non-significance, however the value is drastically reduced from the corresponding asymptotic p-value.

88.4 Example: Odds Ratio of Proportion

Dataset: Vaccine.cydx as described in Section 88.3.

Purpose of the Analysis:

Use the same data to demonstrate the testing of Noninferiority of Odds Ratio in case of two independent binomial samples. **Analysis Steps**

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Discrete) Two Samples > (Parallel Design) Odds Ratio of Proportions
3. In the Main tab, select **Noninferiority** as **Trial Type**. Select all other variables as shown below.

Analysis: Discrete Endpoint Two-Sample Test - Parallel Design - Odds Ratio of Proportions

Data Set: Vaccine.cydx

Main | Advanced

Trial Type: Noninferiority | Response Variable: Resp | Frequency Variable: freq

Population Id: Popn | Response Value: 1

Control: 0 | Treatment: 1

Test Type

Noninferiority Margin: 1.3

Wald Score

4. In the **Advanced** tab, leave the **By Variable 1** and **By Variable 2** blank and keep default value of 0.95 in **Confidence level**.

88 Analysis-Binomial Noninferiority Two-Sample

5. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: Odds Ratio of Proportions

Hypothesis

If $\Psi_0 > 0$ $H_0 : \pi_t(1 - \pi_c) / \pi_c(1 - \pi_t) \geq \Psi_0$ Vs. $H_1 : \pi_t(1 - \pi_c) / \pi_c(1 - \pi_t) < \Psi_0$
 If $\Psi_0 < 0$ $H_0 : \pi_t(1 - \pi_c) / \pi_c(1 - \pi_t) \leq \Psi_0$ Vs. $H_1 : \pi_t(1 - \pi_c) / \pi_c(1 - \pi_t) > \Psi_0$

Input Parameters

Data File: Vaccine.cydx
 Trial Type: Noninferiority
 Test Type: Wald
 Population Variable: Popn(Treatment=1, Control=0)
 Response Variable: Resp(1)
 Frequency Variable: freq
 Noninferiority Margin (Ψ_0): 1.3
 Confidence Level: 0.95

Output

Response Variable: (Resp)
 Total Number of Records: 4
 Number of Records Rejected: 0

Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|----|-------|
| 1 | 12 | 7 | 19 |
| 0 | 3 | 8 | 11 |
| Total | 15 | 15 | 30 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| 0 | 15 | 12 | 0.8 |
| 1 | 15 | 7 | 0.467 |

Test of Hypothesis:

| n | Odds Ratio | Std. Error | Test Statistic | (1-Sided) | | 95% Confidence Interval(1-Sided) | |
|----|------------|------------|----------------|-----------|------|----------------------------------|-------------|
| | | | | p-value | Tail | Lower Limit | Upper Limit |
| 30 | 0.219 | 0.827 | -2.154 | 0.016 | L.E. | 0 | 0.853 |

The output gives Test Statistic value as -2.154 with a 1-sided p-value equal to 0.016. As a result, the vaccination can be considered noninferior to the control.

If you select **Score** as Test Type, the following output is displayed:

Analysis: Binomial Response: Odds Ratio of Proportions

Hypothesis

If $\Psi_0 > 0$ $H_0 : \pi_t(1 - \pi_c) / \pi_c(1 - \pi_t) \geq \Psi_0$ Vs. $H_1 : \pi_t(1 - \pi_c) / \pi_c(1 - \pi_t) < \Psi_0$
 If $\Psi_0 < 0$ $H_0 : \pi_t(1 - \pi_c) / \pi_c(1 - \pi_t) \leq \Psi_0$ Vs. $H_1 : \pi_t(1 - \pi_c) / \pi_c(1 - \pi_t) > \Psi_0$

Input Parameters

Data File: Vaccine.cydx
 Trial Type: Noninferiority
 Test Type: Score
 Population Variable: Popn(Treatment=1, Control=0)
 Response Variable: Resp(1)
 Frequency Variable: freq
 Noninferiority Margin (Ψ_0): 1.3
 Confidence Level: 0.95

Output

Response Variable: (Resp)
 Total Number of Records: 4
 Number of Records Rejected: 0

Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|----|-------|
| 1 | 12 | 7 | 19 |
| 0 | 3 | 8 | 11 |
| Total | 15 | 15 | 30 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| 0 | 15 | 12 | 0.8 |
| 1 | 15 | 7 | 0.467 |

Test of Hypothesis:

| n | Odds Ratio | Std. Error | Test Statistic | (1-Sided) | | 95% Confidence Interval(1-Sided) | |
|----|------------|------------|----------------|-----------|------|----------------------------------|-------------|
| | | | | p-value | Tail | Lower Limit | Upper Limit |
| 30 | 0.219 | 1.229 | 1.527 | 0.937 | L.E. | 0 | 16.387 |

89 Analysis-Binomial Equivalence Two-Samples

89.1 Equivalence: Difference of Proportions

This test arises when the difference is in establishing the bioequivalence of a new compound with an established compound. It is the 2-sided version of the noninferiority test for difference of proportions. Thus if π_c and π_t are the response rates of control and treatment, respectively, then the goal is to test the null hypothesis of inequivalence, $|\pi_t - \pi_c| \geq \delta_0$, against 2-sided alternative hypothesis of equivalence, $|\pi_t - \pi_c| < \delta_0$, for a pre-specified equivalence margin $\delta_0 > 0$.

We test the above null hypothesis by performing two separate one-sided non-inferiority hypothesis tests of the form

$$H_{01}: \pi_c - \pi_t \geq \delta_0 \text{ versus } H_{11}: \pi_c - \pi_t < \delta_0 \quad (89.1)$$

and

$$H_{02}: \pi_t - \pi_c \geq \delta_0 \text{ versus } H_{12}: \pi_t - \pi_c < \delta_0 . \quad (89.2)$$

Each hypothesis test is carried out separately. Hypothesis test H_{01} is performed under the assumption that $\pi_c - \pi_t$ is at its threshold null value $\pi_c - \pi_t = \delta_0$. Similarly hypothesis test H_{02} is tested under the assumption that $\pi_t - \pi_c$ is at its threshold null value $\pi_t - \pi_c = \delta_0$. We reject the null hypothesis of inequivalence and accept the alternative hypothesis of equivalence only if **both** H_{01} and H_{02} are rejected.

89.2 Example: Equivalence: Dataset: Nephrodash.cyd as described in Section 88.1. Difference of Proportions- Asymptotic

Analysis Steps

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Discrete) Two Samples > (Parallel Design) Difference of Proportions
3. In the **Main** tab, select variables as shown below. Do not check **Exact**

Computation checkbox.

Analysis: Discrete Endpoint Two-Sample Test - Parallel Design - Difference of Proportions

Data Set: Nephrodash.cydx

Main Advanced

Trial Type: Equivalence Response Variable: Resp Frequency Variable: Freq
 Population Id: PopID Response Value: 1
 Control: 0 Test Type
 Treatment: 1 Equivalence Margin: 0.1
 Perform Exact Computations

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: Difference of Proportions

$$H_{01} : \pi_c - \pi_t \geq \delta_0 \text{ Vs. } H_{11} : \pi_c - \pi_t < \delta_0$$

$$H_{02} : \pi_t - \pi_c \geq \delta_0 \text{ Vs. } H_{12} : \pi_t - \pi_c < \delta_0$$

Input Parameters

Data File: Nephrodash.cydx
 Trial Type: Equivalence
 Population Variable: PopID(Treatment=1, Control=0)
 Response Variable: Resp(1)
 Frequency Variable: Freq
 Equivalence Margin: 0.1
 Confidence Level: 0.95

89 Analysis-Binomial Equivalence Two-Samples

Output

Response Variable: (Resp)
 Total Number of Records: 4
 Number of Records Rejected: 0
 Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|----|-------|
| 1 | 80 | 83 | 163 |
| 0 | 7 | 5 | 12 |
| Total | 87 | 88 | 175 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| 0 | 87 | 80 | 0.92 |
| 1 | 88 | 83 | 0.943 |

Test of Hypothesis:

| Type | n | Difference of Proportions | Std. Error | | Test Statistic | | 1-Sided p-value | | | | 95% Confidence Inter | |
|------------|-----|---------------------------|------------|-------|----------------|--------|-----------------|------------|------|------------|----------------------|----------|
| | | | SE 01 | SE 02 | t 01 | t 02 | Tail | Under H 01 | Tail | Under H 02 | Lower Limit | Upper Li |
| Asymptotic | 175 | 0.024 | 0.044 | 0.042 | -2.801 | -1.801 | L.E. | 0.003 | L.E. | 0.036 | -0.057 | 0.10 |

The output gives Test Statistic values as -2.801 and -1.801 with 1-sided p-values equal to 0.003 and 0.036 , respectively.

The null hypothesis of inequivalence can be rejected only if both the noninferiority null hypotheses are rejected. Each noninferiority hypothesis is typically tested at the 2.5% level of significance since each test is 1-sided. In the present example a statistically significant p-value ($p = 0.003$) is obtained for the H_{01} non-inferiority tests and a non-significant p-value ($p = 0.036$) is obtained for the H_{02} non-inferiority test. Therefore we cannot reject the null hypothesis of inequivalence.

89.3 Example: Equivalence: Difference of Proportions-Exact

Dataset: **Nephrodash.cyd** as described in Section 88.1.

Analysis Steps

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Discrete) Two Samples > (Parallel Design) Difference of Proportions
3. In the **Main** tab, select variables as shown below. Make sure to check **Exact**

Computation checkbox.

Analysis: Discrete Endpoint Two-Sample Test - Parallel Design - Difference of Proportions

Data Set: Nephrodash.cydx

Main Advanced

| | | |
|---|-------------------------|--------------------------|
| Trial Type: Equivalence | Response Variable: Resp | Frequency Variable: Freq |
| Population Id: PopID | Response Value: 1 | |
| Control: 0 | Test Type | |
| Treatment: 1 | Equivalence Margin: 0.1 | |
| <input checked="" type="checkbox"/> Perform Exact Computation | | |

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: Difference of Proportions

$$H_{01} : \pi_c - \pi_t \geq \delta_0 \text{ Vs. } H_{11} : \pi_c - \pi_t < \delta_0$$

$$H_{02} : \pi_t - \pi_c \geq \delta_0 \text{ Vs. } H_{12} : \pi_t - \pi_c < \delta_0$$

Input Parameters

Data File: Nephrodash.cydx
Trial Type: Equivalence
Population Variable: PopID(Treatment=1, Control=0)
Response Variable: Resp(1)
Frequency Variable: Freq
Equivalence Margin: 0.1
Confidence Level: 0.95

89 Analysis-Binomial Equivalence Two-Samples

Output

Response Variable: (Resp)
 Total Number of Records: 4
 Number of Records Rejected: 0
 Table of Observed Frequencies:

| | 0 | 1 | Total |
|-------|----|----|-------|
| 1 | 80 | 83 | 163 |
| 0 | 7 | 5 | 12 |
| Total | 87 | 88 | 175 |

Summary of the Observed Data:

| Treatment Group | n | Responses | Proportion |
|-----------------|----|-----------|------------|
| 0 | 87 | 80 | 0.92 |
| 1 | 88 | 83 | 0.943 |

Test of Hypothesis:

| Type | n | Difference of Proportions | Std. Error | | Test Statistic | | 1-Sided p-value | | | | 95% Confidence Interval | |
|-------|-----|---------------------------|------------|-------|----------------|--------|-----------------|------------|------|------------|-------------------------|-------------|
| | | | SE_01 | SE_02 | t_01 | t_02 | Tail | Under H_01 | Tail | Under H_02 | Lower Limit | Upper Limit |
| Exact | 175 | 0.024 | 0.044 | 0.042 | -2.801 | -1.801 | L.E. | 0.002 | L.E. | 0.037 | -0.057 | 0.108 |

In this example, a statistically significant p-value ($p = 0.002$) is obtained for the H01 non-inferiority tests and a non-significant p-value ($p = 0.037$) is obtained for the H02 non-inferiority test. Therefore we can not reject the null hypothesis of inequivalence.

90

Analysis-Discrete: Many Proportions

In clinical trials involving categorical endpoints, there are several situations where either the data are coming from many binomial populations or the responses are from multinomial distribution. In case of multiple binomial populations, the interest lies in testing whether the success probability differs across several binomial populations, in particular does it increase or decrease with reference to an index variable. For data coming from multinomial distributions, one is interested in testing if the cell probabilities are according to some theoretical law. **East** can be used to analyze both these types of data. In this chapter we will demonstrate how the tests on many proportions can be executed in **East**.

90.1 Example: Chi-square Test of Specified Proportions

Dataset: Smallt.cyx

Data Description

The dataset has four variables **Category**, **Freq**, **Prob** and **ExpFreq**. The **Category** variable has four categories. **Freq** is the observed frequency for these four categories, and the variable **prob** represents expected probabilities for these categories. Table 90.1 shows the observed counts and the multinomial probabilities under the null hypothesis for a multinomial distribution with four categories.

Table 90.1: Frequency Counts from a Multinomial with 4 Categories

| | Multinomial Categories | | | | Row |
|--------------------|------------------------|-----|-----|-----|-------|
| | 1 | 2 | 3 | 4 | Total |
| Cell Counts | 7 | 1 | 1 | 1 | 10 |
| Cell Probabilities | 0.3 | 0.3 | 0.3 | 0.1 | 1 |

Purpose of the Analysis:

To test whether the observed cell counts are according to the specified Cell probabilities.

Analysis Steps: Based on expected probabilities

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Discrete) Many Samples > (Single Arm Design) Chi-Square for

90 Analysis-Discrete: Many Proportions

Specified Proportions in C categories

This will display several input fields associated with the Chi-square Test in the main window.

3. In the **Main** tab, select **Category** in **Category** and **Freq** in the **Observed Frequency** variable. Since the data consist of expected probabilities, select the **Probability** option and select variable **prob** in **Probability**.

Analysis: Discrete Endpoint Many-Sample Test - Single Arm Design - Chi-Square for Specified Proportions in C Categories

Data Set: Smallt.cydx

Main Advanced

Category: Category

Observed Frequency: Freq

Expected Count Probability

Probability: prob

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Categorical Response: Chi Square Test for Specified Proportions in C Categories

Hypothesis
 $H_0 : (O_1, O_2, \dots, O_C) \sim MN(\pi_1, \pi_2, \dots, \pi_C, N)$
 Vs
 $H_1 : H_0$ is not true for 2-sided test

Input Parameters
 Data File: Smallt.cyx
 Distinct Category: Category
 Observed Frequency: Freq
 Probability: prob
 Confidence Level: 0.95

Output
 Total Number of Records: 4
 Number of Records Rejected: 0
 Table of Observed Frequencies:

| | 1 | 2 | 3 | 4 | Total |
|-------|---|---|---|---|-------|
| 1 | 7 | 1 | 1 | 1 | 10 |
| Total | 7 | 1 | 1 | 1 | 10 |

Estimation of Multinomial Probability:

| Category | Count | Prob (P) | 95% Confidence Interval | |
|----------|-------|----------|-------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| 1 | 7 | 0.7 | 0.329 | 0.917 |
| 2 | 1 | 0.1 | 0.012 | 0.495 |
| 3 | 1 | 0.1 | 0.012 | 0.495 |
| 4 | 1 | 0.1 | 0.012 | 0.495 |

Test of hypothesis for Goodness of Fit:

| Type | p-value | | | |
|------------|-----------|----|------|---------|
| | Statistic | DF | Tail | 2-Sided |
| Asymptotic | 8 | 3 | G.E. | 0.046 |

Note that the output contains estimation of multinomial probabilities as well as the confidence intervals for these based on the observed data. The observed value of chi-square test statistic with degrees of freedom 3 is 8. The 2-sided p-value is 0.046. This p-value is associated with the rejection of $H_0 : \pi_i = \pi_{0i}, i = 1, 2, 3, \dots, C$ in favor of the alternative hypothesis of not following the multinomial distribution with specified proportions.

Analysis Steps: Based on expected frequencies

The test can also be run if the data contains expected frequencies rather than expected probabilities for the categories.

1. Click **Analysis Input/output** tab on the **Status bar** below.
2. In the main tab, select the **Expected Count** option instead of **Probability** and choose **ExpFreq** in **Expected Frequency**.

90 Analysis-Discrete: Many Proportions

3. Click **OK** and the following output is displayed.

Analysis: Categorical Response: Chi Square Test for Specified Proportions in C Categories

Hypothesis
 $H_0: (O_1, O_2, \dots, O_C) \sim MN(\pi_1, \pi_2, \dots, \pi_C, N)$
Vs
 $H_1: H_0$ is not true for 2-sided test

Input Parameters
Data File: Smallt.cydx
Distinct Category: Category
Observed Frequency: Freq
Expected Frequency: Expfreq
Confidence Level: 0.95

Output
Total Number of Records: 4
Number of Records Rejected: 0
Table of Observed Frequencies:

| | 1 | 2 | 3 | 4 | Total |
|-------|---|---|---|---|-------|
| 1 | 7 | 1 | 1 | 1 | 10 |
| Total | 7 | 1 | 1 | 1 | 10 |

Estimation of Multinomial Probability:

| Category | Count | Prob (Pi) | 95% Confidence Interval | |
|----------|-------|-----------|-------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| 1 | 7 | 0.7 | 0.329 | 0.917 |
| 2 | 1 | 0.1 | 0.012 | 0.495 |
| 3 | 1 | 0.1 | 0.012 | 0.495 |
| 4 | 1 | 0.1 | 0.012 | 0.495 |

Test of hypothesis for Goodness of Fit:

| Type | p-value | | | |
|------------|-----------|----|------|---------|
| | Statistic | DF | Tail | 2-Sided |
| Asymptotic | 8 | 3 | G.E. | 0.046 |

As before, the output shows estimates of multinomial probabilities and the asymptotic inference. Since the expected counts in the data were consistent with the probabilities, the inference is the same.

90.2 Example: Two group Chi-square test

Dataset: vari.cydx as described in Section 74.4.1.

Purpose of the Analysis:

To test if the two groups specified in row and column are independent of each other.

Analysis Steps

1. Open the dataset from **Samples** folder.

2. Choose the menu item:

Analysis > (Discrete) Many Samples > (Single Arm Design) Two Group Chi-Square for Proportions in C categories

This will display several input fields associated with the Chi-square Test in the main window.

3. In the **Main** tab, select **Group** in **Row (Group)**, **Category** in **Column(Categories)**, and **Freq** in the **Frequency Variable**.

Analysis: Discrete Endpoint Many-Sample Test - Parallel Design - Chi-Square for Proportions in C Categories

Data Set: Vari.cyx

Main Advanced

Row (Group): Column (Categories): Frequency Variable:

90 Analysis-Discrete: Many Proportions

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Categorical Response: Chi-square Test for Comparing Proportions in C Categories

Hypothesis
 $H_0: \pi_{1j} = \pi_{2j}$ for all $j = 1, 2, 3, \dots, C$
Vs
 H_1 : at least one π_{ij} differs for $i = 1, 2$ and $j = 1, 2, 3, \dots, C$

Input Parameters

Data File: Vari.cyx
Row Variable: Group
Column Variable: Category
Frequency Variable: Freq
Confidence Level: 0.95

Output

Total Number of Records: 6
Number of Records Rejected: 0

Table of Observed Frequencies:

| | 1 | 2 | 3 | 4 | Total |
|-------|----|---|---|---|-------|
| 1 | 21 | 0 | 2 | 0 | 23 |
| 2 | 15 | 3 | 1 | 2 | 21 |
| Total | 36 | 3 | 3 | 2 | 44 |

Pearson Chi-Square Statistic: 6.255
Chi-Square Statistic with Yates CC: NA
Likelihood Ratio Statistic: 8.185

Test of hypothesis for Chi-Square Test

| Type | p-value | | | | 99% CI for exact p-value | |
|------------|-----------|----|------|---------|--------------------------|-------------|
| | Statistic | DF | Tail | 2-Sided | Lower Limit | Upper Limit |
| Asymptotic | 6.255 | 3 | G.E. | 0.1 | 0.061 | 0.074 |

Warning Message: Since the asymptotic p value is not contained in the 99% CI, the Asymptotic Outcomes may be considered unreliable.

Test of hypothesis for Likelihood Ratio Test:

| Type | p-value | | | | 99% CI for exact p-value | |
|------------|-----------|----|------|---------|--------------------------|-------------|
| | Statistic | DF | Tail | 2-Sided | Lower Limit | Upper Limit |
| Asymptotic | 8.185 | 3 | G.E. | 0.042 | 0.073 | 0.087 |

Warning Message: Since the asymptotic p value is not contained in the 99% CI, the Asymptotic Outcomes may be considered unreliable.

Measure of Association:

| Coefficient | Estimate | ASE_MLE | 95% Confidence Interval | |
|------------------|----------|---------|-------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| Phi | 0.377 | 0.095 | 0.249 | 0.505 |
| Pearson's CC | 0.353 | 0.071 | 0.214 | 0.491 |
| Sakoda's CC | 0.499 | 0.1 | 0.303 | 0.695 |
| Tschuprow's CC | 0.286 | 0.038 | 0.212 | 0.361 |
| Cramer's V | 0.377 | 0.086 | 0.208 | 0.546 |
| Uncertainty(C R) | 0.139 | 0.044 | 0.052 | 0.225 |

Note that the output contains inference using chi-square test, likelihood ratio test and several measures of association such as Phi, Pearson's contingency coefficient, Sakoda's as well as Tshuprov coefficient, Cramer's V, and Uncertainty coefficient etc. It also displays a warning in case the asymptotic p-value does not belong to the 99% CI for exact p-value. The observed value of chi-square test statistic with degrees of freedom 3 is 6.255. The 2-sided p-value is 0.1. Accordingly, there is not enough evidence for rejecting the null hypothesis. Therefore, we cannot conclude that Interferon is more effective than placebo in preventing adverse effects.

90.3 Example: Wilcoxon Rank Sum Test for Ordered Categories Data

The Wilcoxon-Rank-Sum test (Lehmann, 1975) is one of the most popular nonparametric tests for detecting a shift in location between two populations. It can accommodate either continuous or ordinal categorical data. It has an asymptotic relative efficiency of 95.5%, relative to the *t* test when the underlying distributions are normal. The Wilcoxon rank sum test is used for comparing two populations that generate either continuous or ordinal categorical responses. The Wilcoxon rank sum statistic is defined by equation R.234.

Dataset: `vari.cydx` as described in Section 74.4.1.

Purpose of the Analysis:

To test that two populations, each generating an ordered categorical response, have the same underlying multinomial distribution for the response variable.

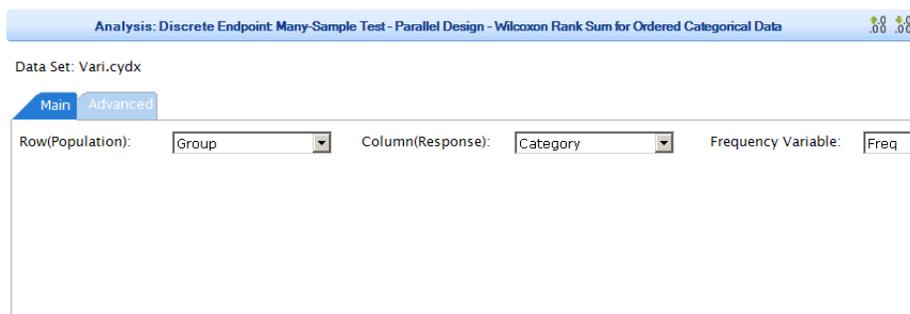
Analysis Steps

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

Analysis > (Discrete) Many Samples > (Single Arm Design) Wilcoxon Rank Sum for Ordered Categorical Data

This will display several input fields associated with the Chi-square Test in the main window.

3. In the **Main** tab, select **Group** in **Row(Population)**, **Category** in **Column(Response)**, and **Freq** in the variable **Frequency Variable**.



90 Analysis-Discrete: Many Proportions

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Wilcoxon Rank Sum Test

Hypothesis
 For $j = 1, 2, \dots, c$ and $k = 1, 2, \dots, s$ define $\pi_k = (\pi_{1k}, \pi_{2k}, \dots, \pi_{ck})$; $\pi'_k = (\pi'_{1k}, \pi'_{2k}, \dots, \pi'_{ck})$; and $V_{jk} = \sum_i \pi_{ik}$, and $V'_{jk} = \sum_i \pi'_{ik}$, where $i=1,2,\dots,j$.

$H_0: \pi_k = \pi'_k$ for all $k = 1, 2, \dots, s$ Vs either $H_1: V_{jk} \geq V'_{jk}$ Or $H_1: V_{jk} \leq V'_{jk}$ for 2-Sided test and
 $H_1: V_{jk} \geq V'_{jk}$ or $H_1: V_{jk} \leq V'_{jk}$ for 1-sided test with strict inequality for at least one j, k .

Input Parameters
 Data File: Vari.cydx
 Population: Group
 Response Variable: Category
 Frequency Variable: Freq
 Confidence Level: 0.95

Output
 Total Number of Records: 6
 Number of Records Rejected: 0
 Summary of Test Statistic:

| Response | Min | Max | Mean | Std.Dev | Observed | Standardized |
|----------|-------|-------|-------|---------|----------|--------------|
| Category | 425.5 | 601.5 | 517.5 | 28.606 | 470.5 | -1.643 |

Mann-Whitney Statistic: 194
 Test of Hypothesis:

| Type | p-value | | | |
|------------|-----------|------|---------|-----------|
| | Statistic | Tail | 1-Sided | 2*1-Sided |
| Asymptotic | -1.643 | L.E. | 0.05 | 0.1 |

Parameter Estimates:

| Type | CMLE | Std.Dev | 95% Confidence Interval | | 2-Sided p-value |
|------------|--------|---------|-------------------------|-------------|-----------------|
| | | | Lower Limit | Upper Limit | |
| Asymptotic | -0.062 | 0.039 | -0.139 | 0.016 | 0.118 |

Estimation of Odds Ratio:

| Column Value | Score | OR | Asymptotic 95% CI Limits | |
|--------------|-------|-----------|--------------------------|-----------|
| | | | Lower | Upper |
| 1 | 18.5 | Base Line | Base Line | Base Line |
| 2 | 38 | 0.3 | 0.066 | 1.357 |
| 3 | 41 | 0.25 | 0.044 | 1.422 |
| 4 | 43.5 | 0.214 | 0.031 | 1.479 |

Note that the output contains asymptotic inference for Wilcoxon Rank Sum Statistic as well as estimation of odds ratios for the categories with the corresponding confidence intervals

90.4 Example: Trend in R ordered proportions

Dataset: Korn_case_data.cydx

Data Description

Data from a prospective study of maternal drinking and congenital sex organ malformations (Graubard and Korn, 1987).

| | Maternal Alcohol Consumption (drinks/day) | | | | |
|--------------|---|-------|-------|-------|-----|
| Malformation | 0 | < 1 | 1 – 2 | 3 – 5 | ≥ 6 |
| Absent | 17066 | 14464 | 788 | 126 | 37 |
| Present | 48 | 38 | 5 | 1 | 1 |

Purpose of the Analysis:

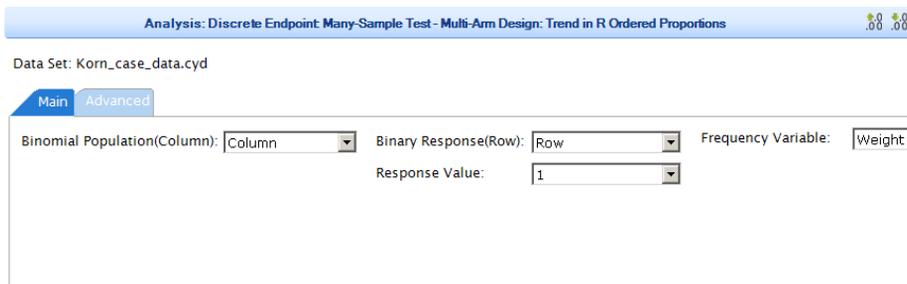
To test if a series of observed proportions all have the same underlying binomial response rate, where the alternative is that these rates are unequal, but ordered in some natural way. In other words, there is a trend in the binomial response rates.

Analysis Steps

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Discrete) Many Samples > (Multi-Arm Design) Trend in R Ordered Proportions

This will display several input fields associated with the Trend in R Ordered Proportions in the main window.

3. In the **Main** tab, select **Column** in **Binomial Population(Column)**, **Row** in **Binary Response(Row)** with **Response Value** of 1. Select **Weight** as the **Frequency Variable**.



90 Analysis-Discrete: Many Proportions

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binomial Response: Cochran-Armitage Trend Test

Hypothesis

$H_0: \pi_1 = \pi_2 \dots \pi_R$

Vs.

(i) Either $H_1: \pi_1 < \pi_2 < \dots \pi_R$ Or $\pi_1 > \pi_2 > \dots \pi_R$ for 2-sided test

Input Parameters

Data File: Korn_case_data.cyd

Population Variable: Column

Response Variable: Row(1)

Frequency Variable: Weight

Confidence Level: 0.95

Output

Total Number of Records: 10

Number of Records Rejected: 0

Table of Observed Frequencies:

| | 1 | 2 | 3 | 4 | 5 | Total |
|-------|-------|-------|-----|-----|----|-------|
| 1 | 17066 | 14464 | 788 | 126 | 37 | 32481 |
| 2 | 48 | 38 | 5 | 1 | 1 | 93 |
| Total | 17114 | 14502 | 793 | 127 | 38 | 32574 |

Summary of the Observed Data:

| Population | 1 | 2 | 3 | 4 | 5 | Total |
|----------------|-------|-------|-------|-------|-------|-------|
| Response Value | 17066 | 14464 | 788 | 126 | 37 | 32481 |
| Proportion | 0.997 | 0.997 | 0.994 | 0.992 | 0.974 | 0.997 |

Summary of Test Statistic:

| Row | Minimum | Maximum | Mean | Std. dev. | Observed Statistics | Standardized |
|-----|---------|---------|-----------|-----------|---------------------|--------------|
| 1 | 16304 | 16621 | 16573.546 | 5.582 | 16566 | -1.352 |

Test of Hypothesis:

| Statistic | p-value | | | | | 99% CI for exact P-Value | |
|-----------|---------|---------|-----------|------|---------|--------------------------|-------------|
| | Tail | 1-Sided | 2*1-Sided | Tail | 2 Sided | Lower Limit | Upper Limit |
| -1.352 | L.E. | 0.088 | 0.176 | G.E. | 0.176 | 0.171 | 0.191 |

Parameter Estimates:

| CMLE | Std.Dev | 95% Confidence Interval | | p-value |
|--------|---------|-------------------------|-------------|---------|
| | | Lower Limit | Upper Limit | 2-Sided |
| -0.228 | 0.168 | -0.558 | 0.102 | 0.176 |

Estimation of Odds Ratio:

| Column Value | Score | OR | 95% Confidence Interval | |
|--------------|-------|-----------|-------------------------|-----------|
| | | | Lower | Upper |
| 1 | 0 | Base Line | Base Line | Base Line |
| 2 | 1 | 0.796 | 0.572 | 1.108 |
| 3 | 2 | 0.634 | 0.328 | 1.227 |
| 4 | 3 | 0.505 | 0.188 | 1.358 |
| 5 | 4 | 0.402 | 0.107 | 1.504 |

Note that the output contains asymptotic inference for Cochran-Armitage Trend Test as well as estimation of odds ratios for the categories with the corresponding confidence intervals. The 2-sided p-value namely, 0.176 indicates that we are unable to reject the null hypothesis of no trend in the proportions across the categories.

90 Analysis-Discrete: Many Proportions

90.5 Example: Chi-Square Test for $R \times 2$ Proportions

Dataset: Fda1.cyx

Data Description

We are grateful to Dr. Mirza W. Ali of the Food and Drug Administration (FDA) for providing this data set. Animals were treated with four dose levels of a carcinogen and then observed (at necropsy) for the presence or absence of a tumor type. The data were stratified by survival time (in weeks) into the four time intervals 0–50, 51–80, 81–104, and terminal sacrifice. Since there were, no tumors found in the first time interval, this stratum may be excluded from data entry. The data for the remaining three strata are given below. We will use the stratum variable as a By variable.

| Stratum 1: 51–80 weeks of survival | | | | | |
|------------------------------------|--------------------|--------|---------|----------|-------|
| Disease Status | Dose of Carcinogen | | | | Total |
| | None | 1 unit | 5 units | 50 units | |
| Tumor Present | 0 | 0 | 0 | 1 | 1 |
| Tumor Absent | 7 | 10 | 6 | 8 | 31 |

| Stratum 2: 81–104 weeks of survival | | | | | |
|-------------------------------------|--------------------|--------|---------|----------|-------|
| Disease Status | Dose of Carcinogen | | | | Total |
| | None | 1 unit | 5 units | 50 units | |
| Tumor Present | 0 | 1 | 0 | 1 | 2 |
| Tumor Absent | 11 | 9 | 13 | 14 | 47 |

| Stratum 3: Sacrificed at end of 104 weeks | | | | | |
|---|--------------------|--------|---------|----------|-------|
| Disease Status | Dose of Carcinogen | | | | Total |
| | None | 1 unit | 5 units | 50 units | |
| Tumor Present | 1 | 1 | 1 | 2 | 5 |
| Tumor Absent | 29 | 26 | 28 | 20 | 103 |

Purpose of the Analysis:

To test if the data come from binomial distributions having same probability of response.

Analysis Steps

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Discrete) Many Samples > (Multi-Arm Design) Chi-square

Test for Rx2 Proportions

This will display several input fields associated with the Chi-square Test for Rx2 Proportions in the main window.

3. In the **Main** tab, select all variables as shown below.

Analysis Inputs

Analysis: Discrete Endpoint Many-Sample Test - Multi-Arm Design - Chi-Square for Rx2 Proportions

Data Set: Fda1.cyd

Main Advanced

Row(Group): column Column(Categories): row Frequency Variable: weight

4. In the **Advanced** tab, select **By Variable 1** as **Stratum**.

Analysis: Discrete Endpoint Many-Sample Test - Multi-Arm Design - Chi-Square for Rx2 Proportions

Data Set: Fda1.cyd

Main Advanced

By Variable 1: stratum Confidence Level: 0.95

By Variable 2:

5. Click **OK** to start analysis. The output is displayed in the main window as

90 Analysis-Discrete: Many Proportions

shown below:

Analysis: Binomial Response: Chi-square Test for Comparing Proportions in R groups and 2 Categ

$H_0: \pi_{1j} = \pi_{2j} \dots \pi_{rj}$ for all $j = 1, 2$
 V_s
 $H_1: \text{at least one } \pi_{ij} \text{ differs for } i = 1, 2, 3, \dots, r \text{ and } j = 1, 2$

Data File: Fda1.cyd
 Row Variable: column
 Column Variable: row
 Frequency Variable: weight
 Confidence Level: 0.95
 By Variable (s): stratum

Output

By: stratum = 1
 Total Number of Records: 8
 Number of Records Rejected: 3
 Table of Observed Frequencies:

| | 1 | 2 | Total |
|-------|---|----|-------|
| 0 | 0 | 7 | 7 |
| 1 | 0 | 10 | 10 |
| 5 | 0 | 6 | 6 |
| 50 | 1 | 8 | 9 |
| Total | 1 | 31 | 32 |

Pearson Chi-Square Statistic: 2.638
 Chi-Square Statistic with Yates CC: NA
 Likelihood Ratio Statistic: 2.621
 Test of hypothesis for Chi-Square Test

| | 1 | 2 | Total |
|-------|---|----|-------|
| 0 | 0 | 11 | 11 |
| 1 | 1 | 9 | 10 |
| 5 | 0 | 13 | 13 |
| 50 | 1 | 14 | 15 |
| Total | 2 | 47 | 49 |

Pearson Chi-Square Statistic: 2.172
 Chi-Square Statistic with Yates CC: NA
 Likelihood Ratio Statistic: 2.862
 Test of hypothesis for Chi-Square Test

| Type | p-value | | | | 99% CI for exact p-value | |
|------------|-----------|----|------|---------|--------------------------|-------------|
| | Statistic | DF | Tail | 2-Sided | Lower Limit | Upper Limit |
| Asymptotic | 2.172 | 3 | G.E. | 0.537 | 0.686 | 0.71 |

Warning Message: Since the asymptotic p value is not contained in the 99% CI, the Asymptotic Outcomes may be consider
 Test of hypothesis for Likelihood Ratio Test:

| Type | p-value | | | | 99% CI for exact p-value | |
|------------|-----------|----|------|---------|--------------------------|-------------|
| | Statistic | DF | Tail | 2-Sided | Lower Limit | Upper Limit |
| Asymptotic | 2.862 | 3 | G.E. | 0.413 | 0.682 | 0.705 |

Warning Message: Since the asymptotic p value is not contained in the 99% CI, the Asymptotic Outcomes may be considered unreliable.
 Measure of Association:

| | 1 | 2 | Total |
|-------|---|----|-------|
| 0 | 0 | 11 | 11 |
| 1 | 1 | 9 | 10 |
| 5 | 0 | 13 | 13 |
| 50 | 1 | 14 | 15 |
| Total | 2 | 47 | 49 |

Pearson Chi-Square Statistic: 2.172
 Chi-Square Statistic with Yates CC: NA
 Likelihood Ratio Statistic: 2.862
 Test of hypothesis for Chi-Square Test

| Type | Statistic | DF | p-value | | 99% CI for exact p-value | |
|------------|-----------|----|---------|---------|--------------------------|-------------|
| | | | Tail | 2-Sided | Lower Limit | Upper Limit |
| Asymptotic | 2.172 | 3 | G E | 0.537 | 0.686 | 0.71 |

Warning Message: Since the asymptotic p value is not contained in the 99% CI, the Asymptotic Outcomes may be considered unreliable.
 Test of hypothesis for Likelihood Ratio Test:

| Type | Statistic | DF | p-value | | 99% CI for exact p-value | |
|------------|-----------|----|---------|---------|--------------------------|-------------|
| | | | Tail | 2-Sided | Lower Limit | Upper Limit |
| Asymptotic | 2.862 | 3 | G E | 0.413 | 0.682 | 0.705 |

Warning Message: Since the asymptotic p value is not contained in the 99% CI, the Asymptotic Outcomes may be considered unreliable.
 Measure of Association:

| Coefficient | Estimate | ASE_MLE | 95% Confidence Interval | |
|------------------|----------|---------|-------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| Phi | 0.211 | 0.039 | 0.134 | 0.287 |
| Pearson's CC | 0.206 | 0.087 | 0.035 | 0.377 |
| Sakoda's CC | 0.291 | 0.124 | 0.049 | 0.534 |
| Tschuprow's CC | 0.16 | 0.041 | 0.08 | 0.24 |
| Cramer's V | 0.211 | 0.093 | 0.028 | 0.393 |
| Uncertainty(CIR) | 0.171 | 0.055 | 0.063 | 0.28 |

By: stratum = 3
 Total Number of Records: 8
 Number of Records Rejected: 0
 Table of Observed Frequencies:

| | 1 | 2 | Total |
|-------|---|-----|-------|
| 0 | 1 | 29 | 30 |
| 1 | 1 | 26 | 27 |
| 5 | 1 | 28 | 29 |
| 50 | 2 | 20 | 22 |
| Total | 5 | 103 | 108 |

Pearson Chi-Square Statistic: 1.25
 Chi-Square Statistic with Yates CC: NA
 Likelihood Ratio Statistic: 1.065
 Test of hypothesis for Chi-Square Test

| Type | Statistic | DF | p-value | | 99% CI for exact p-value | |
|------------|-----------|----|---------|---------|--------------------------|-------------|
| | | | Tail | 2-Sided | Lower Limit | Upper Limit |
| Asymptotic | 1.25 | 3 | G E | 0.741 | 0.804 | 0.824 |

Warning Message: Since the asymptotic p value is not contained in the 99% CI, the Asymptotic Outcomes may be considered unreliable.
 Measure of Association:

| Coefficient | Estimate | ASE_MLE | 95% Confidence Interval | |
|------------------|----------|---------|-------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| Phi | 0.108 | 0.025 | 0.058 | 0.157 |
| Pearson's CC | 0.107 | 0.115 | 0 | 0.333 |
| Sakoda's CC | 0.151 | 0.163 | 0 | 0.471 |
| Tschuprow's CC | 0.082 | 0.052 | 0 | 0.183 |
| Cramer's V | 0.108 | 0.117 | -0.123 | 0.338 |
| Uncertainty(CIR) | 0.026 | 0.053 | 0 | 0.131 |

Note that the output contains asymptotic inference for Chi-square test for Rx2

90 *Analysis-Discrete: Many Proportions*

proportions and Likelihood ratio test as well as the concordance coefficients for each of the value of stratum. All the 2-sided p-values are greater than 0.05 showing no evidence to reject the null hypothesis. This is true for all strata.

90.6 *Example: Chi-square Test for Prop in RxC Tables*

Dataset: Oral.cyx

Data Description

data is obtained on the location of oral lesions, in house to house surveys in three geographic regions of rural India. These data are displayed here in the form of 9×3 contingency table (Table 90.2) in which the counts are the number of patients with oral lesions at that site, in that geographic region.

Table 90.2: Oral Lesions Data Set

| Site of Lesion | Kerala | Gujarat | Andhra |
|----------------|--------|---------|--------|
| Labial Mucosa | 0 | 1 | 0 |
| Buccal Mucosa | 8 | 1 | 8 |
| Commissure | 0 | 1 | 0 |
| Gingiva | 0 | 1 | 0 |
| Hard Palate | 0 | 1 | 0 |
| Soft Palate | 0 | 1 | 0 |
| Tongue | 0 | 1 | 0 |
| Floor of Mouth | 1 | 0 | 1 |
| Alveolar Ridge | 1 | 0 | 1 |

Purpose of the Analysis:

To test if the distribution of the site of the oral lesion is significantly different in the three geographic regions.

Analysis Steps

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Discrete) Many Samples > (Multi-Arm Design) Chi-square for Proportions in RxC Tables

This will display several input fields associated with the Chi-square test for RxC proportions in the main window.

90 Analysis-Discrete: Many Proportions

3. In the **Main** tab select all variables as shown below.

The screenshot shows the SAS software interface for an analysis. At the top, a blue header bar contains the text "Analysis: Discrete Endpoint: Many-Sample Test - Multi-Arm Design: Chi-Square for Proportions in RxC Tables" and two small icons. Below this, the text "Data Set: Oral.cyx" is displayed. The interface has two tabs: "Main" (selected) and "Advanced". Under the "Main" tab, there are three input fields: "Row (Group):" with a dropdown menu set to "Row", "Column (Categories):" with a dropdown menu set to "Column", and "Frequency Variable:" with a text box containing "Weight".

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Categorical Response: Chi-square Test for Comparing Proportions

Hypothesis

$\pi_{i*} = \sum \pi_{ij}$ for $i = 1, 2, \dots, r$ $\pi_{*j} = \sum \pi_{ij}$ for $j = 1, 2, \dots, c$
 $H_0: \pi_{ij} = \pi_{i*} \pi_{*j}$ for all (i, j) pair Vs. $H_1: \pi_{ij} \neq \pi_{i*} \pi_{*j}$ for at least one $i \neq j$
 OR $H_0: \pi_{1j} = \pi_{2j} = \dots = \pi_{rj}$ for all $j = 1, 2, \dots, c$ Vs. $H_1: \text{at least one } \pi_{ij} \text{ differs}$

Input Parameters

Data File: Oral.cydx
 Row Variable: Row
 Column Variable: Column
 Frequency Variable: Weight
 Confidence Level: 0.95

Output

Total Number of Records: 13
 Number of Records Rejected: 0
 Table of Observed Frequencies:

| | 1 | 2 | 3 | Total |
|-------|----|---|----|-------|
| 1 | 0 | 1 | 0 | 1 |
| 2 | 8 | 1 | 8 | 17 |
| 3 | 0 | 1 | 0 | 1 |
| 4 | 0 | 1 | 0 | 1 |
| 5 | 0 | 1 | 0 | 1 |
| 6 | 0 | 1 | 0 | 1 |
| 7 | 0 | 1 | 0 | 1 |
| 8 | 1 | 0 | 1 | 2 |
| 9 | 1 | 0 | 1 | 2 |
| Total | 10 | 7 | 10 | 27 |

Pearson Chi-Square Statistic: 22.099
 Chi-Square Statistic with Yates CC: NA
 Likelihood Ratio Statistic: 23.297
 Test of hypothesis for Chi-Square Test

90 Analysis-Discrete: Many Proportions

| Type | | | p-value | | 99% CI for exact p-value | |
|------------|-----------|----|---------|---------|--------------------------|-------------|
| | Statistic | DF | Tail | 2-Sided | Lower Limit | Upper Limit |
| Asymptotic | 22.099 | 16 | G.E. | 0.14 | 0.021 | 0.029 |

Warning Message: Since the asymptotic p value is not contained in the 99% CI, the Asymptotic Out
Test of hypothesis for Likelihood Ratio Test:

| Type | | | p-value | | 99% CI for exact p-value | |
|------------|-----------|----|---------|---------|--------------------------|-------------|
| | Statistic | DF | Tail | 2-Sided | Lower Limit | Upper Limit |
| Asymptotic | 23.297 | 16 | G.E. | 0.106 | 0.027 | 0.036 |

Warning Message: Since the asymptotic p value is not contained in the 99% CI, the Asymptotic Outcomes may be consid
Measure of Association:

| Coefficient | Estimate | ASE_MLE | 95% Confidence Interval | |
|------------------|----------|---------|-------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| Phi | 0.905 | 0.159 | 0.593 | 1.216 |
| Pearson's CC | 0.671 | 0.036 | 0.601 | 0.741 |
| Sakoda's CC | 0.822 | 0.044 | 0.736 | 0.908 |
| Tschuprow's CC | 0.452 | 0.011 | 0.431 | 0.474 |
| Cramer's V | 0.64 | 0.088 | 0.468 | 0.812 |
| Uncertainty(C R) | 0.397 | 0.099 | 0.203 | 0.592 |

Note that the output contains asymptotic inference for Chi-square test for RxC proportions and Likelihood ratio test as well as the concordance coefficients for each of the value of stratum. The 2-sided p-value is 0.14 and 0.106 for chi-square and likelihood ratio tests, we are unable to reject H_0 . Note that in addition to the inference a warning is displayed as 'Warning: Since the asymptotic p-value is not contained in the 99% CI for exact p-value, the asymptotic outcomes may be considered unreliable.'

The sparseness of data is causing this problem. We recommend you to refer to the **StatXact** software for further details.

91 *Analysis-Binary Regression Analysis*

In this chapter we focus on how to run binary regression analysis in **East**. **East** provides logistic, probit, and complementary log-log regression models for data with a binary response variable. Along with regular maximum likelihood inference for logistic model, **East** provides Firth bias-correction for asymptotic estimates for unstratified logistic regression. Profile likelihood based confidence intervals for estimates are available for unstratified data.

Section 91.1 describes the Logistic Regression model for binary data and how **East** can be used to analyze data. Section 91.3 describes the Firth Procedure. Section 91.4 describes Profile Likelihood Based Confidence Intervals. Section 91.5 describes the Probit Model for Binary Data and Section 91.6 discusses the complementary Log-log Model which is also for binary data.

91.1 *Logistic Regression* *Example: Logistic Regression*

Consider a set of independent binary random variables, Y_1, Y_2, \dots, Y_n . Corresponding to each random variable, Y_j , there is a $(p \times 1)$ vector $\mathbf{x}_j = (x_{1j}, x_{2j}, \dots, x_{pj})'$ of explanatory variables (or covariates). Let π_j be the probability that $Y_j = 1$. Logistic regression models the dependency of π_j on \mathbf{x}_j through the relationship

$$\log \left(\frac{\pi_j}{1 - \pi_j} \right) = \gamma + \mathbf{x}_j' \beta, \quad (91.1)$$

where γ and $\beta \equiv (\beta_1, \beta_2, \dots, \beta_p)'$ are unknown parameters. We usually refer to γ as the constant term.

In this section, we demonstrate how **East** can be used to perform binary logistic regression analysis. Additionally, the asymptotic bias corrected estimates (Firth (1993)) and confidence intervals of the estimates using profile likelihood method (Venzon and Moolgavkar (1988)) based on the normal score function and the penalized score function are also available using **East**.

In addition to fitting the regression coefficients, **East** can also be used to:

- Perform significance testing of regression coefficients using Wald test
- Perform 1st order autocorrelation in residuals using Durbin-Watson test
- Compute collinearity diagnostics
- Compute different types of residuals
- Compute Influential statistics
- Compute predicted values
- Perform variable selection

91 Analysis-Binary Regression Analysis

Example: Logistic Regression

Dataset: LogisticData.cyx

Data Description

This data has been provided by Dr. S. Lai, University of Miami for a hospital based prospective study of perinatal infection and human immunodeficiency virus (HIV-1). Hutto, Parks, Lai, et al. (1991) investigated the possibility that the CD4 and CD8 blood serum levels measured in infants at 6 months of age might be good predictors of eventual HIV infection. In the dataset, CD4 and CD8 assume the values 0, 1, 2. However, these are not the actual blood serum levels. Rather they are coded surrogates for them.

The data on HIV infection rates and blood serum levels are tabulated below:

| Proportion Developing HIV | Serum Levels at 6 Months | |
|---------------------------|--------------------------|-----|
| | CD4 | CD8 |
| 4/7 (57%) | 0 | 0 |
| 1/1 (100%) | 0 | 2 |
| 2/7 (29%) | 1 | 0 |
| 4/12 (33%) | 1 | 1 |
| 2/2 (100%) | 1 | 2 |
| 0/2 (0%) | 2 | 0 |
| 0/13 (0%) | 2 | 1 |
| 1/3 (33%) | 2 | 2 |

Purpose of the Analysis:

We want to fit a Logistic model using the model terms, **CD4** and **CD8**. To specify the Logistic model **HIV = CD4+CD8** to the data.

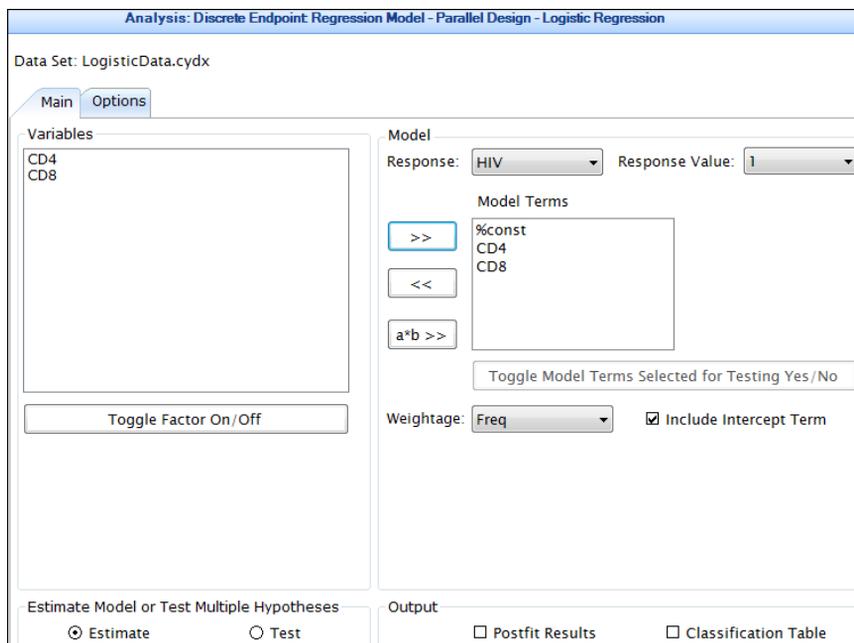
Analysis Steps: Regression based on Logistic Estimate

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

Analysis > (Discrete) Regression > (Parallel Design) Logistic Regression

3. In the **Main** tab, select **HIV** as the **Response** variable with **Response Value 1**, and **Freq** as the **Weightage** variable. Also notice that **% Const** is shown as a

model term. This is because **East**, by default always fits a model which includes the constant term unless you clear the “**Include Intercept Term**” check box. To specify an appropriate model, we would define HIV response rate as a function of CD4 and CD8, both covariates being regarded as ordinal. In the **Variables** box, select **CD4** and **CD8** and click button to include these terms under the **Model Terms**. Leave the default option as **Estimate**.



4. Click **OK** to estimate the regression coefficients. The maximum likelihood estimates, p-values, and confidence intervals for the regression parameters are

91 Analysis-Binary Regression Analysis

computed and displayed in the main window.

Analysis: Binary Regression: Logistic Model

Input Parameters

Data File: LogisticData.cydx
 Model: HIV(Response = 1)=%const+CD4+CD8
 Link Type: LOGIT
 Weight Variable: Freq
 Analysis Type: Estimate:Asymptotic
 Confidence Level: 0.95

Output

Terms in the Model : 3
 # Term(s) Dropped: 0
 Total Number of Records: 47
 Number of Records Rejected: 0
 # Groups: 8
 Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|--------|----|----------|
| Deviance | 4.471 | 5 | 0.484 |
| Likelihood Ratio | 23.652 | 3 | 2.952E-5 |

Parameter Estimates:

| Model Terms | Point Estimate | | | Confidence Interval and p-value for Beta | | |
|-------------|----------------|--------|----------|--|--------|-------------------|
| | Type | Beta | SE(Beta) | 95% CI | | 2*1-Sided p-value |
| | | | | Lower | Upper | |
| %const | MLE | 0.513 | 0.681 | -0.821 | 1.848 | 0.451 |
| CD4 | MLE | -2.542 | 0.839 | -4.187 | -0.897 | 0.002 |
| CD8 | MLE | 1.659 | 0.821 | 0.049 | 3.268 | 0.043 |

The third section is **Summary Statistics**. This section displays the deviance and its degrees of freedom, and the likelihood ratio statistic and degrees of freedom for testing the null hypothesis that the response probability of each observation is 0.5, i.e., all the model parameters, **including the constant term**, are simultaneously 0. The likelihood ratio statistic may be used to test for overall significance of the model. For the present example, the output displays a value of 4.471 on 5 df for the deviance, and a value of

23.652 on 3 df for the likelihood ratio statistic, with a p-value < 0.05 , thereby rejecting the null hypothesis that all the parameters of the model are 0.

The last section **Parameter Estimates**, displays the **Model Term**, **Point Estimate**, and the **Confidence Interval and p-value for Beta**. The **Model Terms** show there are two covariates, **CD4**, **CD8** in the model. The next three columns (under **Point Estimates**) show **MLE as Type**, estimates and standard error of Beta's. For **CD4**, the estimate of Beta is -2.542 . For **CD8**, the estimate of Beta is 1.659 . The next four columns show the inference type, confidence interval of Beta, and the p-value (2*1-sided) for testing Beta = 0. Here the p-value for **CD4** is 0.002.

Analysis Steps: Logistic Estimate in Odds Ratio

Here we would switch from displaying the regression parameters on the log scale (the default) to displaying them on the odds ratio scale.

1. In the **Options** tab, select **Odds Ratio/ Risk Ratio** in the **Output Parameter**.
2. Click **OK** to re-run the estimation, the parameter estimates are all transformed

91 Analysis-Binary Regression Analysis

by exponentiation into odds ratios.

Analysis: Binary Regression: Logistic Model

Input Parameters

Data File: LogisticData.cydx
 Model: HIV(Response = 1)=%const+CD4+CD8
 Link Type: LOGIT
 Weight Variable: Freq
 Analysis Type: Estimate:Asymptotic
 Confidence Level:0.95

Output

Terms in the Model : 3
 # Term(s) Dropped: 0
 Total Number of Records: 47
 Number of Records Rejected: 0
 # Groups: 8
 Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|--------|----|----------|
| Deviance | 4.471 | 5 | 0.484 |
| Likelihood Ratio | 23.652 | 3 | 2.952E-5 |

Parameter Estimates:

| Model Terms | Point Estimate | | | Confidence Interval and p-value for Odds Ratio | | |
|-------------|----------------|-------|----------|--|--------|-------------------|
| | Type | Beta | SE(Beta) | 95% CI | | 2*1-Sided p-value |
| | | | | Lower | Upper | |
| %const | MLE | 1.671 | NA | 0.44 | 6.346 | 0.451 |
| CD4 | MLE | 0.079 | NA | 0.015 | 0.408 | 0.002 |
| CD8 | MLE | 5.252 | NA | 1.05 | 26.257 | 0.043 |

- Now return to the default display by choosing **Beta** as the **Output parameter** in the **Options** tab.

Analysis Steps: Regression based on Factor Variables

In the LogisticData.cydx data set, CD4 and CD8 assume the values 0, 1, 2. However, these are not the actual blood serum levels. Rather they are coded surrogates for them. Thus suppose you are unwilling to treat CD4 and CD8 as ordinal variables, but would

like to treat them as factors. This requires that CD4 and CD8 each be split up into two dummy variables. The **Toggle Factor** option in the **Logistic Regression** dialog box accomplishes this splitting.

1. To accomplish this, press the Shift key on the keyboard, select **CD4** and **CD8** and then click on the **Toggle Factor On/Off** button. Notice that the **Model Terms** section of the window shows < fa > next to both CD4 and CD8. This means that CD4 has been split into two dummy variables, CD4_0 and CD4_1. The CD4_0 variable assumes the value 1 when CD4 is 0, and assumes the value 0 otherwise. The CD4_1 variable assumes the value 1 when CD4 is 1 and 0 otherwise. CD8 has been similarly split.
2. Click **OK** to obtain the unconditional maximum likelihood estimates of the regression coefficients for the model **HIV=CD4+CD8** with CD4 and CD8 declared as factor variables.

91 Analysis-Binary Regression Analysis

Analysis: Binary Regression: Logistic Model

Input Parameters

Data File: LogisticData.cdx
 Model: HIV(Response = 1)=%const+CD4_0+CD4_1+CD8_0+CD8_1
 Link Type: LOGIT
 Weight Variable: Freq
 Analysis Type: Estimate:Asymptotic
 Confidence Level: 0.95

Output

Terms in the Model : 5
 # Term(s) Dropped: 0
 Total Number of Records: 47
 Number of Records Rejected: 0
 # Groups: 8
 Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|-------|----|---------|
| Deviance | ? | ? | ? |
| Likelihood Ratio | ? | ? | ? |

Parameter Estimates:

| Model Terms | Point Estimate | | | Confidence Interval and p-value for Beta | | |
|-------------|----------------|------|----------|--|-------|-------------------|
| | Type | Beta | SE(Beta) | 95% CI | | 2*1-Sided p-value |
| | | | | Lower | Upper | |
| %const | MLE | ? | ? | ? | ? | ? |
| CD4_0 | MLE | ? | ? | ? | ? | ? |
| CD4_1 | MLE | ? | ? | ? | ? | ? |
| CD8_0 | MLE | ? | ? | ? | ? | ? |
| CD8_1 | MLE | ? | ? | ? | ? | ? |

Because the maximum likelihood estimates do not exist for this small data set, convergence is not possible in this case. The **Output** window only contains question marks for all the model terms.

This is not a problem in **East** alone. You will face the same difficulty with any other logistic regression software: **SAS**, **BMDP**, **GLIM** or **Egret**. The question is "Is there any other way to assess the significance of CD4 and CD8 when they are factor variables?" Interested users are referred to the **LogXact** software by **Cytel Inc**.

Analysis Steps: Test Multiple Hypothesis Regression

Suppose you are interested in a simultaneous test that the parameters corresponding to both **CD4** and **CD8** in the previously specified model are equal to 0.

1. Click the **Input Parameters** tab from the status bar below. Select variables **CD4** and **CD8** and click **Toggle Factor On/Off** button.
2. In the bottom left corner of the Input dialog, click the **Test** option. Select **CD4** and **CD8** in the **Model Terms** box and click the **Toggle Model Terms Selected for Testing Yes/No** button.

The screenshot shows the 'Analysis: Discrete Endpoint Regression Model - Parallel Design - Logistic Regression' dialog box. The 'Data Set' is 'LogisticData.cydx'. The 'Variables' list on the left contains 'CD4' and 'CD8'. The 'Model' section shows 'Response: HIV' and 'Response Value: 1'. The 'Model Terms' list contains '%const', 'CD4', and 'CD8'. The 'Test' list contains '%const(No)', 'CD4(Yes)', and 'CD8(Yes)'. A 'Toggle Model Terms Selected for Testing Yes/No' button is present. The 'Weightage' is set to 'Freq' and 'Include Intercept Term' is checked. At the bottom, the 'Estimate Model or Test Multiple Hypotheses' section has 'Test' selected, and the 'Output' section has 'Postfit Results' and 'Classification Table' options.

91 Analysis-Binary Regression Analysis

3. Click **OK** to start the analysis. **East** displays the following Output.

Analysis: Binary Regression: Logistic Model

Input Parameters

Data File: LogisticData.cydx
 Model: HIV(Response = 1)=%const+CD4+CD8
 Link Type: LOGIT
 Weight Variable: Freq
 Analysis Type: Test:Asymptotic
 Confidence Level: 0.95

Output

Terms in the Model : 3
 # Term(s) Dropped: 0
 Total Number of Records: 47
 Number of Records Rejected: 0
 # Groups: 8

Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|--------|----|----------|
| Deviance | 4.471 | 5 | 0.484 |
| Likelihood Ratio | 23.652 | 3 | 2.952E-5 |

Parameter Estimates:

| Model Terms | Point Estimate | | | Confidence Interval and p-value for Beta | | |
|-------------|----------------|--------|----------|--|--------|-------------------|
| | Type | Beta | SE(Beta) | 95% CI | | 2*1-Sided p-value |
| | | | | Lower | Upper | |
| %const | MLE | 0.513 | 0.681 | -0.821 | 1.848 | 0.451 |
| CD4 | MLE | -2.542 | 0.839 | -4.187 | -0.897 | 0.002 |
| CD8 | MLE | 1.659 | 0.821 | 0.049 | 3.268 | 0.043 |

Hypothesis Testing:

Test: <CD4=CD8=0>

| Type of Test | Statistics | DF | p-value |
|------------------|------------|----|----------|
| Score | 13.637 | 2 | 0.001 |
| Likelihood Ratio | 15.747 | 2 | 3.806E-4 |
| Wald | 9.195 | 2 | 0.01 |

The title **Hypothesis Testing Tests** $\langle CD4 = CD8 = 0 \rangle$ appears near the bottom of the Test results output. Below that, you can see the results of three tests: Scores, Likelihood ratio and Wald of the null hypothesis that the regression parameters corresponding to **CD4** and **CD8** are both 0. Since two parameters are being tested, this is a 2 degree of freedom test. All three tests are two-sided. Notice that the p-values based for all the tests are very small indicating that we reject the null hypothesis that the parameters corresponding to both **CD4** and **CD8** are equal to 0.

91.2 Receiver Operating Characteristic (ROC) Curve

Example: ROC curve
ROC Curve vs Classification Table
Example: Classification Table

As part of post-fit diagnostics, you can obtain the computed results that are required for producing an ROC curve. Before we discuss ROC curve in detail a few of the technical terms like **sensitivity** and **specificity** need to be explained.

Terms Explained Consider the example of a medical test carried out on a person to determine whether the person is suffering from HIV disease. Based on the test result, can we compute the probability that the person has the disease. The following table shows the possible alternatives that can occur.

| | Event (Disease Present) | Non-Event (Disease Absent) |
|---------------|-----------------------------------|---------------------------------------|
| Test Positive | Correct Event Prediction (a) | Incorrect Non-Event Prediction (b) |
| Test Negative | Incorrect Event Prediction (c) | Correct Non-Event Prediction (d) |

Suppose a, b, c and d denote the number of persons for whom the test results were as shown in the above table. Then we can define **Sensitivity** and **Specificity** of the test as given below.

The **Sensitivity** of a test is defined as the proportion of Correct Event predictions in the population having the event.

$$\text{Sensitivity} = \frac{a}{a+c}$$

The **Specificity** of a test is defined as the proportion of Correct Non-Event predictions in the population having the non-event.

$$\text{Specificity} = \frac{d}{b+d}$$

In other words, Sensitivity is a measure of **True Positive** and Specificity is a measure of **True Negative** of the test. The measure **False Positive** is given by **1-Specificity** or **1- True Negative** of the test.

91 Analysis-Binary Regression Analysis

When is a test positive?

Deciding on whether a test is positive or not may involve obtaining the value of a single prognostic variable and checking whether the value is more than or less than a pre-defined cut point. Or the test may involve several prognostic variables. A statistical model like **Binary Logistic Regression** may be used in such a situation, to estimate the probability of the disease. If the estimated probability is more than a pre-defined cut point, the test may be taken to be positive for the presence of the disease. For each such cut point of the probability, the sensitivity and specificity will vary. An ROC curve is a graphical representation of the tradeoff between **False Positive** and **True Positive** for various values of the cut point probability.

Example:ROC curve

Dataset: `LogisticData.cyx` as described in Section 91.1.

Purpose of the Analysis:

To fit logistic regression model and produce an ROC curve.

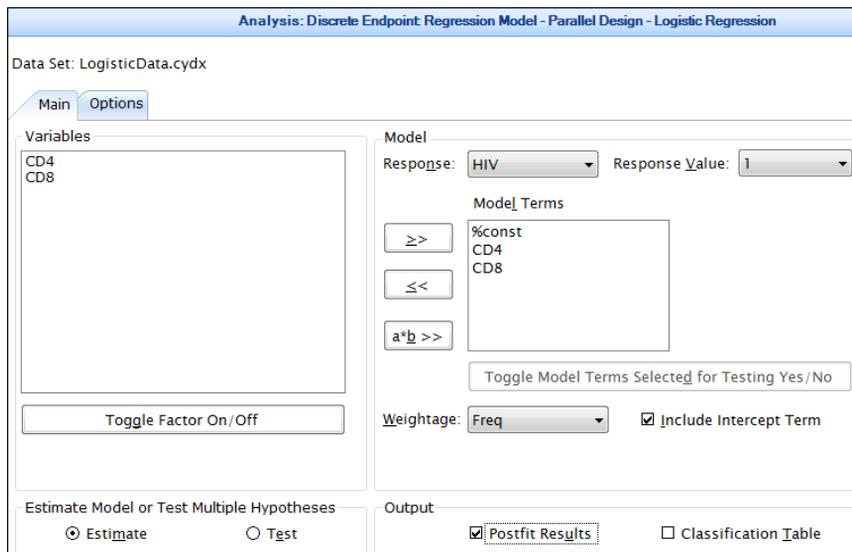
Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

Analysis > (Discrete) Regression > (Parallel Design) Logistic Regression

3. This will display several input fields associated with Logistic Regression in the main window. In the **Main** tab, select **HIV** as the **Response** variable with **Response Value 1**, and **Freq** as the **Weightage** variable. Select **Model Terms** to specify an appropriate model. To begin with, model the HIV response rate as a function of CD4 and CD8, both covariates being regarded as ordinal. In the **Variables** box, select **CD4** and **CD8** as the **Model Terms**. The variables **CD4** and **CD8** will appear in the **Model Terms** box. Leave the default option as

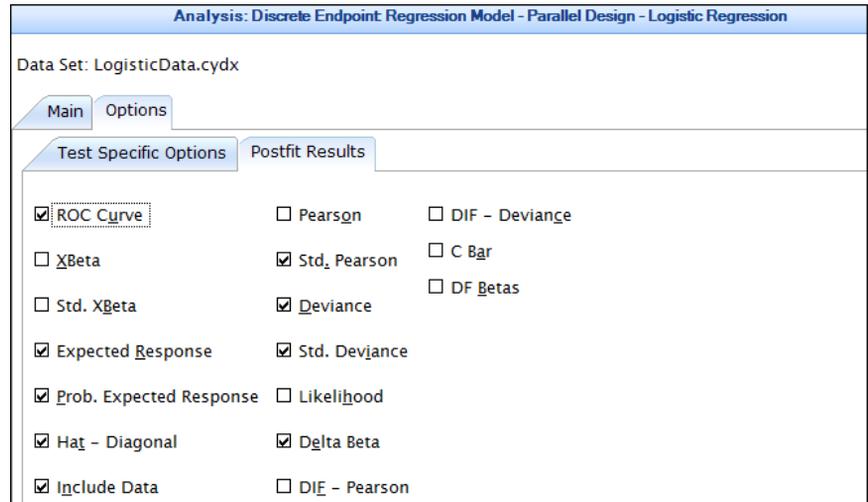
Estimate. In the Output dialog, select the **Postfit Results** checkbox.



4. In the Options tab, click the **Postfit Results** tab. Select the **ROC Curve**

91 Analysis-Binary Regression Analysis

checkbox.



5. Click **OK** to start the analysis. The following output is displayed in the main

window.

Analysis: Binary Regression: Logistic Model**Input Parameters**

Data File: LogisticData.cydx
Model: HIV(Response = 1)=%const+CD4+CD8
Link Type: LOGIT
Weight Variable: Freq
Analysis Type: Estimate:Asymptotic
Confidence Level: 0.95

Output

Terms in the Model : 3
Term(s) Dropped: 0
Total Number of Records: 47
Number of Records Rejected: 0
Groups: 8

Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|--------|----|----------|
| Deviance | 4.471 | 5 | 0.484 |
| Likelihood Ratio | 23.652 | 3 | 2.952E-5 |

Parameter Estimates:

| Model Terms | Point Estimate | | | Confidence Interval and p-value for Beta | | |
|-------------|----------------|--------|----------|--|--------|-------------------|
| | Type | Beta | SE(Beta) | 95% CI | | 2*1-Sided p-value |
| | | | | Lower | Upper | |
| %const | MLE | 0.513 | 0.681 | -0.821 | 1.848 | 0.451 |
| CD4 | MLE | -2.542 | 0.839 | -4.187 | -0.897 | 0.002 |
| CD8 | MLE | 1.659 | 0.821 | 0.049 | 3.268 | 0.043 |

91 Analysis-Binary Regression Analysis

Hosmer-Lemeshow Test:

| Statistics | Value | DF | p-value |
|-----------------|-------|----|---------|
| Hosmer-Lemeshow | 3.396 | 3 | 0.335 |

| Hosmer-Lemeshow Percentile of Risk | Response | | Non-Response | | Total |
|------------------------------------|----------|----------|--------------|----------|-------|
| | Observed | Expected | Observed | Expected | |
| 1 | 0 | 0.691 | 15 | 14.309 | 15 |
| 2 | 2 | 0.814 | 5 | 6.186 | 7 |
| 3 | 5 | 5.57 | 10 | 9.43 | 15 |
| 4 | 4 | 4.379 | 3 | 2.621 | 7 |
| 5 | 3 | 2.547 | 0 | 0.453 | 3 |
| Total | 14 | 14 | 33 | 33 | |

6. The postfit output in the output sheet titled "Regression Diagnostics1" is as follows:

Analysis: Binary Regression: Logistic Model

Regression Diagnostics

| Sequence | GrpSize | ObsvResp | ExptResp | ProbResp | Hat_Diag | StdPearson | Deviance | Std_Deviance |
|----------|---------|----------|----------|----------|----------|------------|----------|--------------|
| 1 | 2 | 0 | 0.021 | 0.01 | 0.044 | -0.147 | -0.203 | -0.208 |
| 2 | 13 | 0 | 0.671 | 0.052 | 0.483 | -1.17 | -1.174 | -1.632 |
| 3 | 7 | 2 | 0.814 | 0.116 | 0.441 | 1.872 | 1.212 | 1.621 |
| 4 | 3 | 1 | 0.667 | 0.222 | 0.36 | 0.579 | 0.441 | 0.551 |
| 5 | 12 | 4 | 4.903 | 0.409 | 0.488 | -0.741 | -0.537 | -0.75 |
| 6 | 7 | 4 | 4.379 | 0.626 | 0.76 | -0.604 | -0.293 | -0.599 |
| 7 | 2 | 2 | 1.568 | 0.784 | 0.363 | 0.93 | 0.987 | 1.236 |
| 8 | 1 | 1 | 0.979 | 0.979 | 0.059 | 0.152 | 0.207 | 0.214 |

7. Notice the column titled **ProbResp** containing the Estimated Response probabilities computed from the fitted model. These probability values are used as the cut points for carrying out the computations that are in the **ROC-Curve**

worksheet, which is shown below.

Analysis: Binary Regression: Logistic Model

ROC-Curve

| Corr_events | Corr_noevents | Incorr_events | Incorr_noevents | Sensitivity | 1-Specificity |
|-------------|---------------|---------------|-----------------|-------------|---------------|
| 14 | 0 | 33 | 0 | 1 | 1 |
| 14 | 2 | 31 | 0 | 1 | 0.939 |
| 14 | 15 | 18 | 0 | 1 | 0.545 |
| 12 | 20 | 13 | 2 | 0.857 | 0.394 |
| 11 | 22 | 11 | 3 | 0.786 | 0.333 |
| 7 | 30 | 3 | 7 | 0.5 | 0.091 |
| 3 | 33 | 0 | 11 | 0.214 | 0 |
| 1 | 33 | 0 | 13 | 0.071 | 0 |

Examine how the computations for ROC-Curve are carried out. First take the cut point probability as 0.01, the first value in the column **ProbResp** from **Regression Diagnostics** results. The rule in using this cut point is that for any individual in the data set, compute the expected probability for response and if this probability is ≥ 0.01 , allot that individual as **Response** or **Event**. These expected or predicted probabilities are already computed and are shown in the column titled **ProbResp**. We can tabulate the prediction results for this rule as shown below.

Cut Point: $z = 0.01$

Rule: An individual is 'Response' or 'Event' if $\text{ProbResp} \geq z$.

Since for all the groups, $\text{ProbResp} \geq 0.01$, all the individuals in all the groups are predicted as 'Response'.

91 Analysis-Binary Regression Analysis

| GrpSize | Observed | | Model ProbResp | Predicted | |
|---------|----------|----------|----------------|-----------|----------|
| | Resp | Non-Resp | | Resp | Non-Resp |
| 2 | 0 | 2 | 0.010251 | 2 | 0 |
| 13 | 0 | 13 | 0.051587 | 13 | 0 |
| 7 | 2 | 5 | 0.11625 | 7 | 0 |
| 3 | 1 | 2 | 0.222194 | 3 | 0 |
| 12 | 4 | 8 | 0.408579 | 12 | 0 |
| 7 | 4 | 3 | 0.625565 | 7 | 0 |
| 2 | 2 | 0 | 0.783934 | 2 | 0 |
| 1 | 1 | 0 | 0.97876 | 1 | 0 |

By comparing the Predicted figures and Observed figures, we can tabulate 'Predicted Correct' numbers as shown below.

| Grp Size | Observed | | Model ProbResp | Predicted | | Predicted Correct | |
|----------|----------|----------|----------------|-----------|----------|-------------------|----------|
| | Resp | Non-Resp | | Resp | Non-Resp | Resp | Non-Resp |
| 2 | 0 | 2 | 0.010251 | 2 | 0 | 0 | 0 |
| 13 | 0 | 13 | 0.051587 | 13 | 0 | 0 | 0 |
| 7 | 2 | 5 | 0.11625 | 7 | 0 | 2 | 0 |
| 3 | 1 | 2 | 0.222194 | 3 | 0 | 1 | 0 |
| 12 | 4 | 8 | 0.408579 | 12 | 0 | 4 | 0 |
| 7 | 4 | 3 | 0.625565 | 7 | 0 | 4 | 0 |
| 2 | 2 | 0 | 0.783934 | 2 | 0 | 2 | 0 |
| 1 | 1 | 0 | 0.97876 | 1 | 0 | 1 | 0 |

By subtracting 'Predicted Correct' numbers from 'Predicted' numbers, 'predicted Incorrect' numbers can be obtained as shown below.

| Grp Size | Observed | | Model ProbResp | Predicted | | Predicted Correct | | Predicted Incorrect | |
|----------|----------|----------|----------------|-----------|----------|-------------------|----------|---------------------|----------|
| | Resp | Non-Resp | | Resp | Non-Resp | Resp | Non-Resp | Resp | Non-Resp |
| 2 | 0 | 2 | 0.010251 | 2 | 0 | 0 | 0 | 2 | 0 |
| 13 | 0 | 13 | 0.051587 | 13 | 0 | 0 | 0 | 13 | 0 |
| 7 | 2 | 5 | 0.11625 | 7 | 0 | 2 | 0 | 5 | 0 |
| 3 | 1 | 2 | 0.222194 | 3 | 0 | 1 | 0 | 2 | 0 |
| 12 | 4 | 8 | 0.408579 | 12 | 0 | 4 | 0 | 8 | 0 |
| 7 | 4 | 3 | 0.625565 | 7 | 0 | 4 | 0 | 3 | 0 |
| 2 | 2 | 0 | 0.783934 | 2 | 0 | 2 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0.97876 | 1 | 0 | 1 | 0 | 0 | 0 |
| Total | | | | | | 14 | 0 | 33 | 0 |

The figures in the last line ‘Total’, 14, 0, 33 and 0 are what you saw in the first row of ROC table.

| | Event (Disease Present) | Non-Event (Disease Absent) |
|---------------|-------------------------------------|--|
| Test Positive | Correct Event Prediction (a=14) | Incorrect Non-Event Prediction (b=33) |
| Test Negative | Incorrect Event Prediction (c=0) | Correct Non-Event Prediction (d=0) |

$$\text{Sensitivity} = \frac{a}{a+c} = \frac{14}{14+0} = 1$$

$$\text{Specificity} = \frac{d}{b+d} = \frac{0}{33+0} = 0$$

Hence,

$$1 - \text{Specificity} = 1 - 0 = 1$$

The above values of Sensitivity and (1-Specificity), 1 and 1 are what you see in the first row of ROC table.

If you carry out similar computations for the fifth group with the cut point of $z = 0.408579$ you will get the following results.

91 Analysis-Binary Regression Analysis

| Grp Size | Observed | | Model ProbResp | Predicted | | Predicted Correct | | Predicted Incorrect | |
|----------|----------|----------|----------------|-----------|----------|-------------------|----------|---------------------|----------|
| | Resp | Non-Resp | | Resp | Non-Resp | Resp | Non-Resp | Resp | Non-Resp |
| 2 | 0 | 2 | 0.010251 | 0 | 2 | 0 | 2 | 0 | 0 |
| 13 | 0 | 13 | 0.051587 | 0 | 13 | 0 | 13 | 0 | 0 |
| 7 | 2 | 5 | 0.11625 | 0 | 7 | 0 | 5 | 0 | 2 |
| 3 | 1 | 2 | 0.222194 | 0 | 3 | 0 | 2 | 0 | 1 |
| 12 | 4 | 8 | 0.408579 | 12 | 0 | 4 | 0 | 8 | 0 |
| 7 | 4 | 3 | 0.625565 | 7 | 0 | 4 | 0 | 3 | 0 |
| 2 | 2 | 0 | 0.783934 | 2 | 0 | 2 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0.97876 | 1 | 0 | 1 | 0 | 0 | 0 |
| Total | | | | | | 11 | 22 | 11 | 3 |

The figures in the last line 'Total', 11, 22, 11, and 3 are what you see in the fifth row of the ROC table.

| | Event (Disease Present) | Non-Event (Disease Absent) |
|---------------|-------------------------------------|--|
| Test Positive | Correct Event Prediction (a=11) | Incorrect Non-Event Prediction (b=11) |
| Test Negative | Incorrect Event Prediction (c=3) | Correct Non-Event Prediction (d=22) |

$$\text{Sensitivity} = \frac{a}{a+c} = \frac{11}{11+3} = 0.785714$$

$$\text{Specificity} = \frac{d}{b+d} = \frac{22}{22+11} = 0.666667$$

Hence,

$$1 - \text{Specificity} = 1 - 0.666667 = 0.333333$$

The above values of Sensitivity and (1-Specificity), 0.785714 and 0.333333, respectively, are what you see in the fifth row of the ROC table.

You have just seen the computations required to obtain the results shown in the ROC table, for 2 cut points. In a similar way, you can check the computations for the remaining 6 cut points.

ROC Curve vs Classification Table

Similar to ROC Curve computations, **East** also provides Classification Table estimates. Though in both the types of analysis, we get information on Sensitivity and Specificity estimates, they differ in the following way:

1. The classification error estimates computed in ROC Curve for any observation is biased because the model used was fitted with data that included that observation. In Classification Table, this bias is eliminated by again estimating the model parameters after leaving out each observation one at a time and then classifying the observation based on new estimates. These new estimates are actually produced as one-step approximations from the computations carried out for the complete data and no separate models are fitted. The formulas used are listed in Appendix W.
2. Classification Table uses Bayes' theorem and computes posterior probabilities in classification, using prior probabilities and probabilities of events.

Example: Classification Table

You can obtain classification table information using the **Classification table** option.

Dataset: **LogisticData.cyx** as described in Section 91.1.

Purpose of the Analysis:

To fit logistic regression and obtain classification table.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

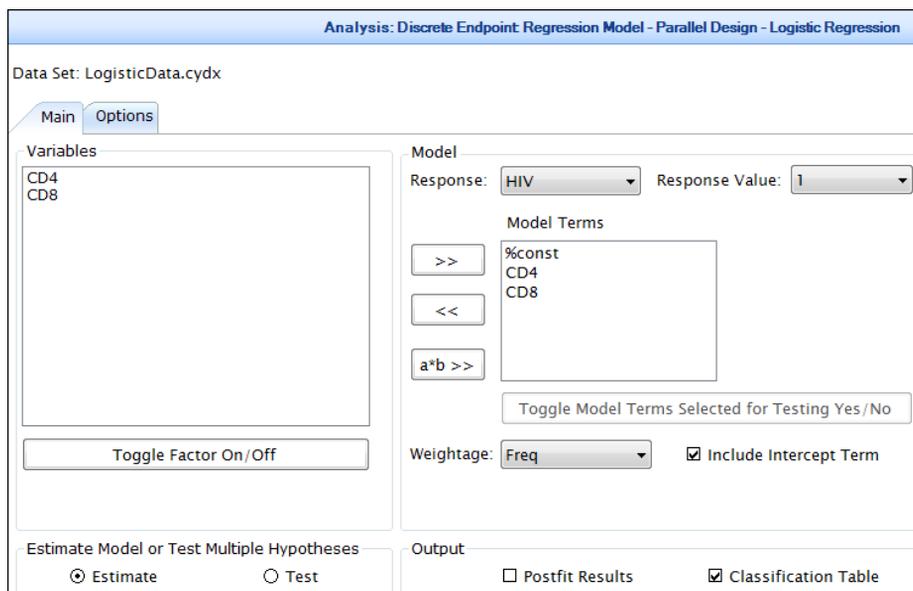
Analysis > (Discrete) Regression > (Parallel Design) Logistic Regression

This will display several input fields associated with Logistic Regression in the main window.

3. In the **Main** tab, select the variables as shown below. Make sure to select the

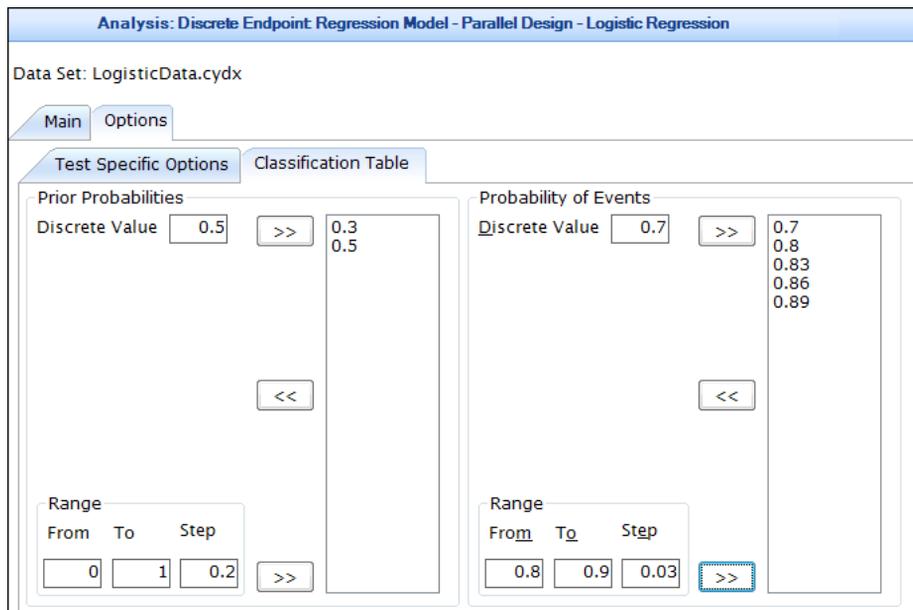
91 Analysis-Binary Regression Analysis

Classification Table checkbox in the Output dialog



- In the **Options** tab, click the **Classification Table** tab. In the ensuing dialog box, specify the set of values for **Prior Probabilities** and **Probability of Events**. You can specify these values as discrete values, by entering each value in the box against 'Discrete value' and then clicking to post it on the right side box. If your set of values is a range of equidistant values, then you can specify the starting value (From), the ending value (To) and the step value (Step) and then click . **East** will compute the individual values in the range and display them on the right side box. You are allowed to specify some values as discrete and some as a range. For this example, in the **Prior Probabilities** section, enter 0.3 and 0.5 in the **Discrete Value** and click . In the **Probability of Events** section, enter 0.7 and click . In the Range, enter 0.8 in the **From** value, 0.9 in the **To** value and 0.03 in the **Step** value. Click button next to the **Range**

ensuing box.



For each combination of the values of ‘prior probability’ and ‘probability of event’, East would produce classification results.

- Click **OK** to start the analysis. The output will be displayed in the main window. In the library along with the **Analysis: Binary Regression: Logistic Model1** node there is another node named the **Classification Table1**. Click the node **Classification Table1** to see the classification table.

Analysis: Binary Regression: Logistic Model

Classification Table

| Seq Number | Prob_event | Cut_off_Prob | Correct_Events | Correct_NoEvents | Incorrect_Events | Incorrect_NoEvents | Percent_Correct |
|------------|------------|--------------|----------------|------------------|------------------|--------------------|-----------------|
| 1 | 0.7 | 0.3 | 11 | 22 | 11 | 3 | 0.75 |
| 2 | 0.7 | 0.5 | 7 | 30 | 3 | 7 | 0.623 |
| 3 | 0.8 | 0.3 | 11 | 22 | 11 | 3 | 0.762 |
| 4 | 0.8 | 0.5 | 7 | 30 | 3 | 7 | 0.582 |
| 5 | 0.83 | 0.3 | 11 | 22 | 11 | 3 | 0.765 |
| 6 | 0.83 | 0.5 | 7 | 30 | 3 | 7 | 0.57 |
| 7 | 0.86 | 0.3 | 11 | 22 | 11 | 3 | 0.769 |
| 8 | 0.86 | 0.5 | 7 | 30 | 3 | 7 | 0.557 |
| 9 | 0.89 | 0.3 | 11 | 22 | 11 | 3 | 0.773 |
| 10 | 0.89 | 0.5 | 7 | 30 | 3 | 7 | 0.545 |

91 Analysis-Binary Regression Analysis

91.3 Firth Procedure

Example: Firth Procedure

Since the MLE is only asymptotically unbiased, various methods have been proposed to reduce the bias. One such approach is due to Firth (1993), which reduces the bias in the MLE by introducing a small bias into the score function. The general idea is to remove the $O(n^{-1})$ term in the expression for the bias of the MLE. This is accomplished by calculating the posterior mode based on Jeffrey's prior. One advantage of the Firth estimator is that it exists when there is complete separation or quasi-complete separation.

Example: Firth Procedure

Dataset: esr.cyx

Data Description

The Firth estimator performs well under separation and near separation and we will illustrate the improvement over the MLE by using a well-known dataset that was originally given by Collett and Jermain (1985) and is also found in Collett (2002).

The response variable was erythrocyte sedimentation rate (ESR), which is used as an indicator of infections and certain types of diseases. The lower the ESR value the better, and as so often it happens in medical applications, the continuous response variable was dichotomized with less than 20 assigned a value of zero and at least 20 assigned a value of one. The two predictor variables are Fibrinogen and γ -globulin.

The data were obtained in a study performed by the Institute of Medical Research, Kuala Lumpur, Malaysia.

Purpose of the Analysis:

To determine if a patient's ESR value is a valuable diagnostic. This is accomplished by trying to determine if there is a relationship between ESR and the two predictors, since the latter are commonly elevated in the presence of inflammatory diseases

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

Analysis > (Discrete) Regression > (Parallel Design) Logistic Regression

3. In the **Main** tab, select **esr** as the **Response** variable with **Response value** as 1. Select **fibrinogn** and **gam_glob** as the **Model Terms**.

4. In the **Options** tab the default asymptotic method is Maximum Likelihood Estimate.
5. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binary Regression: Logistic Model

Input Parameters

Data File: esr.cydx
Model: esr(Response = 1)=%const+fibrinogn+gam_glob
Link Type: LOGIT
Analysis Type: Estimate:Asymptotic
Confidence Level: 0.95

Output

Terms in the Model : 3
Term(s) Dropped: 0
Total Number of Records: 32
Number of Records Rejected: 0
Groups: 32

Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|--------|----|----------|
| Deviance | 22.971 | 29 | 0.778 |
| Likelihood Ratio | 21.39 | 3 | 8.735E-5 |

Parameter Estimates:

| Model Terms | Point Estimate | | | Confidence Interval and p-value for Beta | | |
|-------------|----------------|---------|----------|--|--------|-------------------|
| | Type | Beta | SE(Beta) | 95% CI | | 2*1-Sided p-value |
| | | | | Lower | Upper | |
| %const | MLE | -12.792 | 5.796 | -24.153 | -1.431 | 0.027 |
| fibrinogn | MLE | 1.91 | 0.971 | 0.007 | 3.813 | 0.049 |
| gam_glob | MLE | 0.156 | 0.12 | -0.079 | 0.39 | 0.193 |

6. Click the Input dialog, specify the same model. In the **Options** tab, choose

91 Analysis-Binary Regression Analysis

Penalized MLE for bias correction (Firth's method) as the Type of MLE.

Analysis: Discrete Endpoint: Regression Model - Parallel Design - Logistic Regression

Data Set: esr.cydx

Main Options

Test Specific Options

Parameters

Output Parameter: Confidence Level [0.001 to 0.999]:

Type of MLE: Iteration Limit for Asymptotic Convergence:

Output p - value: Convergence Criteria for Asymptotic Model:

Profile Likelihood

Display Covariance Matrix

7. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binary Regression: Logistic Model

Input Parameters

Data File: esr.cydx
 Model: esr(Response = 1)=%const+fibrinogn+gam_glob
 Link Type: LOGIT
 Analysis Type: Estimate:Asymptotic
 Confidence Level: 0.95

Output

Terms in the Model : 3
 # Term(s) Dropped: 0
 Total Number of Records: 32
 Number of Records Rejected: 0
 # Groups: 32

Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|--------|----|----------|
| Deviance | 23.285 | 29 | 0.763 |
| Likelihood Ratio | 21.076 | 3 | 1.015E-4 |

Parameter Estimates:

| Model Terms | Point Estimate Type | Beta | SE(Beta) | Confidence Interval and p-value for Beta | | |
|-------------|--|---------|----------|--|--------|-------------------|
| | | | | 95% CI | | 2*1-Sided p-value |
| | | | | Lower | Upper | |
| %const | Penalized MLE for bias Correction (Firth's Method) | -10.289 | 4.931 | -19.953 | -0.624 | 0.037 |
| fibrinogn | Penalized MLE for bias Correction (Firth's Method) | 1.471 | 0.815 | -0.127 | 3.069 | 0.071 |
| gam_glob | Penalized MLE for bias Correction (Firth's Method) | 0.127 | 0.109 | -0.086 | 0.34 | 0.243 |

We can see that MLE and Firth estimates differ. However, the Confidence Intervals for Penalized MLE's are shorter than those for MLE's.

91.4 Profile Likelihood Based Confidence Intervals

Example

Classical Wald's confidence intervals are based on the asymptotic normality of the maximum likelihood estimate of a parameter. However, in case of small samples, the properties of the estimator can be very different. A symmetric shape of the likelihood function allows use of Wald's intervals, while an asymmetric shape may result into inaccurate confidence intervals. A more robust construction of confidence intervals is derived from the asymptotic χ^2 distribution of the generalized likelihood ratio test. We have seen in Section 91.3 that Firth's estimator is recommended whenever there is a problem of separation, and is a better alternative to Exact when the latter is not computationally feasible. The problem of separation also leads to inflated standard error which results into an infinite or large Wald's confidence intervals. In such situations, the confidence intervals based on profile likelihood method are a way out. Heinze and Schemper (2002) show that the confidence intervals based on profile likelihood are often preferable to Wald's confidence intervals. Heinze (2006) demonstrated that the confidence intervals based on penalized likelihood equation show excellent behavior in terms of the coverage probability and the higher power.

Example

Dataset: `esr.cydx` as described in Section 91.3.

Purpose of the Analysis:

This example includes the confidence intervals based on profile likelihood method for MLE and PMLE estimates.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

Analysis > (Discrete) Regression > (Parallel Design) Logistic Regression

3. In the **Main** tab, select `esr` as the **Response** variable. Choose a value of 1 as the

91 Analysis-Binary Regression Analysis

Response value. Select **fibrinogn** and **gam_glob** as the **Model Terms**.

Data Set: esr.cydx

Main
Options

Variables

fibrinogn
 gam_glob

Toggle Factor On/Off

Model

Response: esr Response Value: 1

Model Terms

>> %const
 fibrinogn
 gam_glob

<< a*b >>

Toggle Model Terms Selected for Testing Yes/No

Weightage: Include Intercept Term

Estimate Model or Test Multiple Hypotheses

Estimate Test

Output

Postfit Results Classification Table

4. In the **Options** tab, select the **Profile Likelihood** and **Display Covariance Matrix** check boxes.

Data Set: esr.cydx

Main
Options

Test Specific Options

Parameters

Output Parameter: Beta Confidence Level [0.001 to 0.999]: 0.95

Type of MLE: Penalized MLE for bias Correct Iteration Limit for Asymptotic Convergence: 20

Output p - value: Two Sided Convergence Criteria for Asymptotic Model: 0

Profile Likelihood

Display Covariance Matrix

5. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binary Regression: Logistic Model

Input Parameters

Data File: esr.cydx
 Model: esr(Response = 1)=%const+fibrinogn+gam_glob
 Link Type: LOGIT
 Analysis Type: Estimate:Asymptotic
 Confidence Level: 0.95

Output

Terms in the Model : 3
 # Term(s) Dropped: 0
 Total Number of Records: 32
 Number of Records Rejected: 0
 # Groups: 32

Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|--------|----|----------|
| Deviance | 23.285 | 29 | 0.763 |
| Likelihood Ratio | 21.076 | 3 | 1.015E-4 |

Parameter Estimates:

| Model Terms | Type | Point Estimate | | Confidence Interval and p-value for Beta | | |
|-------------|--|----------------|----------|--|--------|-------------------|
| | | Beta | SE(Beta) | 95% CI | | 2*1.Sided p-value |
| | | | | Lower | Upper | |
| %const | Penalized MLE for bias Correction (Firth's Method) | -10.289 | 4.931 | -19.953 | -0.624 | 0.037 |
| fibrinogn | Penalized MLE for bias Correction (Firth's Method) | 1.471 | 0.815 | -0.127 | 3.069 | 0.071 |
| gam_glob | Penalized MLE for bias Correction (Firth's Method) | 0.127 | 0.109 | -0.086 | 0.34 | 0.243 |
| | | | | -0.07 | 0.357 | |

Covariance Matrix:

| | %const | fibrinogn | gam_glob |
|-----------|--------|-----------|----------|
| %const | 24.315 | -2.318 | -0.464 |
| fibrinogn | -2.318 | 0.665 | 0.009 |
| gam_glob | -0.464 | 0.009 | 0.012 |

Note that for every estimate now there are two confidence intervals. The later one is based on Profile likelihood. You can as well have the profile based Confidence Intervals when the Penalized MLE option is chosen.

91 Analysis-Binary Regression Analysis

91.5 Probit Regression

91.5.0 Example

The probit model is a generalized linear model that uses the inverse cumulative distribution function (cdf) from the standard normal distribution as a link function. Let y_i be a binary response for subject i , $i = 1, \dots, n$, such that $y_i = 1$ if subject i experiences a "success" and $y_i = 0$ otherwise. Further, let π_i and \mathbf{x}_i be the probability of a response and a vector of covariates for subject i , $i = 1, \dots, n$, respectively. A probit model for y_i is

$$\Phi^{-1}(\pi_i) = \beta_0 + \beta'_b \mathbf{x}_i,$$

where Φ is the standard normal cdf. Here, as in the case of logistic regression, the link function Φ^{-1} maps the (0,1) scale for π_i onto the scale of the entire real line for the linear predictor $\beta_0 + \beta'_b \mathbf{x}_i$. Also similar to the logistic case, the probit link is symmetric around 0.5 in the sense that $\Phi^{-1}(\pi) = -\Phi^{-1}(1 - \pi)$. Thus, the response curve for the probability of a response π is symmetric around 0.5.

Example: Probit Regression

Dataset: Devtox.cydx.

Data Description

This data set contains 1,512 observations of which you can only see the first few. Use the horizontal and vertical scroll bars or the and keys to examine the data set. There are 8 variables, **ID**, **Dose**, **Death**, **Weight**, **Malf**, **Sex**, **Impl** and **LittSz**, and 1,512 cases (1,512 implantations in 112 litters). The explanation of each variable represents and their codes are described below:

| Variable | Description | Code |
|----------|--|------------------|
| Dose | dose administered in g/kg body weight | 0, 0.5, 1 or 2 |
| Death | fetal death | 1=Yes, 0=No |
| Weight | fetal weight in grams | |
| Malf | fetal malformation | 1=Yes, 0=No |
| Sex | gender of the rat | 1=Male, 2=Female |
| Impl | number of implantations in the litter | |
| LittSz | number of live offspring in the litter | |

Purpose of the Analysis:

We will analyze a single binary outcome, **death**, in a developmental toxicity study of a substance conducted in rats, through a probit model. We want to fit a probit model using the model terms, **Dose**, **Impl** and their interaction **Dose*Impl**. To specify the probit model:

$$\text{Death} = \text{Dose} + \text{Impl} + \text{Dose} * \text{Impl}$$

to the data.

Analysis Steps: Probit Regression - Estimate Model

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

Analysis > (Discrete) Regression > (Parallel Design) Probit Regression

3. In the **Main** tab, select **Death** as the **Response** variable with Response value as
 1. Select **Dose** and **Impl** as the **Model Terms**. Add an interaction term **Dose*Impl**: click on **Dose** in the **Variables** section, Press the Ctrl key on the

91 Analysis-Binary Regression Analysis

keyboard, click on **Impl** in the **Variables** section, and click the **a*b** > button.

The screenshot shows a software interface for a binary regression analysis. The data set is 'Devtox.cyx'. The 'Variables' list on the left includes Id, Dose, Weight, Malf, Sex, Impl, and LittSz. The 'Model' section on the right has 'Response' set to 'Death' and 'Response Value' set to '1'. The 'Model Terms' list includes %const, Dose, Impl, and Dose*Impl. The 'a*b >>' button is highlighted. At the bottom, there are options for 'Estimate Model or Test Multiple Hypotheses' (with 'Estimate' selected) and 'Output' (with 'Postfit Results' and 'Classification Table' options).

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binary Regression: Probit Model

Input Parameters

Data File: Devtox.cydx
 Model: Death(Response = 1)=%const+Dose+Impl+Dose*Impl
 Link Type: Probit
 Analysis Type: Estimate:Asymptotic
 Confidence Level: 0.95

Output

Terms in the Model : 4
 # Term(s) Dropped: 0
 Total Number of Records: 1512
 Number of Records Rejected: 0
 # Groups: 34

Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|----------|----|----------|
| Deviance | 890.746 | 30 | 2.979E-9 |
| Likelihood Ratio | 1205.331 | 4 | 0 |

Parameter Estimates:

| Model Terms | Point Estimate | | | Confidence Interval and p-value for Beta | | |
|-------------|----------------|--------|----------|--|--------|-------------------|
| | Type | Beta | SE(Beta) | 95% CI | | 2*1-Sided p-value |
| | | | | Lower | Upper | |
| %const | MLE | -2.762 | 0.684 | -4.102 | -1.422 | 5.36E-5 |
| Dose | MLE | 1.08 | 0.442 | 0.214 | 1.946 | 0.014 |
| Impl | MLE | 0.07 | 0.048 | -0.023 | 0.163 | 0.141 |
| Dose*Impl | MLE | -0.044 | 0.031 | -0.104 | 0.016 | 0.151 |

The third section is **Summary Statistics**. This section displays the deviance and its degrees of freedom, and the likelihood ratio statistic and degrees of freedom for testing the null hypothesis that the response probability of each observation is 0.5, i.e., all the model parameters, **including the constant term**, are simultaneously 0. The likelihood ratio statistic may be used to test for overall significance of the model. For the present example, the output displays a value of 890.7457 on 30 df for the deviance, and a value of 1205.3314 on 4 df for the likelihood ratio statistic, thereby rejecting the null

91 Analysis-Binary Regression Analysis

hypothesis that all the parameters of the model are 0.

The last section, **Parameter Estimates**, displays the **Model Term**, **Point Estimate** and the **Confidence Interval and p-value for Beta**. The **Model Terms** show there are three covariates, **Dose**, **Impl** and **Dose*Impl**, in the model. The next three columns (the **Point Estimates**) show **MLE** as **Type**, estimates and standard error of Beta's. For **Dose**, the estimate of Beta is 1.08. For **Impl**, the estimate of Beta is 0.07. For **Dose*Impl**, the estimate of Beta is -0.044. The next four columns show the inference type, confidence interval of Beta, and the p-value (2*1-sided) for testing Beta = 0. Here the p-value for **Dose** is 0.014.

Analysis Steps: Probit Regression - Test Multiple Hypothesis Model

Suppose you are interested in a simultaneous test that the parameters corresponding to both **Impl** and **Dose*Impl** in the previously specified model are equal to 0.

1. Invoke the **Analysis Input** tab from the status bar below. In the Input dialog, click the **Test** option. Use the Toggle **Selected for Estimation or Testing** button in the **Model Terms** box to select **Impl** and **Dose*Impl** for testing and deselect **Dose**.

Data Set: Devtox.cydx

Main
Options

Variables

- Id
- Dose
- Weight
- Malr
- Sex
- Impl
- LittSz

Toggle Factor On/Off

Model

Response: Death Response Value: 1

| | Model Terms | Test |
|--------|-------------|----------------|
| >> | %const | %const(No) |
| << | Dose | Dose(No) |
| a*b >> | Impl | Impl(Yes) |
| | Dose*Impl | Dose*Impl(Yes) |

Toggle Model Terms Selected for Testing Yes/No

Weightage:

Estimate Model or Test Multiple Hypotheses

Estimate Test

Output

Postfit Results Classification Table

2. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binary Regression: Probit Model

Input Parameters

Data File: Devtox.cydx
 Model: Death(Response = 1)=%const+Dose+Impl+Dose*Impl
 Link Type: Probit
 Analysis Type: Test:Asymptotic
 Confidence Level: 0.95

Output

Terms in the Model : 4
 # Term(s) Dropped: 0
 Total Number of Records: 1512
 Number of Records Rejected: 0
 # Groups: 34

Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|----------|----|----------|
| Deviance | 890.746 | 30 | 2.979E-9 |
| Likelihood Ratio | 1205.331 | 4 | 0 |

Parameter Estimates:

| Model Terms | Point Estimate | | | Confidence Interval and p-value for Beta | | |
|-------------|----------------|--------|----------|--|--------|-------------------|
| | Type | Beta | SE(Beta) | 95% CI | | 2*1-Sided p-value |
| | | | | Lower | Upper | |
| %const | MLE | -2.762 | 0.684 | -4.102 | -1.422 | 5.36E-5 |
| Dose | MLE | 1.08 | 0.442 | 0.214 | 1.946 | 0.014 |
| Impl | MLE | 0.07 | 0.048 | -0.023 | 0.163 | 0.141 |
| Dose*Impl | MLE | -0.044 | 0.031 | -0.104 | 0.016 | 0.151 |

Hypothesis Testing:

Test: <Impl=Dose*Impl=0>

| Type of Test | Statistics | DF | p-value |
|------------------|------------|----|---------|
| Score | 2.241 | 2 | 0.326 |
| Likelihood Ratio | 2.235 | 2 | 0.327 |
| Wald | 2.238 | 2 | 0.327 |

91 Analysis-Binary Regression Analysis

The title **Hypothesis Testing Tests** $\langle Impl = Dose * Impl = 0 \rangle$ appears near the bottom of the Test results worksheet. Below that, you can see the results of three asymptotic tests (based on the Scores, likelihood ratio, and Wald statistics, respectively) of the null hypothesis that the regression parameters corresponding to **Impl** and **Dose*Impl** are both 0. Since two parameters are being tested, this is a 2 degree of freedom test. All three tests are two-sided. Notice that the test statistics and p-values based on the Score test (2.241 and 0.326) are very similar to those based on the Likelihood Ratio and Wald tests. The p-values are quite large, indicating that we cannot reject the null hypothesis that the parameters corresponding to both **Impl** and **Dose*Impl** are equal to 0.

Since we are only interested in a positive trend, it is appropriate to perform 1-sided tests.

1. In the **Options** tab change the **Output p- value** to **One-sided**.
2. Click **OK** to estimate the model once more. The output is displayed in the main window.

Analysis: Binary Regression: Probit Model

Input Parameters

Data File: Devtox.cyx
Model: Death(Response = 1)=%const+Dose+Impl+Dose*Impl
Link Type: Probit
Analysis Type: Test:Asymptotic
Confidence Level: 0.95

Output

Terms in the Model : 4
Term(s) Dropped: 0
Total Number of Records: 1512
Number of Records Rejected: 0
Groups: 34
Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|----------|----|----------|
| Deviance | 890.746 | 30 | 2.979E-9 |
| Likelihood Ratio | 1205.331 | 4 | 0 |

Parameter Estimates:

| Model Terms | Point Estimate | | | Confidence Interval and p-value for Beta | | |
|-------------|----------------|--------|----------|--|--------|-----------------|
| | Type | Beta | SE(Beta) | 95% CI | | 1-Sided p-value |
| | | | | Lower | Upper | |
| %const | MLE | -2.762 | 0.684 | -INF | -1.638 | 2.68E-5 |
| Dose | MLE | 1.08 | 0.442 | 0.354 | INF | 0.007 |
| Impl | MLE | 0.07 | 0.048 | -0.008 | INF | 0.071 |
| Dose*Impl | MLE | -0.044 | 0.031 | -INF | 0.006 | 0.075 |

Hypothesis Testing:

Test: <Impl=Dose*Impl=0>

| Type of Test | Statistics | DF | p-value |
|------------------|------------|----|---------|
| Score | 2.241 | 2 | 0.326 |
| Likelihood Ratio | 2.235 | 2 | 0.327 |
| Wald | 2.238 | 2 | 0.327 |

Note that since we specified **One-sided** p-values, **East** reports 1-sided p-values as well as corresponding 1-sided confidence bounds. Since the remaining analyses will all be 2-sided, re-set the option for **output p-value** to **Two-sided** in the **Options** tab.

Post-Fit Analysis Now that we have fit a model to the data, let us obtain regression diagnostics to evaluate the fit. To do so, invoke **Analysis Inputs** from the lower status bar. Select the **Postfit Results** check box. Click **OK** to run the analysis. In the Library, there will be two more nodes named **Regression Diagnostics1** and **ROC Curve-1**

91 Analysis-Binary Regression Analysis

which essentially forms the post-fit analysis. The main output is as follows:

Analysis: Binary Regression: Probit Model

Input Parameters

Data File: Devtox.cydx
 Model: Death(Response = 1)=%const+Dose+Impl+Dose*Impl
 Link Type: Probit
 Analysis Type: Test:Asymptotic
 Confidence Level: 0.95

Output

Terms in the Model : 4
 # Term(s) Dropped: 0
 Total Number of Records: 1512
 Number of Records Rejected: 0
 # Groups: 34

Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|----------|----|----------|
| Deviance | 890.746 | 30 | 2.979E-9 |
| Likelihood Ratio | 1205.331 | 4 | 0 |

Parameter Estimates:

| Model Terms | Point Estimate | | | Confidence Interval and p-value for Beta | | |
|-------------|----------------|--------|----------|--|--------|-------------------|
| | Type | Beta | SE(Beta) | 95% CI | | 2*1-Sided p-value |
| | | | | Lower | Upper | |
| %const | MLE | -2.762 | 0.684 | -4.102 | -1.422 | 5.36E-5 |
| Dose | MLE | 1.08 | 0.442 | 0.214 | 1.946 | 0.014 |
| Impl | MLE | 0.07 | 0.048 | -0.023 | 0.163 | 0.141 |
| Dose*Impl | MLE | -0.044 | 0.031 | -0.104 | 0.016 | 0.151 |

Hypothesis Testing:

Test: <Impl=Dose*Impl=0>

| Type of Test | Statistics | DF | p-value |
|------------------|------------|----|---------|
| Score | 2.241 | 2 | 0.326 |
| Likelihood Ratio | 2.235 | 2 | 0.327 |
| Wald | 2.238 | 2 | 0.327 |

Hosmer-Lemeshow Test:

| Statistics | Value | DF | p-value |
|-----------------|--------|----|---------|
| Hosmer-Lemeshow | 15.375 | 7 | 0.031 |

| Hosmer-Lemeshow Percentile of Risk | Response | | Non-Response | | Total |
|------------------------------------|------------|----------------|--------------|-----------------|-------|
| | Observed | Expected | Observed | Expected | |
| 1 | 9 | 7.16 | 203 | 204.84 | 212 |
| 2 | 10 | 11.058 | 229 | 227.942 | 239 |
| 3 | 10 | 9.977 | 165 | 165.023 | 175 |
| 4 | 16 | 11.162 | 146 | 150.838 | 162 |
| 5 | 11 | 16.587 | 180 | 174.413 | 191 |
| 6 | 12 | 17.089 | 158 | 152.911 | 170 |
| 7 | 27 | 32.529 | 148 | 142.471 | 175 |
| 8 | 44 | 31.126 | 111 | 123.874 | 155 |
| 9 | 5 | 7.348 | 28 | 25.652 | 33 |
| Total | 144 | 144.036 | 1368 | 1367.964 | |

The post-fit output in the output sheet titled 'Regression Diagnostics' is as follows:

91 Analysis-Binary Regression Analysis

Analysis: Binary Regression: Probit Model

Regression Diagnostics

| Sequence | GrpSize | ObsvResp | ExptResp | ProbResp | Hat_Diag | StdPearson | Deviance | Std_Deviance |
|----------|---------|----------|----------|----------|----------|------------|----------|--------------|
| 1 | 36 | 2 | 0.985 | 0.027 | 0.017 | 1.046 | 0.912 | 0.92 |
| 2 | 78 | 2 | 2.503 | 0.032 | 0.025 | -0.327 | -0.335 | -0.339 |
| 3 | 98 | 5 | 3.672 | 0.037 | 0.024 | 0.715 | 0.671 | 0.679 |
| 4 | 120 | 6 | 5.228 | 0.044 | 0.039 | 0.352 | 0.338 | 0.345 |
| 5 | 22 | 1 | 0.993 | 0.045 | 0.015 | 0.007 | 0.007 | 0.007 |
| 6 | 1 | 0 | 0.049 | 0.049 | 0.005 | -0.227 | -0.316 | -0.317 |
| 7 | 96 | 3 | 4.789 | 0.05 | 0.043 | -0.857 | -0.898 | -0.918 |
| 8 | 32 | 1 | 1.614 | 0.05 | 0.021 | -0.501 | -0.532 | -0.537 |
| 9 | 39 | 3 | 2.146 | 0.055 | 0.011 | 0.603 | 0.568 | 0.571 |
| 10 | 34 | 3 | 1.977 | 0.058 | 0.044 | 0.767 | 0.699 | 0.716 |
| 11 | 70 | 3 | 4.24 | 0.061 | 0.014 | -0.626 | -0.654 | -0.659 |
| 12 | 6 | 0 | 0.381 | 0.063 | 0.015 | -0.642 | -0.887 | -0.893 |
| 13 | 120 | 14 | 7.987 | 0.067 | 0.031 | 2.238 | 2.005 | 2.037 |
| 14 | 9 | 1 | 0.663 | 0.074 | 0.011 | 0.432 | 0.402 | 0.404 |
| 15 | 10 | 0 | 0.774 | 0.077 | 0.009 | -0.92 | -1.269 | -1.275 |
| 16 | 17 | 1 | 1.358 | 0.08 | 0.016 | -0.323 | -0.335 | -0.337 |
| 17 | 11 | 1 | 0.893 | 0.081 | 0.007 | 0.118 | 0.116 | 0.116 |
| 18 | 84 | 8 | 7.156 | 0.085 | 0.032 | 0.335 | 0.324 | 0.329 |
| 19 | 18 | 0 | 1.57 | 0.087 | 0.03 | -1.332 | -1.813 | -1.84 |
| 20 | 78 | 2 | 6.966 | 0.089 | 0.018 | -1.99 | -2.298 | -2.319 |
| 21 | 42 | 4 | 3.93 | 0.094 | 0.008 | 0.037 | 0.037 | 0.037 |
| 22 | 45 | 4 | 4.409 | 0.098 | 0.011 | -0.206 | -0.208 | -0.209 |

ROC Curve and Classification Table The use of **ROC Curve and Classification table** for Probit model is similar to what is described in sections 91.2 and 91.2 for **Logistic Regression** model. The ROC output in the output sheet titled 'ROC Curve-1' is as follows:

Analysis: Binary Regression: Probit Model

ROC-Curve

| Corr_events | Corr_noevents | Incorr_events | Incorr_noevents | Sensitivity | 1-Specificity |
|-------------|---------------|---------------|-----------------|-------------|---------------|
| 144 | 0 | 1368 | 0 | 1 | 1 |
| 142 | 34 | 1334 | 2 | 0.986 | 0.975 |
| 140 | 110 | 1258 | 4 | 0.972 | 0.92 |
| 135 | 203 | 1165 | 9 | 0.938 | 0.852 |
| 129 | 317 | 1051 | 15 | 0.896 | 0.768 |
| 128 | 338 | 1030 | 16 | 0.889 | 0.753 |
| 128 | 339 | 1029 | 16 | 0.889 | 0.752 |
| 125 | 432 | 936 | 19 | 0.868 | 0.684 |
| 124 | 463 | 905 | 20 | 0.861 | 0.662 |
| 121 | 499 | 869 | 23 | 0.84 | 0.635 |
| 118 | 530 | 838 | 26 | 0.819 | 0.613 |
| 115 | 597 | 771 | 29 | 0.799 | 0.564 |
| 115 | 603 | 765 | 29 | 0.799 | 0.559 |
| 101 | 709 | 659 | 43 | 0.701 | 0.482 |
| 100 | 717 | 651 | 44 | 0.694 | 0.476 |
| 100 | 727 | 641 | 44 | 0.694 | 0.469 |
| 99 | 743 | 625 | 45 | 0.688 | 0.457 |
| 98 | 753 | 615 | 46 | 0.681 | 0.45 |
| 90 | 829 | 539 | 54 | 0.625 | 0.394 |
| 90 | 847 | 521 | 54 | 0.625 | 0.381 |
| 88 | 923 | 445 | 56 | 0.611 | 0.325 |
| 84 | 961 | 407 | 60 | 0.583 | 0.298 |
| 80 | 1002 | 366 | 64 | 0.556 | 0.268 |
| 78 | 1032 | 336 | 66 | 0.542 | 0.246 |
| 76 | 1081 | 287 | 68 | 0.528 | 0.21 |
| 61 | 1084 | 284 | 83 | 0.424 | 0.208 |
| 59 | 1116 | 252 | 85 | 0.41 | 0.184 |

91.6 Complementary Log Log Model

The complementary log-log model also falls within the generalized linear model framework. The model uses the complementary log-log function to link the probability

91 Analysis-Binary Regression Analysis

of response to a linear combination of the covariates. Using the notation from the previous section, the complementary log-log model is

$$\log[-\log(1 - \pi_i)] = \beta_0 + \beta'_b \mathbf{x}_i.$$

The model gets its name from the fact that the log-log function is applied to $(1 - \pi)$, or the complement of the probability of a success. Thus, the model applies a log-log link to the probability that $Y_i = 0$. Unlike the logistic and probit models, the complementary log-log model implies that the probability of a response is asymmetric around 0.5. That is, the model specifies that this probability approaches 0 relatively slowly but approaches 1 relatively quickly. See Agresti (2002; Section 6.6.4) for graphical comparison of these rates in relation to the logit and probit models. As a result, the model will fit data that exhibit asymmetric rates of change in the probability of success better than the corresponding logistic and probit models, and is preferable in such cases.

Example: Complementary Log Log Model

Dataset: Seropos.cydx

Data Description

Consider the Serological Malaria data that have been discussed by Draper, Voller, and Carpenter (1972), and by Collett (1991). A serologic survey was carried out in 1971 in two areas of Amazonas, Brazil. An indirect fluorescent antibody test was used to detect the presence of antibodies to a malarial parasite in the villagers. The data reproduced in Table below refers to the proportion of individuals in each of seven age groups who were found to be seropositive.

Table: Seropositivity rates for villagers in Amozonas, Brazil in 1971

| Age group | Mid-point of age range in years | Proportion seropositive |
|-------------|---------------------------------|-------------------------|
| 0-11 months | 0.5 | 3/10 (30.00%) |
| 1-2 years | 1.5 | 1/10 (10.00%) |
| 2-4 years | 3.0 | 5/29 (17.24%) |
| 5-9 years | 7.0 | 39/69 (56.52%) |
| 10-14 years | 12.0 | 31/51 (60.78%) |
| 15-19 years | 17.0 | 8/15 (53.33%) |
| ≥ 20 years | 30.0 | 91/108 (84.26%) |

Analysis Steps: Clog Log Regression - Estimate Model

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

Analysis > (Discrete) Regression > (Parallel Design) Clog Log Regression

3. In the **Main** tab, select **AgeGroup** as the variable to be included as the **Model Terms**. Select **Seropositive** as the **Response** variable. Enter 1 as **Response Value**. Select **Frequency** as the **Weightage** variable. Click the **Estimate** option.
4. Click **OK** to start the analysis. The output displayed in the main window.

Analysis: Binary Regression: Complementary Log Log Model

Input Parameters

Data File: Seropos.cydx
Model: Seropositive(Response = 1)=%const+AgeGroup
Link Type: Complementary Log Log
Weight Variable: Frequency
Analysis Type: Estimate:Asymptotic
Confidence Level: 0.95

Output

Terms in the Model : 2
Term(s) Dropped: 0
Total Number of Records: 292
Number of Records Rejected: 0
Groups: 7
Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|---------|----|-----------|
| Deviance | 338.362 | 5 | 1.382E-9 |
| Likelihood Ratio | 52.927 | 2 | 4.587E-10 |

Parameter Estimates:

| Model Terms | Point Estimate | | | Confidence Interval and p-value for Beta | | |
|-------------|----------------|--------|----------|--|--------|-------------------|
| | Type | Beta | SE(Beta) | 95% CI | | 2*1-Sided p-value |
| | | | | Lower | Upper | |
| %const | MLE | -0.881 | 0.156 | -1.187 | -0.576 | 1.78E-8 |
| AgeGroup | MLE | 0.051 | 0.007 | 0.037 | 0.065 | 3.131E-12 |

91 Analysis-Binary Regression Analysis

The third section is **Summary Statistics**. This section displays the deviance and its degrees of freedom, and the likelihood ratio statistic and degrees of freedom for testing the null hypothesis that the response probability of each observation is 0.5, i.e., all the model parameters, **including the constant term**, are simultaneously 0. The likelihood ratio statistic has a chi-squared distribution under the null hypothesis and can be used to test for overall significance of the model. For the present example, the output displays a value of 338.362 on 5 df for the deviance, and a value of 52.927 on 2 df for the likelihood ratio statistic, thereby rejecting the null hypothesis that all the parameters of the model are 0.

The last section, **Parameter Estimates**, displays the **Model Term, Point Estimate** and the **Confidence Interval and p-value for Beta**. The **Model Term** shows one covariate **AgeGroup** in the model. The next three columns the **Point Estimate**) show **MLE as Type**, estimates and standard error of **Betas**. For **AgeGroup**, the estimate of **Beta** is 0.0511.

The next four columns shows the inference type, confidence interval of **Beta** and the p-value(2*1-sided) for testing **Beta** equal to 0. Here the p-value for **AgeGroup** is < 0.0001.

A node **Analysis: Binary Regression: Complementary Log Log Model** is created in the Library.

Analysis Steps: Clog Log Regression - Test Multiple Hypotheses

Suppose you wish to test the null hypothesis that the parameter corresponding to **AgeGroup** in the model is equal to 0.

1. Click the **Analysis Input** tab from the status bar below. In the Input dialog, select the **Test** option. Use the **Toggle Model term Selected for Testing** button

in the **Model Terms** box to select **AgeGroup** for testing.

Analysis: Discrete Endpoint: Regression Model - Parallel Design - CLog Log Regression

Data Set: Seropos.cydx

Main Options

Variables

AgeGroup

Toggle Factor On/Off

Model

Response: Seropositive Response Value: 1

| Model Terms | Test |
|-------------|---------------|
| %const | %const(No) |
| AgeGroup | AgeGroup(Yes) |

>> << a*b >>

Toggle Model Terms Selected for Testing Yes/No

Weightage: Frequency

Estimate Model or Test Multiple Hypotheses

Estimate Test

Output

Postfit Results Classification Table

91 Analysis-Binary Regression Analysis

2. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Binary Regression: Complementary Log Log Model

Input Parameters

Data File: Seropos.cydx
Model: Seropositive(Response = 1)=%const+AgeGroup
Link Type: Complementary Log Log
Weight Variable: Frequency
Analysis Type: Test:Asymptotic
Confidence Level: 0.95

Output

Terms in the Model : 2
Term(s) Dropped: 0
Total Number of Records: 292
Number of Records Rejected: 0
Groups: 7
Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|---------|----|-----------|
| Deviance | 338.362 | 5 | 1.382E-9 |
| Likelihood Ratio | 52.927 | 2 | 4.587E-10 |

Parameter Estimates:

| Model Terms | Point Estimate | | | Confidence Interval and p-value for Beta | | |
|-------------|----------------|--------|----------|--|--------|-------------------|
| | Type | Beta | SE(Beta) | 95% CI | | 2*1-Sided p-value |
| | | | | Lower | Upper | |
| %const | MLE | -0.881 | 0.156 | -1.187 | -0.576 | 1.78E-8 |
| AgeGroup | MLE | 0.051 | 0.007 | 0.037 | 0.065 | 3.131E-12 |

Hypothesis Testing:

Test: <AgeGroup=0>

| Type of Test | Statistics | DF | p-value |
|------------------|------------|----|-----------|
| Score | 53.902 | 1 | 5.999E-10 |
| Likelihood Ratio | 52.294 | 1 | 4.62E-10 |
| Wald | 50.242 | 1 | 6.389E-10 |

The title **Hypothesis testing Tests** $\langle \text{AgeGroup} = 0 \rangle$ appears near the bottom of the **Test** worksheet. Below that, you can see the results of three tests: Score, likelihood ratio, and Wald respectively of the null hypothesis that the regression parameter corresponding to **AgeGroup** is 0. Since a single parameter is being tested, this is a 1 degree of freedom test. All three tests are 2-sided. See Agresti (2002). Notice that all three p-values are very small (< 0.0001) indicating that we reject the null hypothesis that the parameter corresponding to **AgeGroup** is 0.

Estimation Results The estimation output currently displays the point estimates, confidence intervals, and two-sided p-values for the parameters corresponding to **AgeGroup**. These statistics were computed by the maximum likelihood method (See Agresti). Let us look at the individual items computed as estimation output. Specifically, look at the output corresponding to **AgeGroup** in the **Estimate** worksheet. The MLE for the β coefficient, its standard error, its confidence interval, and the p-value are all displayed.

Post-Fit Analysis Now that we have fit a model to the data, let us obtain regression diagnostics to evaluate the fit. Click the **Input Parameters** tab from the status bar below. Select the **Postfit Results** check box. Click **OK** to start the analysis. The output is displayed in the main window. In the Library along with the node **Analysis: Binary Regression:Complementary Log Log Model2**, there is two more nodes named **Regression Diagnostics1** and **ROC Curve-1** which essentially form the post-fit

91 Analysis-Binary Regression Analysis

analysis. The main output is as follows:

Analysis: Binary Regression: Complementary Log Log Model

Input Parameters

Data File: Seropos.cyx
 Model: Seropositive(Response = 1)=%const+AgeGroup
 Link Type: Complementary Log Log
 Weight Variable: Frequency
 Analysis Type: Test:Asymptotic
 Confidence Level: 0.95

Output

Terms in the Model : 2
 # Term(s) Dropped: 0
 Total Number of Records: 292
 Number of Records Rejected: 0
 # Groups: 7

Summary Statistics:

| Statistics | Value | DF | p-value |
|------------------|---------|----|-----------|
| Deviance | 338.362 | 5 | 1.382E-9 |
| Likelihood Ratio | 52.927 | 2 | 4.587E-10 |

Parameter Estimates:

| Model Terms | Point Estimate | | | Confidence Interval and p-value for Beta | | |
|-------------|----------------|--------|----------|--|--------|-------------------|
| | Type | Beta | SE(Beta) | 95% CI | | 2*1-Sided p-value |
| | | | | Lower | Upper | |
| %const | MLE | -0.881 | 0.156 | -1.187 | -0.576 | 1.78E-8 |
| AgeGroup | MLE | 0.051 | 0.007 | 0.037 | 0.065 | 3.131E-12 |

Hypothesis Testing:

Test: <AgeGroup=0>

| Type of Test | Statistics | DF | p-value |
|------------------|------------|----|-----------|
| Score | 53.902 | 1 | 5.999E-10 |
| Likelihood Ratio | 52.294 | 1 | 4.62E-10 |
| Wald | 50.242 | 1 | 6.389E-10 |

The post-fit output in the output sheet titled ‘Regression Diagnostics’ is as follows:

Analysis: Binary Regression: Complementary Log Log Model

Regression Diagnostics

| Sequence | GrpSize | ObsvResp | ExptResp | ProbResp | Hat_Diag | StdPearson | Deviance | Std_Deviance | DeltaBeta | %const | AgeGroup |
|----------|---------|----------|----------|----------|----------|------------|----------|--------------|-----------|--------|----------|
| 1 | 10 | 3 | 3.462 | 0.346 | 0.053 | -0.315 | -0.31 | -0.319 | 0.006 | 1 | 0.5 |
| 2 | 10 | 1 | 3.606 | 0.361 | 0.05 | -1.76 | -1.894 | -1.943 | 0.162 | 1 | 1.5 |
| 3 | 29 | 5 | 11.105 | 0.383 | 0.13 | -2.501 | -2.472 | -2.65 | 0.936 | 1 | 3 |
| 4 | 69 | 39 | 30.833 | 0.447 | 0.229 | 2.252 | 1.97 | 2.244 | 1.505 | 1 | 7 |
| 5 | 51 | 31 | 27.252 | 0.534 | 0.111 | 1.116 | 1.058 | 1.122 | 0.156 | 1 | 12 |
| 6 | 15 | 8 | 9.407 | 0.627 | 0.024 | -0.76 | -0.741 | -0.75 | 0.014 | 1 | 17 |
| 7 | 108 | 91 | 92.101 | 0.853 | 0.183 | -0.331 | -0.296 | -0.328 | 0.025 | 1 | 30 |

Note the following items of information:

- 7 records of data, corresponding to the 7 groups;
- the group size, observed response, and expected response, in each group;
- the Pearson residual for each group;
- the Pregibon (1981) $\Delta\beta$ leverage value for each group;
- the value of the covariate vector for each group.

ROC Curve and Classification Table The use of **ROC Curve and Classification table** for Complementary Log Log model is similar to what is described in sections 91.2 and 91.2 for Logistic Regression model.

Analysis: Binary Regression: Complementary Log Log Model

ROC-Curve

| Corr_events | Corr_noevents | Incorr_events | Incorr_noevents | Sensitivity | 1-Specificity |
|-------------|---------------|---------------|-----------------|-------------|---------------|
| 178 | 0 | 114 | 0 | 1 | 1 |
| 175 | 7 | 107 | 3 | 0.983 | 0.939 |
| 174 | 16 | 98 | 4 | 0.978 | 0.86 |
| 169 | 40 | 74 | 9 | 0.949 | 0.649 |
| 130 | 70 | 44 | 48 | 0.73 | 0.386 |
| 99 | 90 | 24 | 79 | 0.556 | 0.211 |
| 91 | 97 | 17 | 87 | 0.511 | 0.149 |

92 *Analysis- Multiple Comparison Procedures for Binary Data*

It is often the case that multiple objectives are to be addressed in one single trial. These objectives are formulated into a family of hypotheses. Type I error rate is inflated when one considers the inferences together as a family. Failure to compensate for multiplicities can have adverse consequences. For example, a drug could be approved when actually it is not better than Placebo.

Probability of making at least one type I error is known as family wise error rate (FWER). Multiple comparison (MC) procedures provide a guard against inflation of type I error due to multiple testing. All the MC procedures available in **East** strongly control FWER. Strong control of FWER refers to preserving the probability of incorrectly claiming at least one null hypothesis. To contrast strong control with weak control of FWER, the latter controls the FWER under the assumption that all hypotheses are true.

East supports several p-value based MC procedures for binary data. We have seen how to simulate data under different MC procedures with specified response rates and types of variance such as pooled or unpooled in chapter 27. In this chapter we explain how to analyze binary data with different MC procedures available in **East**. For MC procedures in **East**, we can either provide the dataset containing observations under each arm or the raw p-values to obtain the adjusted p-values.

92.1 Available Procedures **East** supports following MC procedures based on binary endpoint.

| PROCEDURE | REFERENCE |
|---------------------|----------------------------------|
| Bonferroni | Bonferroni CE. (1935) |
| Sidak | Sidak Z. (1967) |
| Weighted Bonferroni | Benjamini Y, Hochberg Y. (1997) |
| Holm’s Step Down | Holm S. (1979) |
| Hochberg’s Step Up | Hochberg Y. (1988) |
| Hommel’s Step Up | Hommel G. (1988) |
| Fixed Sequence | Westfall PH, Krishen A. (2001) |
| Fallback | Wiens B. (2003) |

East supports three p-value based single step MC procedures - Bonferroni procedure, Sidak procedure and Weighted Bonferroni procedure. Whereas, Hocheberg Procedure and Holm procedure are available as Data-driven step-up MC procedures.

Fixed-sequence stepwise procedure and fallback procedure are also part of **East** multiple comparison procedures for binary end points.

92.2 Single step MC procedures

Example: Bonferroni procedure

Example: Sidak Procedure

Example: Weighted Bonferroni Procedure

East supports three p-value based single step MC procedures. These are:

- Bonferroni procedure
- Sidak procedure and
- Weighted Bonferroni procedure

For the Bonferroni procedure, H_i is rejected if $p_i < \frac{\alpha}{k-1}$ and the adjusted p-value is given as $\min(1, (k-1)p_i)$.

For the Sidak procedure, H_i is rejected if $p_i < 1 - (1 - \alpha)^{\frac{1}{k-1}}$ and the adjusted p-value is given as $1 - (1 - p_i)^{k-1}$.

For the weighted Bonferroni procedure, H_i is rejected if $p_i < w_i\alpha$ and the adjusted p-value is given as $\min(1, \frac{p_i}{w_i})$. Here w_i denotes the proportion of α allocated to the H_i such that $\sum_{i=1}^{k-1} w_i = 1$. Note that, if $w_i = \frac{1}{k-1}$, then the Bonferroni procedure is reduced to the regular Bonferroni procedure.

Example: Bonferroni procedure

Dataset: HIV-study.cyx

Data Description

Throughout this chapter we will use the data from a dose finding HIV Study. It was a randomized, double-blind, parallel-group, placebo-controlled, multi-center trial to assess the efficacy and safety of 125mg(L), 250 mg(M), and 500 mg(H) orally twice daily of a new drug for a treatment of HIV associated diarrhea. The primary efficacy endpoint is clinical response, defined as two or less watery bowel movements per week, during at least two of the four weeks of the 4-week efficacy assessment period. The efficacy is evaluated by comparing the proportion of responders in the placebo group to the proportion of responders in the three treatment groups at a 1-sided alpha of 0.025.

The data set consists of two variables. The first variable, **Trt_group**, takes four values as "P", "L", "M", and "H". The "P" value represents the placebo group, "L" the low dose (125 mg) group, "M" the middle dose (250 mg) group, and "H" the high dose (500 mg) group. The second variable, **response**, is a binary indicator of whether or not each subject was a responder (1 represents a responder, 0 represents a non-responder).

92 *Analysis- Multiple Comparison Procedures for Binary Data*

Purpose of the Analysis:

To analyze the data of the dose finding HIV trial using Bonferroni procedure.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

Analysis > (Discrete) Many Samples > (Multiple Comparisons) Pairwise Comparison to Control - Differences of Proportions

3. In the **Main** tab, select the **raw data** option. In the ensuing box with label **Treatment Variable and its Control Arm**, select **Trt_group** and select the **P** option next to it. Select **response** as the **Select Response Variable** with a **Response Value** of **1**. Under the dropdown box for selecting the response variable there are two options, pooled variance and un-pooled variance. For this example, select the **Pooled Variance** option. If **Pooled Variance** is selected, the software will use the pooled variance estimate in calculating the standard error of the test statistics. If **Un-pooled Variance** is selected, the software will use the un-pooled variance estimate in calculating the standard error of the test statistics. The technical details on variance estimates are provided in the technical appendix **H**. Select **Bonferroni** from the **Select MCP** drop-down list.

Analysis: Discrete Endpoint: Many-Sample Test - Multi-Arm Design - Pairwise Comparisons to Control - Differences of Proportions

Data Set: HIV-study.cyd

Main Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
Select Treatment Variable:
Trt_group
 H
 L
 M
 P

Response Variable
Select Response Variable: response
Select Value: 1
 Pooled Variance Un-pooled Variance

Frequency: [dropdown]

Multiple Comparison Procedure
Select MCP: Bonferroni

4. In the **Advanced** tab leave the **By Variable** input boxes blank. Enter **0.95** for

92 Analysis- Multiple Comparison Procedures for Binary Data

Confidence Level and select **Right-tail** for **Rejection Region**.

Analysis: Discrete Endpoint Many-Sample Test - Multi-Arm Design - Pairwise Comparisons to Control - Differences of Proportions

Data Set: HIV-study.cyd

Main
Advanced

By Variable 1: Confidence Level:

By Variable 2: Rejection Region:

5. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions

Hypothesis

$H_i : \pi_i - \pi_0 \leq 0$ Vs. $K_i : \pi_i - \pi_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: HIV-study.cyd
 Treatment Variable: Trt_group
 Response Variable: response(1)
 Control Arm: P
 Variance Type: Pooled
 Multiple Comparison Procedure: Bonferroni
 Confidence Level: 0.95
 Test Type: z-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Proportion | Std. Err. of Diff. of Proportions | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Proportions(One sided) | |
|------|-------------|------------|-----------------------------------|----------------|----------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 125 | 0.344 | NA | NA | NA | NA | NA | NA |
| H | 125 | 0.544 | 0.063 | 3.182 | 7.306E-4 | 0.002 | 0.097 | 1 |
| L | 125 | 0.392 | 0.061 | 0.787 | 0.216 | 0.647 | -0.052 | 1 |
| M | 125 | 0.44 | 0.062 | 1.555 | 0.06 | 0.18 | -0.006 | 1 |

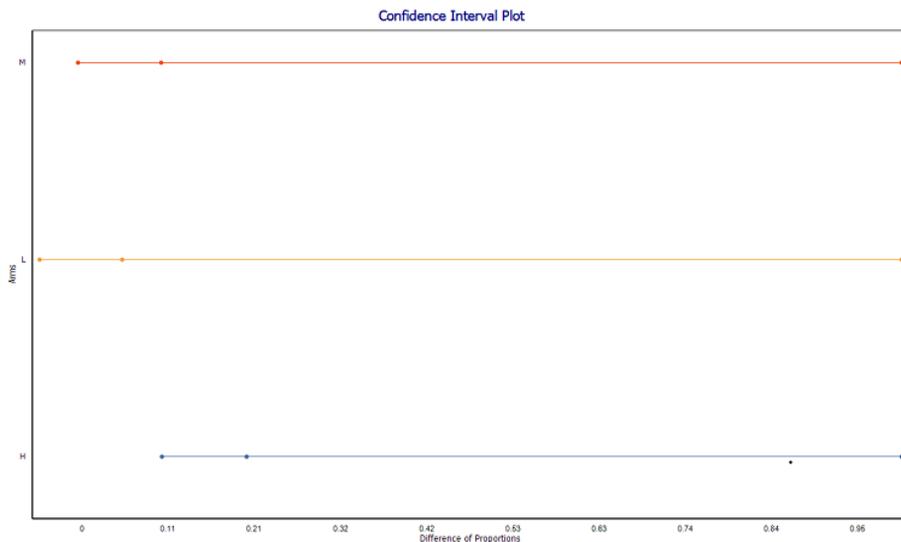
Adjusted Global p-value: 0.002
 Total Number of Records: 500
 Number of Records Rejected: 0
 Total Number of Arms (k): 4

The **Output** section gives us a table with our results. The sample size of each group and the sample mean of our response variable (response) are given. The Std. Err. of Diff. of Means column gives us the standard error of the difference of means (not the standard error of the mean) for comparing that specific treatment to placebo. The next column gives us the test statistic. The two columns after that give us the naive and adjusted (using Bonferroni's procedure) p-values. The technical appendix **H** contains the technical details on Bonferroni's procedure. You can refer to it to see how the

p-values are calculated. From these results we can see that after adjusting for multiplicity there is a significant difference, at the $\alpha = 0.05$ level, in the proportion of clinical response between placebo and the high dose (adjusted p-value = 0.002). We did not find any evidence of a difference between placebo and the low dose (adjusted p-value = 0.647), and placebo group and the middle dose (adjusted p-value = 0.180). Also, the naive p-values are all less than or equal to the adjusted p-values, as expected.

The final two columns of the table give us the lower and upper bounds for the 95% one sided confidence intervals. The last section shows us the adjusted global p-value, total number of records, number of records rejected, and total number of arms.

In **Library**, there would also be another node labeled **Confidence Interval Plot1**. Double click this node to display a Confidence Interval plot.



Example: Sidak Procedure

Dataset: HIV-study.cyx as described in Section 92.2.

92 Analysis- Multiple Comparison Procedures for Binary Data

Purpose of the Analysis:

To analyze the data of the dose finding HIV trial using Sidak procedure.

Analysis Steps:

1. Click **Analysis Inputs** tab on the status bar below. This will display several input fields associated with multiple comparison test in the main window.
2. In the **Main** tab, select **Sidak** in the **Select MCP** drop-down. Leave all other parameters as selected for the Bonferroni procedure above.

Analysis: Discrete Endpoint Many-Sample Test - Multi-Arm Design - Pairwise Comparisons to Control - Differences of Proportions

Data Set: HIV-study.cyd

Main Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
Select Treatment Variable:
Trt_group

H
 L
 M
 P

Response Variable
Select Response Variable: response
Select Value: 1

Pooled Variance Un-pooled Variance

Frequency:

Multiple Comparison Procedure
Select MCP: Sidak

3. Click **OK** to start the analysis. The output, as shown below, is displayed in the

Output Preview Area.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions

Hypothesis

$H_i : \pi_i - \pi_0 \leq 0$ Vs. $K_i : \pi_i - \pi_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: HIV-study.cyd
 Treatment Variable: Trt_group
 Response Variable: response(1)
 Control Arm: P
 Variance Type: Pooled
 Multiple Comparison Procedure: Sidak
 Confidence Level: 0.95
 Test Type: z-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Proportion | Std. Err. of Diff. of Proportions | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Proportions(One sided) | |
|------|-------------|------------|-----------------------------------|----------------|----------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 125 | 0.344 | NA | NA | NA | NA | NA | NA |
| H | 125 | 0.544 | 0.063 | 3.182 | 7.306E-4 | 0.002 | 0.097 | 1 |
| L | 125 | 0.392 | 0.061 | 0.787 | 0.216 | 0.518 | -0.052 | 1 |
| M | 125 | 0.44 | 0.062 | 1.555 | 0.06 | 0.169 | -0.006 | 1 |

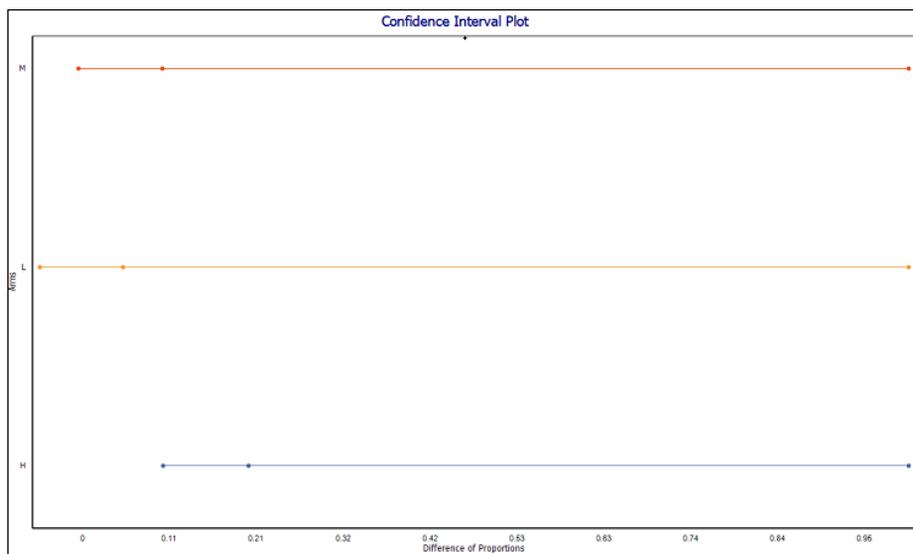
Adjusted Global p-value: 0.002
 Total Number of Records: 500
 Number of Records Rejected: 0
 Total Number of Arms (k): 4

The interpretation of the above output is similar to what was described for the output of Bonferroni procedure in section 92.2.

In addition to the above output, East also creates a confidence interval plot. A node labeled Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions2 is displayed in the Library. Under this node there is another node

92 Analysis- Multiple Comparison Procedures for Binary Data

labeled as **Confidence Interval Plot2**. To open the plot, double click this node.



Example: Weighted Bonferroni Procedure

Dataset: HIV-study.cyx as described in Section 92.2.

Purpose of the Analysis:

To analyze the data of the dose finding HIV trial using Weighted Bonferroni procedure.

Analysis Steps:

1. Click **Analysis Inputs** tab on the status bar below. This will display several input fields associated with multiple comparison test in the main window.
2. In the **Main** tab, select **Weighted Bonferroni** in the **Select MCP** drop-down. After selecting Weighted Bonferroni a table is displayed below the dropdown box. The table is to specify the proportion of alpha allocated to each comparison. By default **East** distributes the proportion of alpha equally among the treatment groups. For this example, enter **0.2** for group **L**, **0.3** for group **M**, and **0.5** for group **H**. Ideally, the sum of these values must add up to one. If the sum of these values do not add up to 1, **East** will automatically scale them to add up to 1.

Leave all other parameters as selected for the Bonferroni procedure above.

Analysis: Discrete Endpoint Many-Sample Test - Multi-Arm Design - Pairwise Comparisons to Control - Differences of Proportions

Data Set: HIV-study.cyd

Main Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
 Select Treatment Variable:
 Trt_group H L M P

Response Variable
 Select Response Variable: response
 Select Value: 1
 Pooled Variance Un-pooled Variance

Frequency: []

Multiple Comparison Procedure
 Select MCP: Weighted Bonferroni
 No. of Treatment Arms: 3

Table of Treatment specific parameters:

| Arm | Proportion of Alpha |
|-----|---------------------|
| L | 0.2 |
| M | 0.3 |
| H | 0.5 |

3. Click **OK** to start the analysis. The output, as shown below, is displayed in the

92 Analysis- Multiple Comparison Procedures for Binary Data

Output Preview Area.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions

Hypothesis

$H_1 : \pi_i - \pi_0 \leq 0$ Vs. $K_1 : \pi_i - \pi_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: HIV-study.cyd
 Treatment Variable: Trt_group
 Response Variable: response(1)
 Control Arm: P
 Variance Type: Pooled
 Multiple Comparison Procedure: Weighted Bonferroni
 Confidence Level: 0.95
 Test Type: z-test
 Rejection Region: Right-Tail
 Table For Proportion of Alpha:

| Arms | Proportion of Alpha |
|------|---------------------|
| L | 0.2 |
| M | 0.3 |
| H | 0.5 |

Output

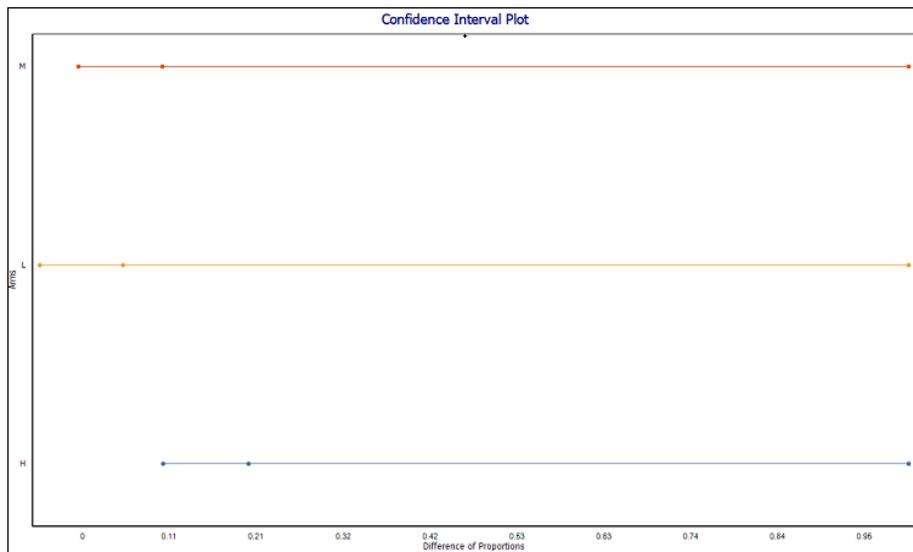
Inference for each Treatment Level versus Control:

| Arm | Sample Size | Proportion | Std. Err. of Diff. of Proportions | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Proportions(One sided) | |
|------|-------------|------------|-----------------------------------|----------------|----------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 125 | 0.344 | NA | NA | NA | NA | NA | NA |
| H | 125 | 0.544 | 0.063 | 3.182 | 7.306E-4 | 0.001 | 0.097 | 1 |
| L | 125 | 0.392 | 0.061 | 0.787 | 0.216 | 1 | -0.052 | 1 |
| M | 125 | 0.44 | 0.062 | 1.555 | 0.06 | 0.2 | -0.006 | 1 |

Adjusted Global p-value: 0.001
 Total Number of Records: 500
 Number of Records Rejected: 0
 Total Number of Arms (k): 4

The interpretation of the above output is similar to what was described for the output of Bonferroni procedure in section 92.2. In addition to the above output, East also creates a confidence interval plot. A node labeled **Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions3** is displayed in the Library. Under this node there is another node labeled as **Confidence Interval Plot3**. To open

the plot, double click the **Confidence Interval Plot** node.



92.3 Data-driven step-down MC procedure

92.3.0 Holm's Procedure

In the single step MC procedures, the decision to reject any hypothesis does not depend on the decision to reject other hypotheses. On the other hand, in the stepwise procedures decision of one hypothesis test can influence the decisions on the other tests of hypotheses. There are two types of stepwise procedures. One type of procedures proceed in data-driven order. The other type proceeds in a fixed order set a priori. Stepwise tests in a data-driven order can proceed in step-down or step-up manner.

East supports Holm step-down MC procedure which starts with the most significant comparison and continue as long as tests are significant until the test for certain hypothesis fails. The testing procedure stops at the first time a non-significant comparison occurs and all remaining hypotheses will be retained. In i -th step, $H_{(k-i)}$ is rejected if $p_{(k-i)} \leq \frac{\alpha}{i}$ and go to the next step.

Example: *Holm's Step Down Procedure*

Dataset: **HIV-study.cyx** as described in Section 92.2.

Purpose of the Analysis:

To analyze the data of the dose finding HIV trial using Holm's Step Down Procedure.

92 Analysis- Multiple Comparison Procedures for Binary Data

Analysis Steps:

1. Click **Analysis Inputs** tab on the status bar below. This will display several input fields associated with multiple comparison test in the main window.
2. In the **Main** tab, select **Holm's step down** in the **Select MCP** drop-down. Leave all other parameters as selected for the Bonferroni procedure above.

Analysis: Discrete Endpoint: Many-Sample Test - Multi-Arm Design - Pairwise Comparisons to Control - Differences of Proportions

Data Set: HIV-study.cyd

Main Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
 Select Treatment Variable:
 Trt_group

H
 L
 M
 P

Response Variable
 Select Response Variable: response
 Select Value: 1

Pooled Variance Un-pooled Variance

Frequency:

Multiple Comparison Procedure
 Select MCP: Holm's step down

3. Click **OK** to start the analysis. The output, as shown below, is displayed in the

Output Preview Area.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions

Hypothesis

$H_1 : \pi_i - \pi_0 \leq 0$ Vs. $K_1 : \pi_i - \pi_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: HIV-study.cyd
 Treatment Variable: Trt_group
 Response Variable: response(1)
 Control Arm: P
 Variance Type: Pooled
 Multiple Comparison Procedure: Holm's step down
 Confidence Level: 0.95
 Test Type: z-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Proportion | Std. Err. of Diff. of Proportions | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Proportions(One sided) | |
|------|-------------|------------|-----------------------------------|----------------|----------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 125 | 0.344 | NA | NA | NA | NA | NA | NA |
| H | 125 | 0.544 | 0.063 | 3.182 | 7.306E-4 | 0.002 | 0.097 | 1 |
| L | 125 | 0.392 | 0.061 | 0.787 | 0.216 | 0.216 | -0.052 | 1 |
| M | 125 | 0.44 | 0.062 | 1.555 | 0.06 | 0.12 | -0.006 | 1 |

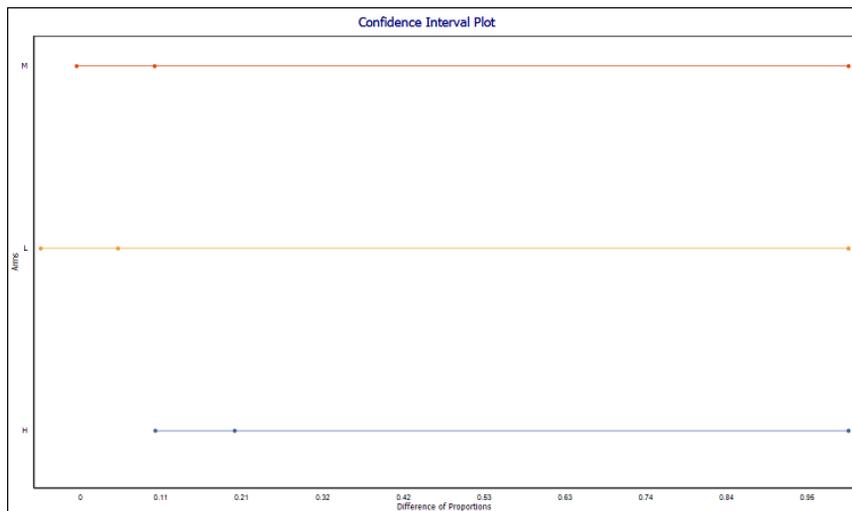
Adjusted Global p-value: 0.002
 Total Number of Records: 500
 Number of Records Rejected: 0
 Total Number of Arms (k): 4

The interpretation of the above output is similar to what was described for the output of Bonferroni procedure in section 92.2.

In addition to the above output, East also creates a confidence interval plot. A node labeled **Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions4** is displayed in the Library. Under this node there is another node labeled as **Confidence Interval Plot4**.

92 Analysis- Multiple Comparison Procedures for Binary Data

To open the plot, double click the **Confidence Interval Plot4** node.



92.4 Data-driven step-up MC procedures

92.4.0 Hochberg's Procedure

92.4.0 Hommel's Step up

Step-up tests start with the least significant comparison and continue as long as tests are not significant until the first time when a significant comparison occurs and all remaining hypotheses will be rejected.

East supports two such MC procedures - Hochberg step-up and Hommel step-up procedures.

In the Hochberg step-up procedure, in i^{th} step $H_{(k-i)}$ is retained if $p_{(k-i)} > \frac{\alpha}{i}$.

In the Hommel step-up procedure, in i^{th} step $H_{(k-i)}$ is retained if $p_{(k-j)} > \frac{i-j+1}{i}\alpha$ for $j = 1, \dots, i$. Fixed sequence test and fallback test are the types of tests, which proceed, in a prespecified order.

Example: Hochberg's Step Up Procedure

Dataset: HIV-study.cydx as described in Section 92.2.

Purpose of the Analysis:

To analyze the data of the dose finding HIV trial using Hochberg's Step Up Procedure.

Analysis Steps:

1. Click **Analysis Inputs** tab on the status bar below. This will display several input fields associated with multiple comparison test in the main window.
2. In the **Main** tab, select **Hochberg's step up** in the **Select MCP** drop-down. Leave all other parameters as selected for the Bonferroni procedure above.

The screenshot shows a software window titled "Analysis: Discrete Endpoint Many-Sample Test - Multi-Arm Design - Pairwise Comparisons to Control - Differences of Proportions". The "Data Set" is "HIV-study.cyd". The "Main" tab is selected, and the "Advanced" tab is also visible. The "raw Data" option is selected. Under "Treatment Variable and its Control Arm", "Trt_group" is selected in the dropdown, and radio buttons for H, L, M, and P are shown, with P being selected. Under "Response Variable", "response" is selected in the dropdown, and "1" is selected in the "Select Value" dropdown. The "Pooled Variance" option is selected. The "Frequency" dropdown is empty. On the right, the "Multiple Comparison Procedure" section shows "Hochberg's step up" selected in the "Select MCP" dropdown.

3. Click **OK** to start the analysis. The output, as shown below, is displayed in the **Output Preview Area**.

92 Analysis- Multiple Comparison Procedures for Binary Data

Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions

Hypothesis

$H_1 : \pi_i - \pi_0 \leq 0$ Vs. $K_i : \pi_i - \pi_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: HIV-study.cyd
 Treatment Variable: Trt_group
 Response Variable: response(1)
 Control Arm: P
 Variance Type: Pooled
 Multiple Comparison Procedure: Hochberg's step up
 Confidence Level: 0.95
 Test Type: z-test
 Rejection Region: Right-Tail

Output

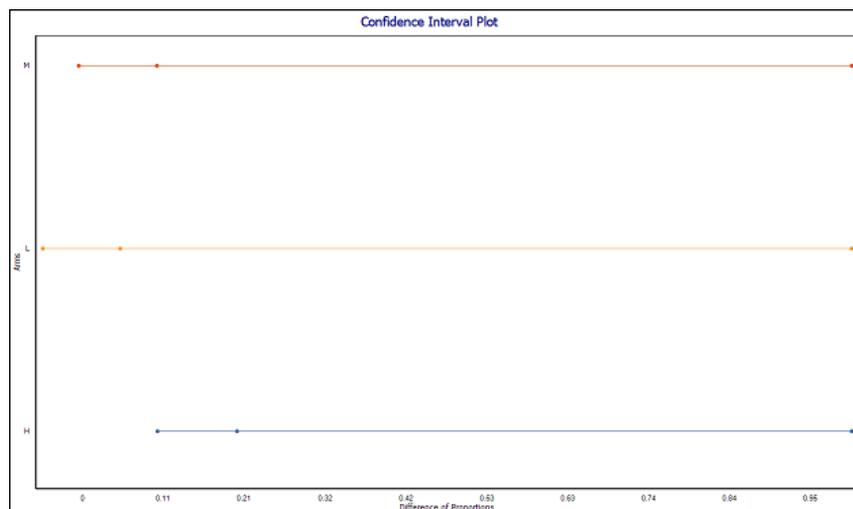
Inference for each Treatment Level versus Control:

| Arm | Sample Size | Proportion | Std. Err. of Diff. of Proportions | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Proportions(One sided) | | |
|------|-------------|------------|-----------------------------------|----------------|----------|----------|--|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | | Upper Limit |
| Ctrl | 125 | 0.344 | NA | NA | NA | NA | NA | | NA |
| H | 125 | 0.544 | 0.063 | 3.182 | 7.306E-4 | 0.002 | 0.097 | | 1 |
| L | 125 | 0.392 | 0.061 | 0.787 | 0.216 | 0.216 | -0.052 | | 1 |
| M | 125 | 0.44 | 0.062 | 1.555 | 0.06 | 0.12 | -0.006 | | 1 |

Adjusted Global p-value: 0.002
 Total Number of Records: 500
 Number of Records Rejected: 0
 Total Number of Arms (k): 4

The interpretation of the above output is similar to what was described for the output of Bonferroni procedure in section 92.2. In addition to the above output, East also creates a confidence interval plot. A node labeled **Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions5** is displayed in the Library. Under this node there is another node labeled as **Confidence Interval Plot5**. To open

the plot, double click the **Confidence Interval Plot5** node.



Example: Hommel's Step up Procedure

Dataset: HIV-study.cyx as described in Section 92.2.

Purpose of the Analysis:

To analyze the data of the dose finding HIV trial using Hochberg's Step Up Procedure.

Analysis Steps:

1. Click **Analysis Inputs** tab on the status bar below. This will display several input fields associated with multiple comparison test in the main window.
2. In the **Main** tab, select **Hommel's step up** in the **Select MCP** drop-down. Leave

92 Analysis- Multiple Comparison Procedures for Binary Data

all other parameters as selected for the Bonferroni procedure above.

3. Click **OK** to start the analysis. The output, as shown below, is displayed in the **Output Preview Area**.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions

Hypothesis

$H_1: \pi_i - \pi_0 \leq 0$ Vs. $K_i: \pi_i - \pi_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: HIV-study.cyd
 Treatment Variable: Trt_group
 Response Variable: response(1)
 Control Arm: P
 Variance Type: Pooled
 Multiple Comparison Procedure: Hommel's step up
 Confidence Level: 0.95
 Test Type: z-test
 Rejection Region: Right-Tail

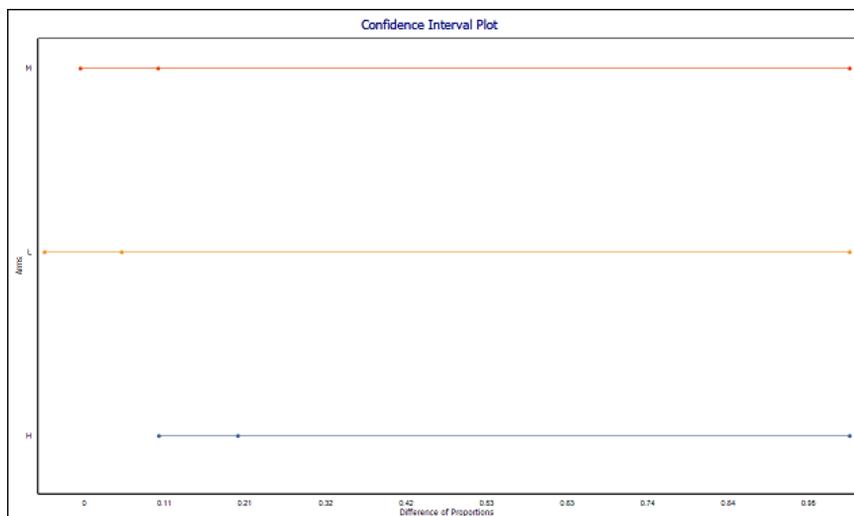
Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Proportion | Std. Err. of Diff. of Proportions | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Proportions(One sided) | | |
|------|-------------|------------|-----------------------------------|----------------|----------|----------|--|--------|-------------|
| | | | | | Naive | Adjusted | Lower Limit | | Upper Limit |
| Ctrl | 125 | 0.344 | NA | NA | NA | NA | | NA | NA |
| H | 125 | 0.544 | 0.063 | 3.182 | 7.306E-4 | 0.002 | | 0.097 | 1 |
| L | 125 | 0.392 | 0.061 | 0.787 | 0.216 | 0.216 | | -0.052 | 1 |
| M | 125 | 0.44 | 0.062 | 1.555 | 0.06 | 0.12 | | -0.006 | 1 |

Adjusted Global p-value: 0.002
 Total Number of Records: 500
 Number of Records Rejected: 0
 Total Number of Arms (k): 4

The interpretation of the above output is similar to what was described for the output of Bonferroni procedure in section 92.2. In addition to the above output, **East** also creates a confidence interval plot. A node labeled **Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions6** is displayed in the Library. Under this node there is another node labeled as **Confidence Interval Plot6**. To open the plot, double click the **Confidence Interval Plot6** node.



92.5 Fixed-seq stepwise MC procedures

In data-driven stepwise procedures, we don't have any control on the order of the hypotheses to be tested. However, sometimes based on our preference or prior knowledge we might want to fix the order of tests a priori. Fixed sequence test and fallback test are the types of tests which proceed in a pre-specified order. **East** supports both of these procedures.

Assume that H_1, H_2, \dots, H_{k-1} are ordered hypotheses and the order is pre-specified so that H_1 is tested first followed by H_2 and so on. Let p_1, p_2, \dots, p_{k-1} be the associated raw marginal p-values. In the fixed sequence testing procedure, for $i = 1, \dots, k - 1$, in i -th step, if $p_i < \alpha$, reject H_i and go to the next step; otherwise retain H_i, \dots, H_{k-1} and stop.

Fixed sequence testing strategy is optimal when early tests in the sequence have largest treatment effect and performs poorly when early hypotheses have small treatment effect or are nearly true (Westfall and Krishen 2001). The drawback of fixed sequence

92 Analysis- Multiple Comparison Procedures for Binary Data

test is that once a hypothesis is not rejected no further testing is permitted. This will lead to lower power to reject hypotheses tested later in the sequence.

Fallback test alleviates the above undesirable feature for fixed sequence test. Let w_i be the proportion of α for testing H_i such that $\sum_{i=1}^{k-1} w_i = 1$. In the fixed sequence testing procedure, in i -th step ($i = 1, \dots, k-1$), test H_i at $\alpha_i = \alpha_{i-1} + \alpha w_i$ if H_{i-1} is rejected and at $\alpha_i = \alpha w_i$ if H_{i-1} is retained. If $p_i < \alpha_i$, reject H_i ; otherwise retain it. Unlike the fixed sequence testing approach, the fallback procedure can continue testing even if a non-significant outcome is encountered by utilizing the fallback strategy. If a hypothesis in the sequence is retained, the next hypothesis in the sequence is tested at the level that would have been used by the weighted Bonferroni procedure. With $w_1 = 1$ and $w_2 = \dots = w_{k-1} = 0$, the fallback procedure simplifies to fixed sequence procedure.

Example: Fixed Sequence Procedure

Dataset: `HIV-study.cyx` as described in Section 92.2.

Purpose of the Analysis:

To analyze the data of the dose finding HIV trial using Fixed Sequence Procedure.

Analysis Steps:

1. Click **Analysis Inputs** tab on the status bar below. This will display several input fields associated with multiple comparison test in the main window.
2. In the **Main** tab, select **Fixed sequence** in the **Select MCP** drop-down. After selecting Fixed sequence a table will appear below the dropdown box. The table has two columns - **Arm** and **Test Sequence**. In the column **Test Sequence**, you have to specify the order in which the hypotheses will be tested. Specify 1 for the arm that will be compared first with Placebo, 2 for the arm that will be compared next and so on. By default **East** specifies 1 to the first arm, 2 to the second arm and so on. This default order implies that Dose1 will be compared first with Placebo, then Dose2 will be compared followed by comparison of Dose3 vs. Placebo. However, if we believe that efficacy of drug increases with dose, then the dose groups should be compared in descending order of dose. For this example, assign the high dose a sequential priority of 1, the middle dose as 2, and the low dose as 3. Leave all other parameters as selected for the

Bonferroni procedure above.

Analysis: Discrete Endpoint: Many-Sample Test - Multi-Arm Design - Pairwise Comparisons to Control - Differences of Proportions

Data Set: HIV-study.cyd

Main Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
 Select Treatment Variable:
 Trt_group H
 L
 M
 P

Response Variable
 Select Response Variable: response
 Select Value: 1
 Pooled Variance Un-pooled Variance

Frequency:

Multiple Comparison Procedure
 Select MCP: Fixed Sequence
 No. of Treatment Arms: 3

Table of Treatment specific parameters:

| Arm | Test Sequence |
|-----|---------------|
| H | 1 |
| M | 2 |
| L | 3 |

3. Click **OK** to start the analysis. The output, as shown below, is displayed in the

92 Analysis- Multiple Comparison Procedures for Binary Data

Output Preview Area.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions

Hypothesis

$H_1: \pi_i - \pi_0 \leq 0$ Vs. $K_i: \pi_i - \pi_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: HIV-study.cyd
 Treatment Variable: Trt_group
 Response Variable: response(1)
 Control Arm: P
 Variance Type: Pooled
 Multiple Comparison Procedure: Fixed Sequence
 Confidence Level: 0.95
 Test Type: z-test
 Rejection Region: Right-Tail
 Table For Test Sequence :

| Arms | Test Sequence |
|------|---------------|
| H | 1 |
| M | 2 |
| L | 3 |

Output

Inference for each Treatment Level versus Control:

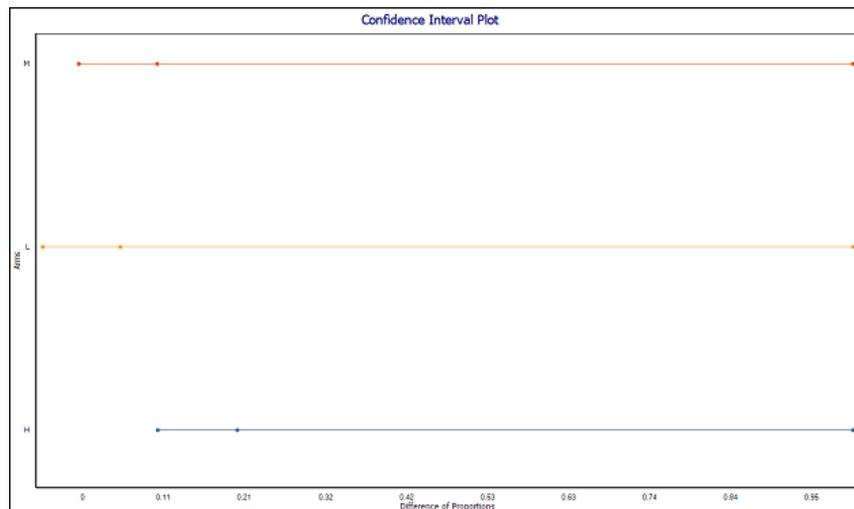
| Arm | Sample Size | Proportion | Std. Err. of Diff. of Proportions | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Proportions(One sided) | |
|------|-------------|------------|-----------------------------------|----------------|----------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 125 | 0.344 | NA | NA | NA | NA | NA | NA |
| H | 125 | 0.544 | 0.063 | 3.182 | 7.306E-4 | 7.306E-4 | 0.097 | 1 |
| L | 125 | 0.392 | 0.061 | 0.787 | 0.216 | 0.216 | -0.052 | 1 |
| M | 125 | 0.44 | 0.062 | 1.555 | 0.06 | 0.06 | -0.006 | 1 |

Adjusted Global p-value: 7.306E-4
 Total Number of Records: 500
 Number of Records Rejected: 0
 Total Number of Arms (k): 4

The interpretation of the above output is similar to what was described for the output of Bonferroni procedure in section 92.2.

In addition to the above output, East also creates a confidence interval plot. A node labeled **Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions7** is displayed in the Library. Under this node there is another node

labeled as **Confidence Interval Plot7**. To open the plot, double click this node.



Example: Fallback Procedure

Dataset: HIV-study.cyx as described in Section 92.2.

Purpose of the Analysis:

To analyze the data of the dose finding HIV trial using Fallback Procedure.

Analysis Steps:

1. Click **Analysis Inputs** tab on the status bar below. This will display several input fields associated with multiple comparison test in the main window.
2. In the **Main** tab, select **Fallback** in the **Select MCP** drop-down. After selecting Fallback, a table will appear under the dropdown box. This table is for you to specify the sequential priority for testing and the proportion of alpha allocated to each comparison. See the technical appendix H for details about this procedure. For this example, let's assign the high dose a sequential priority of 1, the middle dose 2, and the low dose 3. Also, for the proportion of alpha, let's allocate 0.3 to the low group, 0.3 to the middle group, and 0.4 to the high group. Leave all

92 Analysis- Multiple Comparison Procedures for Binary Data

other parameters as selected for the Bonferroni procedure above.

Analysis: Discrete Endpoint: Many-Sample Test - Multi-Arm Design - Pairwise Comparisons to Control - Differences of Proportions

Data Set: HIV-study.cyd

Main Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
 Select Treatment Variable: Trt_group
 H
 L
 M
 P

Response Variable
 Select Response Variable: response
 Select Value: 1
 Pooled Variance Un-pooled Variance
 Frequency: [dropdown]

Multiple Comparison Procedure
 Select MCP: Fallback
 No. of Treatment Arms: 3

Table of Treatment specific parameters:

| Arm | Proportion of Alpha | Test Sequen... |
|-----|---------------------|----------------|
| H | 0.4 | 1 |
| M | 0.3 | 2 |
| L | 0.3 | 3 |

3. Click **OK** to start the analysis. The output, as shown below, is displayed in the

main window.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions

Hypothesis
 $H_i : \pi_i - \pi_0 \leq 0$ Vs. $K_i : \pi_i - \pi_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: HIV-study.cyd
 Treatment Variable: Trt_group
 Response Variable: response(1)
 Control Arm: P
 Variance Type: Pooled
 Multiple Comparison Procedure: Fallback
 Confidence Level: 0.95
 Test Type: z-test
 Rejection Region: Right-Tail

Table For Proportion of Alpha and Test Sequence :

| Arms | Proportion of Alpha | Test Sequence |
|------|---------------------|---------------|
| H | 0.4 | 1 |
| M | 0.3 | 2 |
| L | 0.3 | 3 |

Output

Inference for each Treatment Level versus Control:

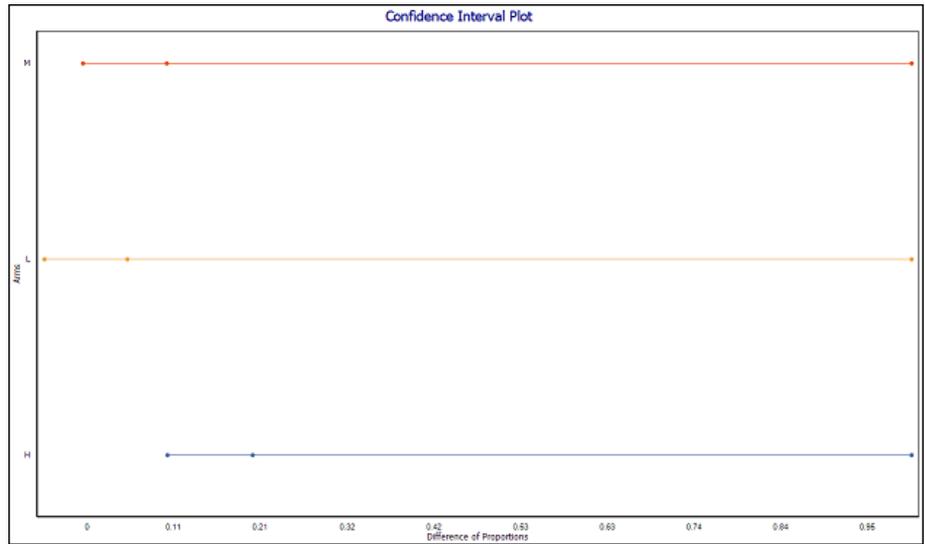
| Arm | Sample Size | Proportion | Std. Err. of Diff. of Proportions | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Proportions(One sided) | |
|------|-------------|------------|-----------------------------------|----------------|----------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 125 | 0.344 | NA | NA | NA | NA | NA | NA |
| H | 125 | 0.544 | 0.063 | 3.182 | 7.306E-4 | 0.002 | 0.097 | 1 |
| L | 125 | 0.392 | 0.061 | 0.787 | 0.216 | 0.216 | -0.052 | 1 |
| M | 125 | 0.44 | 0.062 | 1.555 | 0.06 | 0.086 | -0.006 | 1 |

Adjusted Global p-value: 0.002
 Total Number of Records: 500
 Number of Records Rejected: 0
 Total Number of Arms (k): 4

The interpretation of the above output is similar to what was described for the output of Bonferroni procedure in section 92.2. In addition to the above output, East also creates a confidence interval plot. A node labeled **Analysis: Multiple Pairwise Comparisons with Control: Difference of Proportions8** is displayed in the Library. Under this node there is another node labeled as **Confidence Interval Plot8**. To open

92 Analysis- Multiple Comparison Procedures for Binary Data

the plot, double click the **Confidence Interval Plot** node.



93

Analysis-Comparison of Multiple Comparison Procedures for Continuous Data- Analysis

In this chapter, we will use the hypertension trial example to illustrate the different multiple testing procedures. There are two scenarios. One scenario has increasing dose-response profile and the other one has decreasing dose-response profile. The data sets are available in the Samples subfolder in the **East** installation directory with file names Hypertension-trial.cyd and Hypertension-trial 2.cyd. The trial was conducted to compare the effects of four doses of the new drug. The doses are labeled as D1, D2, D3, and D4 from the lowest dose D1 to the highest dose D4. Table 93.1 and 93.2 display the mean treatment effects of each dose group against placebo group, standard errors, t statistics, raw p -values and 97.5% lower confidence limits for the two scenarios.

Table 93.1: Summary Statistics for scenario 1

| Dose | Mean Effect | Standard Error | t statistics | p -value | 97.5% Lower Confidence Limit |
|------|-------------|----------------|----------------|------------|------------------------------|
| D1 | -0.6957 | 1.9634 | -0.3543 | 0.638138 | -4.5831 |
| D2 | 4.5498 | 1.9245 | 2.3642 | 0.009838 | 0.7395 |
| D3 | 4.9252 | 1.9634 | 2.5085 | 0.00673 | 1.0378 |
| D4 | 6.6268 | 1.9245 | 3.4434 | 0.000396 | 2.8164 |

Table 93.2: Summary Statistics for scenario 2

| Dose | Mean Effect | Standard Error | t statistics | p -value | 97.5% Lower Confidence Limit |
|------|-------------|----------------|----------------|------------|------------------------------|
| D1 | 8.3574 | 1.9817 | 4.2173 | 0.000024 | 4.4354 |
| D2 | 4.979 | 1.9817 | 2.5125 | 0.006631 | 1.057 |
| D3 | 4.5469 | 1.9817 | 2.2944 | 0.011717 | 0.6249 |
| D4 | 0.9544 | 1.9817 | 0.4816 | 0.315461 | -2.9676 |

Table 93.3 displays the adjusted p -values for all the multiplicity adjustment methods. The numbers highlighted in red are significant at 0.025 level.

Single step Dunnett test finds two significant doses in both scenario 1 and 2. Using Bonferroni test, only Dose 4 is superior to placebo in Scenario 1 and only Dose 1 is superior to placebo in Scenario 2. Also, note that the adjusted p -values by single step Dunnett test are all smaller than those by Bonferroni test. This is because single step

93 Analysis-Comparison of Multiple Comparison Procedures for Continuous Data- Analysis

Table 93.3: Adjusted p values for scenario 1

| MCP procedure | D1 | D2 | D3 | D4 |
|---------------------------------------|----------|----------|----------|----------|
| Single step Dunnett | 0.895822 | 0.032812 | 0.022962 | 0.001488 |
| Step down Dunnett | 0.638138 | 0.018357 | 0.018067 | 0.001488 |
| Bonferroni | 1 | 0.039351 | 0.02692 | 0.001584 |
| Sidak | 0.982854 | 0.038774 | 0.026649 | 0.001583 |
| Holm | 0.638138 | 0.02019 | 0.02019 | 0.001584 |
| Hochberg | 0.638138 | 0.019676 | 0.019676 | 0.001584 |
| Hommel | 0.638138 | 0.019676 | 0.014757 | 0.001584 |
| Fixed sequence (D1,D2,D3,D4) | 0.638138 | 0.638138 | 0.638138 | 0.638138 |
| Fallback (D1,D2,D3,D4, equal weights) | 1 | 0.039351 | 0.02692 | 0.001584 |

Dunnett test is a parametric test, which takes into account the joint distribution of the test statistics.

Dunnett step down test finds three significant doses in both scenario 1 and 2. It is a closed test based on single step Dunnett procedure and is uniformly more powerful than single step Dunnett test. This can be seen from the fact that all adjusted p-values by Dunnett step down test are smaller than those by single step Dunnett test. The relationship between Dunnett step down test and Holm test is similar to that between single step Dunnett and Bonferroni test. Dunnett step down test is a parametric procedure of Holm test and is uniformly more powerful than Holm test which is confirmed by the smaller p -values adjusted by step down Dunnett test than those adjusted by Holm test.

Sidak test gives similar adjusted p-values to those provided by Bonferroni test. These two test have very similar performance.

Holm test rejects three doses in both scenarios and all the adjusted p-values by Holm test are smaller than or equal to those by Bonferroni test. This is because Holm test is a closed test based on Bonferroni procedure and consequently it is uniformly more powerful than Bonferroni test.

Hochberg and Hommel procedures also reject the same three hypotheses in both scenarios. However, their adjusted p -values for all the doses are smaller than or equal to those by Holm procedure. This is the well-known fact that Hochberg and Hommel procedures are uniformly more powerful than Holm test. Hommel procedure is

Table 93.4: Adjusted p values for scenario 2

| MCP procedure | D1 | D2 | D3 | D4 |
|---------------------------------------|----------|----------|----------|----------|
| Single step Dunnett | 0.000092 | 0.022646 | 0.038631 | 0.60987 |
| Step down Dunnett | 0.000092 | 0.017788 | 0.02176 | 0.315461 |
| Bonferroni | 0.000094 | 0.026523 | 0.046866 | 1 |
| Sidak | 0.000094 | 0.026261 | 0.046049 | 0.78042 |
| Holm | 0.000094 | 0.019893 | 0.023433 | 0.315461 |
| Hochberg | 0.000094 | 0.019893 | 0.023433 | 0.315461 |
| Hommel | 0.000094 | 0.017575 | 0.023433 | 0.315461 |
| Fixed sequence (D1,D2,D3,D4) | 0.000024 | 0.006631 | 0.011717 | 0.315461 |
| Fallback (D1,D2,D3,D4, equal weights) | 0.000094 | 0.013262 | 0.015622 | 0.315461 |

uniformly more powerful than Hochberg procedure. Their performances are similar which can be seen from the similar adjusted p -values with Hommel adjusted p -values being slightly smaller than the Hochberg ones. Note that Hochberg and Hommel tests control the FWER when the joint distribution of the test statistics have a certain type of positive dependence so called multivariate totally positive of order two (Sarkar and Chang 1997, Sarkar 1998). For negatively correlated test statistics, Hochberg and Hommel procedures might not control the FWER.

Fixed sequence test fails to reject all the doses in Scenario 1 where all the adjusted p values are more than 0.5. However, this test rejects dose 1, 2 and 3 in Scenario 2. Further note that the fixed sequence test performs uniformly better than all other procedures since the adjusted p -values are smaller than all those by other procedures. This illustrates an important feature of the fixed sequence test. This test performs best when the testing order is in line with the magnitudes of the underlying true treatment effects. In other words, if the hypotheses being tested earlier in the sequence have larger treatment effects, the fixed sequence procedure is more powerful. On the other hand, if the treatment effects are not monotone with respect to the testing order, this test performs poorly.

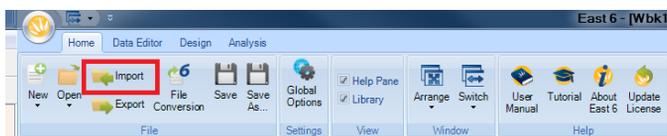
Fallback procedure rejects dose 4 in Scenario 1 like Bonferroni and Sidak procedures. However, it rejects three doses in Scenario 2, dose 1, 2, and 3. The adjusted p -values generated by fallback test are smaller than those produced by Holm, Hochberg and Hommel tests. This implies that fallback test with equal weights performs better than Holm, Hochberg and Hommel tests when the testing order is in line with the magnitudes of the treatment effects. Also, note that fallback test is more robust than

93 *Analysis-Comparison of Multiple Comparison Procedures for Continuous Data- Analysis*

fixed sequence test, especially when the testing order is not consistent with the order of the true treatment effects as in Scenario 1 where fallback finds one significance whereas fixed sequence does not find any significant results.

94 *Analysis-Multiple Endpoints for Binary Data*

In Chapter 28, we have seen how to evaluate different gatekeeping procedures for correlated binary outcome through intensive simulations. In this chapter, we will illustrate how to analyze a trial with binary outcome with gatekeeping multiple comparison procedures. Consider the same example used in Chapter 27: a randomized, placebo-controlled, double blind, parallel treatment clinical trial designed to compare two treatments for migraine. In this study, Telcagepant (300mg), an antagonist of the CGRP receptor associated with migraine, and zolmitriptan (5mg) the standard treatment against migraine, are compared against a placebo. The five co-primary endpoints include pain freedom, pain relief, absence of photophobia (sensitivity to light), absence of phonophobia (sensitivity to sound), and absence of nausea two hours post treatment. Three co-secondary endpoints included more sustained measurements of pain freedom, pain relief, and total migraine freedom for up to a 24 hour period. For illustration purpose, we consider three primary endpoints, pain freedom (PF), absence of phonophobia (phono) and absence of photophobia (photo) at two hours post treatment. Only one endpoint from the secondary family, sustained pain freedom (SPF), will be included in the example. The data set is saved in the installation folder of EAST as Migraine.csv. To analyze this data set, we need to import the data into EAST by clicking on the Import icon as seen in the following screen.



Select the Migraine.csv file and click OK to see the data set displayed in EAST. The

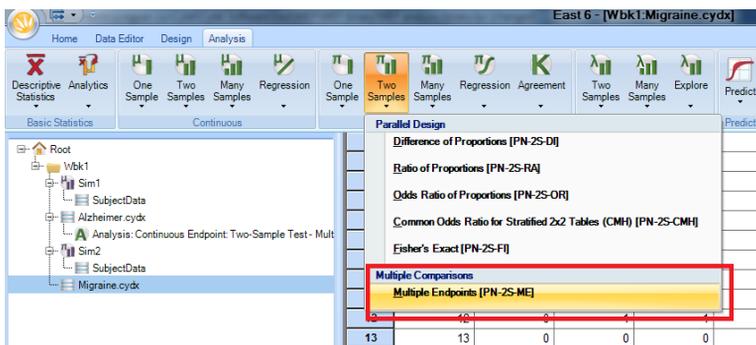
94 Analysis-Multiple Endpoints for Binary Data

following screen shows a snapshot of the data set.

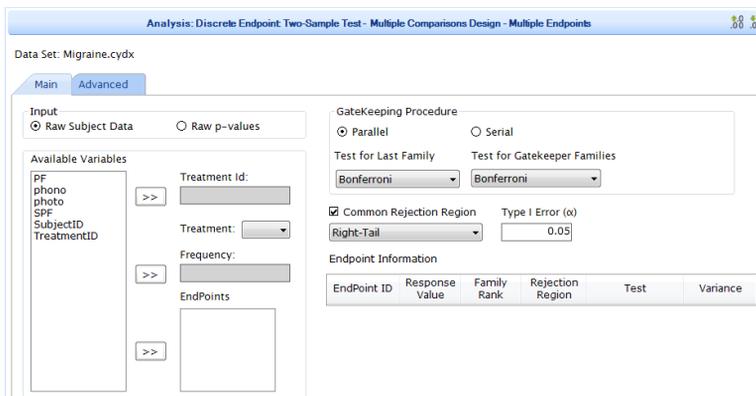
| SubjectID | TreatmentID | PF | phono | photo | SPF |
|-----------|-------------|----|-------|-------|-----|
| 1 | 0 | 0 | 1 | 1 | 0 |
| 2 | 0 | 0 | 1 | 0 | 0 |
| 3 | 0 | 0 | 0 | 1 | 0 |
| 4 | 0 | 0 | 1 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 1 | 0 |
| 7 | 0 | 1 | 1 | 1 | 0 |
| 8 | 0 | 0 | 1 | 1 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0 | 0 | 1 | 1 | 1 |
| 11 | 0 | 0 | 0 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | 0 |
| 13 | 0 | 0 | 1 | 0 | 0 |
| 14 | 0 | 0 | 0 | 0 | 0 |
| 15 | 0 | 0 | 0 | 1 | 0 |
| 16 | 0 | 0 | 0 | 0 | 0 |
| 17 | 0 | 1 | 1 | 1 | 1 |
| 18 | 0 | 1 | 1 | 1 | 0 |
| 19 | 0 | 0 | 0 | 0 | 0 |
| 20 | 0 | 0 | 0 | 1 | 0 |
| 21 | 0 | 0 | 0 | 0 | 0 |

Now click on the Analysis menu on the top of EAST window, select Two Samples for discrete outcome and then select Multiple Comparisons-Multiple Endpoints from the

dropdown list.



The main input dialog window pops up as seen in the following screen.



EAST can analyze two types of data: (1) raw subject level data, (2) raw p-values. For the migraine example, the data is raw subject level data so we select the left radio button. The left bottom panel of the screen displays all the variables contained in the data set. We need to specify which variable contains the information on treatment group ID for each subject and further specify which one is active treatment group. The next input is to identify all the endpoints to be analyzed. For this example, PF, phono and photo constitute the primary family of endpoints. SPF constitutes the secondary family. Suppose we need to analyze the data using serial gatekeeping procedure. After

94 Analysis-Multiple Endpoints for Binary Data

filling in all inputs, the screen looks as follows

The screenshot shows the 'Advanced' tab of the software interface. The 'Data Set' is 'Migraine.cyx'. The 'Input' section has 'Raw Subject Data' selected. The 'Available Variables' list contains 'SubjectID'. The 'Treatment ID' is 'TreatmentID', 'Treatment' is '1', and 'Frequency' is empty. The 'EndPoints' list contains 'PF', 'phono', 'photo', and 'SPF'. The 'GateKeeping Procedure' is 'Serial'. The 'Test for Last Family' is 'Bonferroni'. The 'Common Rejection Region' is checked and set to 'Right-Tail' with a 'Type I Error (α)' of 0.05. The 'Endpoint Information' table is as follows:

| Endpoint ID | Response Value | Family Rank | Rejection Region | Test | Variance |
|-------------|----------------|-------------|------------------|---------------|----------|
| PF | 1 | 1 | Right-Tail | Diff of Props | Pooled |
| phono | 1 | 1 | Right-Tail | Diff of Props | Pooled |
| photo | 1 | 1 | Right-Tail | Diff of Props | Pooled |
| SPF | 1 | 2 | Right-Tail | Diff of Props | Pooled |

Now click on OK button on the right bottom of the screen to run the analysis. The

following screen displays the detailed output of this analysis.

Analysis: Discrete Endpoint: Two-Sample Test - Multiple Comparisons Design - Multiple Endpoint

Data File: Migraine.cydx
 Data Input Method: Raw Data
 Treatment Variable: TreatmentID
 Total Number of Endpoints: 4
 Type I Error (α): 0.05
 Gatekeeping Procedure: Serial
 Test for Last Family: Bonferroni

Output

Total Number of Records: 100
 Number of Records Rejected: 0
 Summary of each Endpoint:

| EndPoint ID | Test | Variance Type | Sample Size | Control Proportion | Treatment Proportion | Est. of Delta | Std. Err. of Est. of Delta | Test Statistic | 95% Confidence Interval of Delta (Two-Sided) | |
|-------------|---------------|---------------|-------------|--------------------|----------------------|---------------|----------------------------|----------------|--|-------------|
| | | | | | | | | | Lower Limit | Upper Limit |
| PF | Diff of Props | Pooled | 100 | 0.12 | 0.26 | 0.14 | 0.078 | 1.784 | -0.014 | 0.294 |
| phono | Diff of Props | Pooled | 100 | 0.4 | 0.66 | 0.26 | 0.1 | 2.605 | 0.064 | 0.456 |
| photo | Diff of Props | Pooled | 100 | 0.32 | 0.48 | 0.16 | 0.098 | 1.633 | -0.032 | 0.352 |
| SPF | Diff of Props | Pooled | 100 | 0.04 | 0.24 | 0.2 | 0.069 | 2.882 | 0.064 | 0.336 |

Inference of Gatekeeping Procedures:

| Family Rank | EndPoint ID | Rejection Region | Test Statistic | p-Value | | Status |
|-------------|-------------|------------------|----------------|---------|----------|------------------|
| | | | | Raw | Adjusted | |
| 1 | PF | Right-Tail | 1.784 | 0.037 | 0.051 | Unable to Reject |
| | phono | Right-Tail | 2.605 | 0.005 | 0.051 | Unable to Reject |
| | photo | Right-Tail | 1.633 | 0.051 | 0.051 | Unable to Reject |
| Not Passed | | | | | | |
| 2 | SPF | Right-Tail | 2.882 | 0.002 | 0.051 | Not Tested |

*Using Bonferroni test for the last family in the Serial gatekeeping procedure, the treatment arm is not significantly different from the control arm on the co-primary endpoints. As a result, the other non-primary endpoints are not tested.

The first table shows the summary statistics for each endpoint including mean for each treatment group, estimate of treatment effect, standard error of the effect estimate, test statistic and marginal two-sided confidence interval. The second table shows the inference summary including raw p-values, multiplicity adjusted p-values with the gatekeeping procedure and significance status. It also shows whether the primary family is passed as the serial gatekeeper for the secondary family of endpoints.

95 *Analysis-Agreement*

This chapter discusses Cohen’s Kappa and the Weighted Kappa measures. These two measures are used to assess the level of agreement between two observers classifying a sample of objects on the same categorical scale. The joint ratings of the observers are displayed on a square $r \times r$ contingency table.

95.1 Available Measures

A reference for each measure of agreement is provided in the table shown below:

| Measure Of Agreement | References |
|----------------------|------------------|
| Cohen’s Kappa | Agresti (2002) |
| Weighted Kappa | Liebetrau (1983) |

Note the following special features of these procedures.

- For every possible option, in addition to the option specific output, you also get the maximum likelihood point estimate of the measure of agreement (MLE), its asymptotic standard error (ASE_MLE), a confidence interval for the measure of agreement, and asymptotic 1 and 2-sided p-values for testing the null hypothesis that Kappa (or weighted Kappa) equals zero.
- Negative values of Kappa are possible, reflecting agreement weaker than might be expected by chance, but are rare in practice.

95.2 When to Use Each Measure

The two measures in this chapter capture the extent to which two sets of observers classifying the same set of objects agree.

- **Cohen’s Kappa:** Use Cohen’s Kappa when the classification of each object by the two observers is on a nominal scale.
- **Weighted Kappa:** Use the Weighted Kappa when the classification of each object by the two observers is on an ordered scale.

95.3 Example: Cohen’s Kappa

Dataset: Radio_Case_data.cyx

Data Description

It is hypothetical data concerning two radiologists who rated 85 patients with respect to liver lesions. The ratings were designated on an ordinal scale as "Normal", "Benign", "Suspected", and "Cancer". The following table provides the data:

| Rater1/Rater2 | Normal | Benign | Suspected | Cancer |
|---------------|--------|--------|-----------|--------|
| Normal | 21 | 12 | 0 | 0 |
| Benign | 4 | 17 | 1 | 0 |
| Suspected | 3 | 9 | 15 | 2 |
| Cancer | 0 | 0 | 0 | 1 |

Purpose of the Analysis:

To calculate Cohen’s Kappa estimates based on the selected dataset.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

Analysis > (Discrete) Agreement > (Parallel Design) Cohen’s Kappa

This will display several input fields associated with Cohen’s Kappa test in the main window.

3. In the **Main** tab, select the variables as shown below:

Analysis: Discrete Endpoint Agreement Test - Paired Design - Cohen's Kappa

Data Set: Radio_Case_data.cyd

Main
Advanced

Variable 1(Rater 1): Column
 Weighted Kappa
Frequency Variable: Weight

Variable 2(Rater 2): Row

95 Analysis-Agreement

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Categorical Response: Cohen's Kappa

Hypothesis

H_0 : Agreement between two raters is purely from Random variation Vs. (i) H_1 ; Agreement between two raters is not purely from Random variation, for 2-Sided test
 Either (ii) H_1 ; Agreement between two raters is greater than that is expected from Random variation only
 Or (iii) H_1 ; Agreement between two raters is less than that is expected from Random variation only for 1-Sided test

Input Parameters

Data File: Radio_Case_data.cyd
 Rater1: Column
 Rater2: Row
 Frequency Variable: Weight
 Coefficient : Weighted Kappa
 Confidence Level: 0.95

Output

Total Number of Records: 10
 Number of Records Rejected: 0

Table of Observed Frequencies:

| | 1 | 2 | 3 | 4 | Total |
|-------|----|----|----|---|-------|
| 1 | 21 | 4 | 3 | 0 | 28 |
| 2 | 12 | 17 | 9 | 0 | 38 |
| 3 | 0 | 1 | 15 | 0 | 16 |
| 4 | 0 | 0 | 2 | 1 | 3 |
| Total | 33 | 22 | 29 | 1 | 85 |

Summary of Test Statistic:

| Coefficient | Estimate | ASE(MLE) | 95% Confidence Interval | |
|----------------|----------|----------|-------------------------|-------------|
| | | | Lower Limit | Upper Limit |
| Weighted Kappa | 0.671 | 0.068 | 0.538 | 0.805 |

Test of Hypothesis:

| Type | Statistic | p-value | | | | |
|--------------------|-----------|---------|-----------|-----------|------|-----------|
| | | Tail | 1 Sided | 2*1-Sided | Tail | 2 Sided |
| Asymptotic | Observed | G.E. | 2.454E-10 | 4.907E-10 | G.E. | 4.907E-10 |
| 99% CI Lower Limit | | | 0 | | | 0 |
| 99% CI Upper Limit | | | 4.604E-4 | | | 4.604E-4 |

East displays estimate of Kappa to be 0.671 which indicates moderate agreement between two radiologist. The asymptotic 1-sided as well as 2-sided p-value is very low. The hypothesis of no agreement is rejected at 5% two sided level of significance.

96 *Analysis-Survival Data*

96.1 *Superiority*

In this section, we explore how we can use **East** to compare two survival curves. **East** provides the option of using Log Rank Test for this purpose.

Here, our endpoint of interest is time-to-event. Some situations in medical research could be: study of a new-anticancer agent on patient survival; study of an anti-depressant drug on shortening the interval between diagnosis of depression and response to treatment and so on. More formally, we are interested in comparing the hazard rate parameters λ_t and λ_c between the treatment and control populations. Define $\delta = \ln(\lambda_t/\lambda_c)$. The null hypothesis $H_0 : \delta = 0$ is tested against a 2-sided alternative $H_1 : \delta \neq 0$ or against a one-sided alternative $H_1 : \delta < 0$ or $H_1 : \delta > 0$.

where

$$\lambda_t(u) = \frac{f_t(u)}{1 - F_t(u)}$$

and

$$\lambda_c(u) = \frac{f_c(u)}{1 - F_c(u)}$$

associated with the survival distributions F_t and F_c , respectively. Then the Logrank test is especially effective for detecting the proportional hazards alternative hypothesis.

Under the null hypothesis, $\log \delta = 0$. If $\log \delta$ is positive, population F_c prolongs survival relative to population F_t , while if $\log \delta$ is negative, population F_t prolongs survival relative to population F_c .

96.2 *Example: Survival Superiority Two Samples:Logrank*

Dataset: Cancer.cyx

Data Description

This data is from a small lung cancer clinical trial involving a new and control drug. The dataset has three variables **Drug**, **Response** and **Censored**.

The variable **Drug** acts as an identifier of the population to which the observation belongs. The value **1** corresponds to the control group and value **2** corresponds to the treatment group.

The **Response** variable provides survival time (in days).

96 Analysis-Survival Data

The variable **Censored** gives information about which observation is censored. The value **0** corresponds to censoring and the value **1** corresponds to non-censoring.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

Analysis > (Events) Two Samples > (Parallel Design) Logrank

This will display several input fields associated with Logrank Test in the main window.

3. In the **Main** tab, select **Superiority** as **Trial Type** and **Drug** as **Population Id.** Enter **1** as **Control** and **2** as **Treatment**. Select **Response** as **Response** variable and **Censored** as **Censor** variable with **Censor Value** as 0 and **Complete** as 1. This data does not have a frequency variable, so leave it blank.

4. In the **Advanced** tab leave the fields **By Variable 1** and **By Variable 2** blank. Keep the default value **0.95** for **Confidence Level**.

5. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Time to Event Response: Logrank Test

Let $\delta = \ln(\lambda_1 / \lambda_0)$
 $H_0 : \delta = 0$ Vs. $H_1 : \delta \neq 0$ for 2-Sided test
 Either $H_1 : \delta > 0$ Or $H_1 : \delta < 0$ for 1-Sided test

Data File: Cancer.cyd
 Trial Type: Superiority
 Population ID: Drug(Treatment=2, Control=1)
 Response Variable: Response
 Censor: Censored(Censor Value=0, Complete=1)
 Confidence Level: 0.95

Output

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | |
|--------------|-----------------|--------|--------|----------|--------|
| | | Count | % | Count | % |
| 1 | 5 | 2 | 40 | 3 | 60 |
| 2 | 9 | 9 | 100 | 0 | 0 |
| Total | 14 | 11 | 78.571 | 3 | 21.429 |

Parameter Estimates:

| Hazard Ratio (HR) | 95% Confidence Interval(2-Sided) | |
|-------------------|----------------------------------|-------------|
| | Lower Limit | Upper Limit |
| 11.168 | 1.252 | 99.626 |

Test of Hypothesis:

| Log Rank Score | Std. Error | Standardized Test Statistic | (1-Sided) | | (2-Sided) |
|----------------|------------|-----------------------------|-----------|---------|-----------|
| | | | Tail | p-value | p-value |
| 4.076 | 1.439 | 2.833 | G.E. | 0.002 | 0.005 |

Estimated Hazard Rates:

| | |
|-------------------------------|----------|
| Control (λ_0) | 6.207E-4 |
| Treatment (λ_0 * HR) | 0.007 |

East calculates 2-sided as well as 1-sided p-values. 2-sided p-value for this test is 0.005 and 1-sided p-value is 0.002. At 5% significance level, the null hypothesis is rejected.

96.3 Example :Survival Superiority Two Samples: Wilcoxon-Gehan

Dataset: Cancer.cydx

Data Description

96 Analysis-Survival Data

This data is from a small lung cancer clinical trial involving a new and control drug. The dataset has three variables **Drug**, **Response** and **Censored**.

The variable **Drug** acts as an identifier of the population to which the observation belongs. The value **1** corresponds to the control group and value **2** corresponds to the treatment group.

The **Response** variable provides survival time (in days).

The variable **Censored** gives information about which observation is censored. The value **0** corresponds to censoring and the value **1** corresponds to non-censoring.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item: **Analysis > (Events) Two Samples > (Parallel Design)**

Wilcoxon

This will display several input fields associated with Wilcoxon-Gehan Test in the main window.

3. In the **Main** tab, select **Superiority** as **Trial Type** and **Drug** as **Population Id**. Enter **1** as **Control** and **2** as **Treatment**. Select **Response** as **Response** variable and **Censored** as **Censor** variable with **Censor Value** as **0** and **Complete** as **1**. This data does not have a frequency variable, so leave it blank. Choose **Test Statistic** as **Wilcoxon-Gehan**.

Data Set: Cancer.cyd

| | | | | | |
|----------------|-------------|--------------------|----------|---------------------|----------------|
| Main | | Advanced | | | |
| Trial Type: | Superiority | Response Variable: | Response | Frequency Variable: | |
| Population ID: | Drug | Censor Indicator: | Censored | Test Statistic: | Wilcoxon-Gehan |
| Control: | 1 | Censored: | 0 | | |
| Treatment: | 2 | Complete: | 1 | | |

4. In the **Advanced** tab leave the fields **By Variable 1** and **By Variable 2** blank.

Keep the default value 0.95 for **Confidence Level**.

Analysis: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank

Data Set: Cancer.cyd

Main Advanced

By Variable 1: Confidence Level:

By Variable 2:

5. Click **OK** to start the analysis. The output is displayed in the main window.



Analysis: Time to Event Response: Logrank Test

Let $\delta = \ln (\lambda_t / \lambda_c)$

$H_0 : \delta = 0$ Vs. $H_1 : \delta \neq 0$ for 2-Sided test

Either $H_1 : \delta > 0$ Or $H_1 : \delta < 0$ for 1-Sided test

Data File: Cancer.cyd
Trial Type: Superiority
Population ID: Drug(Treatment=2, Control=1)
Response Variable: Response
Censor: Censored (Censor Value=0, Complete=1)
Type of Test Statistic: Wilcoxon-Gehan
Confidence Level: 0.95

Output

Summary of Observed Data:

| Treatment ID | No.of Subjects | Events | | Censored | | Average Follow-up Time |
|--------------|----------------|--------|----------|----------|---------|------------------------|
| | | Count | % | Count | % | |
| 1 | 5 | 2 | 40.000% | 3 | 60.000% | 644.4 |
| 2 | 9 | 9 | 100.000% | 0 | 0.000% | 178 |
| Total | 14 | 11 | 78.571% | 3 | 21.429% | 344.571 |

96 Analysis-Survival Data

Test of Hypothesis:

| Wilcoxon Gehan Score | Std. Error | Standardized Test Statistic | (1-Sided) | | (2-Sided) |
|----------------------|------------|-----------------------------|-----------|---------|-----------|
| | | | Tail | p-value | p-value |
| 39 | 14.542 | 2.682 | G.E. | 0.004 | 0.007 |

Parameter Estimates from Cox Model:

| Hazard Ratio (HR) | ln(Hazard Ratio) (δ) | Std. Error of Estimate of δ | Wald Statistic | 95% Confidence Interval(2-Sided) | |
|-------------------|-------------------------------|------------------------------------|----------------|----------------------------------|-------------|
| | | | | Lower Limit | Upper Limit |
| 11.168 | 2.413 | 1.08 | 2.235 | 1.346 | 92.679 |

Estimated Hazard Rates:

| | |
|---------------------------------------|----------|
| Control (λ_c) | 6.207E-4 |
| Treatment ($\lambda_c + \text{HR}$) | 0.007 |

East calculates 2-sided as well as 1-sided p-values. 2-sided p-value for this test is 0.007 and 1-sided p-value is 0.004. At 5% significance level, the null hypothesis is rejected.

96.4 Example:Survival Superiority Two Samples: Harrington-Fleming

Dataset: Cancer.cyx

Data Description

This data is from a small lung cancer clinical trial involving a new and control drug. The dataset has three variables **Drug**, **Response** and **Censored**.

The variable **Drug** acts as an identifier of the population to which the observation belongs. The value **1** corresponds to the control group and value **2** corresponds to the treatment group.

The **Response** variable provides survival time (in days).

The variable **Censored** gives information about which observation is censored. The value **0** corresponds to censoring and the value **1** corresponds to non-censoring.

Analysis Steps:

1. Open the dataset from **Samples** folder.

- Choose the menu item:

**Analysis > (Events) Two Samples > (Parallel Design)
 Harrington-Fleming-Sup**

This will display several input fields associated with Harrington-Fleming Test in the main window.

- In the **Main** tab, select **Superiority** as **Trial Type** and **Drug** as **Population Id**. Enter **1** as **Control** and **2** as **Treatment**. Select **Response** as **Response** variable and **Censored** as **Censor** variable with **Censor Value** as **0** and **Complete** as **1**. This data does not have a frequency variable, so leave it blank. Choose **Test Statistic** as **Harrington-Fleming**. Leave the default values of p and q as 1 each.

Analysis: Survival Endpoint Two-Sample Test - Parallel Design - Logrank

Data Set: Cancer.cyd

Main Advanced

Superiority Response Variable: Response Frequency Variable:

Drug Censor Indicator: Censored Test Statistic: Harrington-Fleming P: 1 Q: 1

1 Censored: 0
 2 Complete: 1

- In the **Advanced** tab leave the fields **By Variable 1** and **By Variable 2** blank. Keep the default value **0.95** for **Confidence Level**.

Data Set: Cancer.cyd

Main Advanced

By Variable 1: Confidence Level: 0.95

By Variable 2:

- Click **OK** to start the analysis. The output is displayed in the main window.

96 Analysis-Survival Data

Analysis: Time to Event Response: Logrank Test

Let $\delta = \ln(\lambda_1 / \lambda_0)$

$H_0 : \delta = 0$ Vs. $H_1 : \delta \neq 0$ for 2-Sided test

Either $H_1 : \delta > 0$ Or $H_1 : \delta < 0$ for 1-Sided test

Data File: Cancer.cyd
Trial Type: Superiority
Population ID: Drug(Treatment=2, Control=1)
Response Variable: Response
Censor: Censored (Censor Value=0, Complete=1)
Type of Test Statistic: Harrington-Fleming
Weight Parameter p: 1
Weight Parameter q: 1
Confidence Level: 0.95

Output

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | | Average Follow-up Time |
|--------------|-----------------|--------|----------|----------|---------|------------------------|
| | | Count | % | Count | % | |
| 1 | 5 | 2 | 40.000% | 3 | 60.000% | 644.4 |
| 2 | 9 | 9 | 100.000% | 0 | 0.000% | 178 |
| Total | 14 | 11 | 78.571% | 3 | 21.429% | 344.571 |

Test of Hypothesis:

| Harrington Fleming Score | Std. Error | Standardized Test Statistic | (1-Sided) | | (2-Sided) |
|--------------------------|------------|-----------------------------|-----------|---------|-----------|
| | | | Tail | p-value | p-value |
| 0.612 | 0.271 | 2.262 | G.E. | 0.012 | 0.024 |

Parameter Estimates from Cox Model:

| Hazard Ratio (HR) | ln(Hazard Ratio) (δ) | Std. Error of Estimate of δ | Wald Statistic | 95% Confidence Interval(2-Sided) | |
|-------------------|-------------------------------|------------------------------------|----------------|----------------------------------|-------------|
| | | | | Lower Limit | Upper Limit |
| 11.168 | 2.413 | 1.08 | 2.235 | 1.346 | 92.679 |

Estimated Hazard Rates:

| | |
|--------------------------------|----------|
| Control (λ_c) | 6.207E-4 |
| Treatment ($\lambda_c * HR$) | 0.007 |

East calculates 2-sided as well as 1-sided p-values. 2-sided p-value for this test is 0.024 and 1-sided p-value is 0.012. At 5% significance level, the null hypothesis is rejected.

96.5 Example: Survival Noninferiority two Samples:Logrank

Dataset: Cancer.cyx as described in section 96.2.

Purpose of the Analysis:

This section will illustrate through a worked example how to analyze data generated from a two-sample noninferiority study with a time-to-event trial endpoint. The noninferiority margin is generally determined by performing a meta-analysis on past clinical trials of the active control versus placebo. Regulatory agencies then require the sponsor of the clinical trial to demonstrate that a fixed percentage of the active control effect (usually 50%) is retained by the new treatment. In a noninferiority trial the goal is to establish that an experimental treatment is **no worse than** the standard treatment, rather than attempting to establish that it is superior. The between-treatment difference

96 Analysis-Survival Data

is expressed in terms of the hazard ratio,

$$\rho = \frac{\lambda_t}{\lambda_c},$$

or equivalently, in terms of the log hazard ratio

$$\delta = \ln(\rho) = \ln\left(\frac{\lambda_t}{\lambda_c}\right).$$

Where ρ_0 is the noninferiority margin for the hazard ratio, whereas, $\delta_0 = \ln(\rho_0)$ is the noninferiority margin for log hazard ratio.

We perform the comparison of the two treatments by testing

$$H_0: \delta \geq \delta_0$$

against the one-sided alternative

$$H_1: \delta < \delta_0,$$

when $\delta_0 (\geq 0)$

Or

$$H_0: \delta \leq \delta_0$$

against the one-sided alternative

$$H_1: \delta > \delta_0,$$

when $\delta_0 (\leq 0)$.

Analysis Steps:

1. Click **Analysis Inputs** tab on the status bar below. This will display several input fields associated with Logrank Test in the main window.
2. In the **Main** tab, select **Noninferiority** as **Trial Type**. Enter noninferiority margin as $\ln(0.511692)$ which is -0.67 . Select **Drug** in the **Population Id** field with 1 as **Control** and 2 as **Treatment**. Select **Response** as **Response** variable. Select **Censored** as **Censor** variable with **Censor value** as 0. This data does not have a frequency variable, so leave the **Frequency Variable** blank. Choose the Test Statistic **LogRank**

3. In the **Advanced** tab, leave the fields **By Variable 1** and **By Variable 2** blank. Enter 0.975 as the value of **Confidence Level**.

96 Analysis-Survival Data

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Time to Event Response: Logrank Test

Let $\delta = \ln(\lambda_1 / \lambda_c)$
 If $\delta_0 < 0$ then $H_0 : \delta \leq \delta_0$ Vs. $H_1 : \delta > \delta_0$
 If $\delta_0 > 0$ then $H_0 : \delta \geq \delta_0$ Vs. $H_1 : \delta < \delta_0$

Data File: Cancer.cyd
 Trial Type: Noninferiority
 Population ID: Drug(Treatment=2, Control=1)
 Response Variable: Response
 Censor: Censored(Censor Value=0, Complete=1)
 Confidence Level: 0.975
 Noninferiority Margin: -0.67

Output

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | |
|--------------|-----------------|--------|--------|----------|--------|
| | | Count | % | Count | % |
| 1 | 5 | 2 | 40 | 3 | 60 |
| 2 | 9 | 9 | 100 | 0 | 0 |
| Total | 14 | 11 | 78.571 | 3 | 21.429 |

Parameter Estimates:

| Hazard Ratio (HR) | 97.5% Confidence Interval(1-Sided) | |
|-------------------|------------------------------------|-------------|
| | Lower Limit | Upper Limit |
| 11.168 | 1.252 | Infinity |

Test of Hypothesis:

| Log Rank Score | Std. Error | Standardized Test Statistic | (1-Sided) | |
|----------------|------------|-----------------------------|-----------|----------|
| | | | Tail | p-value |
| 5.563 | 1.447 | 3.843 | G.E. | 6.067E-5 |

Estimated Hazard Rates:

| | |
|--------------------------------|----------|
| Control (λ_c) | 6.207E-4 |
| Treatment ($\lambda_c * HR$) | 0.007 |

With the low 1-sided p-values the noninferiority of the drug over control is established.

96.6 Example: Survival Noninferiority two Samples-Wilcoxon

Dataset: Cancer.cyd

Data Description

This data is from a small lung cancer clinical trial involving a new and control drug. The dataset has three variables **Drug**, **Response** and **Censored**.

The variable **Drug** acts as an identifier of the population to which the observation belongs. The value **1** corresponds to the control group and value **2** corresponds to the treatment group.

The **Response** variable provides survival time (in days).

The variable **Censored** gives information about which observation is censored. The value **0** corresponds to censoring and the value **1** corresponds to non-censoring.

Analysis Steps:

1. Click **Analysis Inputs** tab on the status bar below. This will display several input fields associated with Logrank Test in the main window.
2. In the **Main** tab, select **Noninferiority** as **Trial Type**. Enter noninferiority margin as $\ln(0.511692)$ which is -0.67 . Select **Drug** in the **Population Id** field with 1 as **Control** and 2 as **Treatment**. Select **Response** as **Response** variable. Select **Censored** as **Censor** variable with **Censor value** as 0. This data does not have a frequency variable, so leave the **Frequency Variable** blank. Choose the **Test Statistic** as **Wilcoxon-Gehan**.

Data Set: Cancer.cyd

Main Advanced

Trial Type: Noninferiority Response Variable: Response Frequency Variable:

Noninf. Margin (ln(HR₀)): -0.67

Population ID: Drug Censor Indicator: Censored Test Statistic: Wilcoxon-Gehan

Control: 1 Censored: 0

Treatment: 2 Complete: 1

3. In the **Advanced** tab, leave the fields **By Variable 1** and **By Variable 2** blank. Enter 0.975 as the value of **Confidence Level**.

Data Set: Cancer.cyd

Main Advanced

By Variable 1: Confidence Level: 0.975

By Variable 2:

96 Analysis-Survival Data

- Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Time to Event Response: Logrank Test

Let $\delta = \ln(\lambda_t / \lambda_c)$

If $\delta_0 < 0$ then $H_0 : \delta \leq \delta_0$ Vs. $H_1 : \delta > \delta_0$

If $\delta_0 > 0$ then $H_0 : \delta \geq \delta_0$ Vs. $H_1 : \delta < \delta_0$

Data File: Cancer.cyd
Trial Type: Noninferiority
Population ID: Drug(Treatment=2, Control=1)
Response Variable: Response
Censor: Censored (Censor Value=0, Complete=1)
Type of Test Statistic: Wilcoxon-Gehan
Confidence Level: 0.975
Noninferiority Margin: -0.67

Output

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | | Average Follow-up Time |
|--------------|-----------------|--------|----------|----------|---------|------------------------|
| | | Count | % | Count | % | |
| 1 | 5 | 2 | 40.000% | 3 | 60.000% | 644.4 |
| 2 | 9 | 9 | 100.000% | 0 | 0.000% | 178 |
| Total | 14 | 11 | 78.571% | 3 | 21.429% | 344.571 |

Test of Hypothesis:

| Wilcoxon Gehan Score | Std. Error | Standardized Test Statistic | (1-Sided) | |
|----------------------|------------|-----------------------------|-----------|----------|
| | | | Tail | p-value |
| 53.806 | 15.985 | 3.366 | G.E. | 3.812E-4 |

Parameter Estimates from Cox Model:

| Hazard Ratio (HR) | ln(Hazard Ratio) (δ) | Std. Error of Estimate of δ | Wald Statistic | 97.5% Confidence Interval(1-Sided) | |
|-------------------|-------------------------------|------------------------------------|----------------|------------------------------------|-------------|
| | | | | Lower Limit | Upper Limit |
| 11.168 | 2.413 | 1.08 | 2.856 | 1.346 | INF |

Estimated Hazard Rates:

| | |
|-------------------------------|----------|
| Control (λ_c) | 6.207E-4 |
| Treatment (λ_c * HR) | 0.007 |

With the low 1-sided p-values the noninferiority of the drug over control is established.

96.7 Example: Survival Noninferiority two Samples:Harrington-Fleming

Dataset: Cancer.cyx

Data Description

This data is from a small lung cancer clinical trial involving a new and control drug. The dataset has three variables **Drug**, **Response** and **Censored**.

The variable **Drug** acts as an identifier of the population to which the observation belongs. The value **1** corresponds to the control group and value **2** corresponds to the treatment group.

The **Response** variable provides survival time (in days).

The variable **Censored** gives information about which observation is censored. The value **0** corresponds to censoring and the value **1** corresponds to non-censoring.

Analysis Steps:

1. Click **Analysis Inputs** tab on the status bar below. This will display several input fields in the main window.
2. In the **Main** tab, select **Noninferiority** as **Trial Type**. Enter noninferiority margin as $\ln(0.511692)$ which is -0.67 . Select **Drug** in the **Population Id** field

96 Analysis-Survival Data

with 1 as **Control** and 2 as **Treatment**. Select **Response** as **Response** variable. Select **Censored** as **Censor** variable with **Censor value** as 0. This data does not have a frequency variable, so leave the **Frequency Variable** blank. Choose the **Test Statistic** as **Harrington-Fleming**.

Data Set: Cancer.cyd

Main Advanced

Trial Type: Noninferiority Response Variable: Response Frequency Variable:

Noninf. Margin (ln(HR₀)): -0.67

Population ID: Drug Censor Indicator: Censored Test Statistic: Harrington-Fleming

Control: 1 Censored: 0

Treatment: 2 Complete: 1

- In the **Advanced** tab, leave the fields **By Variable 1** and **By Variable 2** blank. Enter 0.975 as the value of **Confidence Level**.

Data Set: Cancer.cyd

Main Advanced

By Variable 1: Confidence Level: 0.975

By Variable 2:

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Time to Event Response: Logrank Test

Let $\delta = \ln(\lambda_t / \lambda_c)$

If $\delta_0 < 0$ then $H_0 : \delta \leq \delta_0$ Vs. $H_1 : \delta > \delta_0$

If $\delta_0 > 0$ then $H_0 : \delta \geq \delta_0$ Vs. $H_1 : \delta < \delta_0$

Data File: Cancer.cyd
Trial Type: Noninferiority
Population ID: Drug(Treatment=2, Control=1)
Response Variable: Response
Censor: Censored (Censor Value=0, Complete=1)
Type of Test Statistic: Harrington-Fleming
Weight Parameter p: 1
Weight Parameter q: 1
Confidence Level: 0.975
Noninferiority Margin: -0.67

Output

96 Analysis-Survival Data

Summary of Observed Data:

| Treatment ID | No. of Subjects | Events | | Censored | | Average Follow-up Time |
|--------------|-----------------|--------|----------|----------|---------|------------------------|
| | | Count | % | Count | % | |
| 1 | 5 | 2 | 40.000% | 3 | 60.000% | 644.4 |
| 2 | 9 | 9 | 100.000% | 0 | 0.000% | 178 |
| Total | 14 | 11 | 78.571% | 3 | 21.429% | 344.571 |

Test of Hypothesis:

| Harrington Fleming Score | Std. Error | Standardized Test Statistic | (1-Sided) | |
|--------------------------|------------|-----------------------------|-----------|----------|
| | | | Tail | p-value |
| 0.804 | 0.241 | 3.334 | G.E. | 4.273E-4 |

Parameter Estimates from Cox Model:

| Hazard Ratio (HR) | ln(Hazard Ratio) (δ) | Std. Error of Estimate of δ | Wald Statistic | 97.5% Confidence Interval(1-Sided) | |
|-------------------|-------------------------------|------------------------------------|----------------|------------------------------------|-------------|
| | | | | Lower Limit | Upper Limit |
| 11.168 | 2.413 | 1.08 | 2.856 | 1.346 | INF |

Estimated Hazard Rates:

| | |
|--------------------------------|----------|
| Control (λ_c) | 6.207E-4 |
| Treatment ($\lambda_e * HR$) | 0.007 |

With the low 1-sided p-values the noninferiority of the drug over control is established.

96.8 Example: Survival Multi-arm-Kaplan Meier Estimator

Dataset: Cancer.cydx as described in section 96.2.

Purpose of the Analysis:

The Kaplan-Meier estimator also known as the product limit estimator is an estimator for estimating the survival function from lifetime data. In medical research, it is often used to measure the fraction of patients living for a certain amount of time after treatment.

A plot of the Kaplan-Meier estimate of the survival function is a series of horizontal steps of declining magnitude which, when a large enough sample is taken, approaches the true survival function for that population. The value of the survival function between successive distinct sampled observations is assumed to be constant.

An important advantage of the Kaplan-Meier estimator is that the method can take into account some types of censored data, particularly right-censoring, which occurs if a patient withdraws from a study, that is, lost from the sample before the final outcome is observed.

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

Analysis > (Events) Explore > (Multi-Arm Design) Kaplan Meier

This will display several input fields associated with Kaplan Meier Test in the main window.

3. In the **Main** tab, select **Drug** as **Population ID**, **Response** as **Response** variable. Select **Censored** as **Censor** variable with **Censor value** as 0. Leave the **Frequency Variable** field blank.

Analysis: Survival Endpoint: Explore Test - Multi-Arm Design - Kaplan-Meier

Data Set: Cancer.cyd

Main Advanced

Population Id: Response: Frequency Variable:

Censor:

Censor Value:

Complete:

96 Analysis-Survival Data

4. Click **OK** to start the analysis. The output is displayed in the main window.

Analysis: Time to Event Response: Kaplan-Meier

Input Parameters

Data File: Cancer.cyd
 Population Variable: Drug
 Response Variable: Response
 Censor Variable: Censored(Censor Value=0, Complete=1)
 Confidence Level: 0.95

Output

Response variable: Response
 Total Number of Records: 14
 Number of Records Rejected: 0
 Summary:

| Drug | No. of Subjects | Events | | Censored | |
|------|-----------------|--------|-----|----------|----|
| | | Count | % | Count | % |
| 1 | 5 | 2 | 40 | 3 | 60 |
| 2 | 9 | 9 | 100 | 0 | 0 |

Quartiles for Survival Time (Using Interpolation):

| Drug | Q1 | | | Q2 (Median) | | | Q3 | | |
|------|----------|----------|----------|-------------|---------|----------|-------|--------|----------|
| | Value | 95% CI | | Value | 95% CI | | Value | 95% CI | |
| | | Lower | Upper | | Lower | Upper | | Lower | Upper |
| 1 | 1344.667 | 1344.667 | 1344.667 | 910.333 | 286.657 | 1647.561 | 355 | -INF | 1213.228 |
| 2 | 253.25 | 69.462 | INF | 103.5 | 6.313 | 268.809 | 9 | -INF | 199.759 |

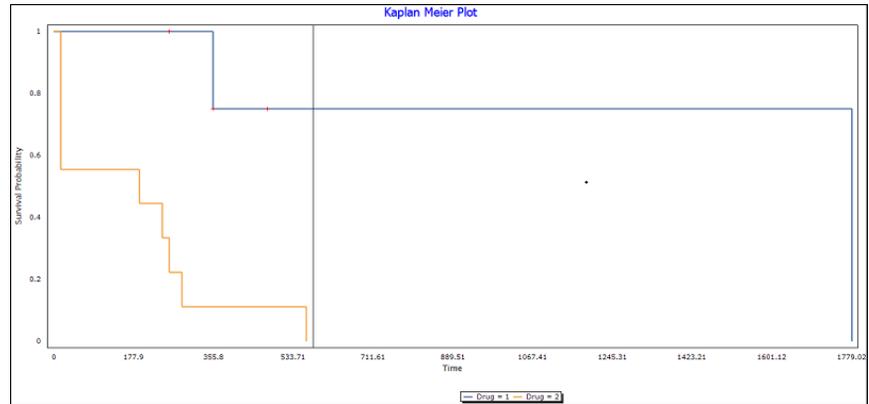
Quartiles for Survival Time (Using Actual Data):

| Drug | Q1 | | | Q2 (Median) | | | Q3 | | |
|------|-------|--------|-------|-------------|--------|-------|-------|--------|-------|
| | Value | 95% CI | | Value | 95% CI | | Value | 95% CI | |
| | | Lower | Upper | | Lower | Upper | | Lower | Upper |
| 1 | 1779 | 355 | 1779 | 1779 | 355 | 1779 | 355 | 355 | 1779 |
| 2 | 257 | 16 | 563 | 191 | 16 | 285 | 16 | 16 | 191 |

Failure Times:

| Drug | Time (t) | No. of Subjects (n) | Events (d) | Censored (c) | Survival (S(t)) | Var (S(t)) |
|------|----------|---------------------|------------|--------------|-----------------|------------|
| 1 | 0 | 5 | 0 | 0 | 1 | 0 |
| 1 | 257 | 5 | 0 | 1 | 1 | 0 |
| 1 | 355 | 4 | 1 | 1 | 0.75 | 0.047 |
| 1 | 476 | 2 | 0 | 1 | 0.75 | 0.047 |
| 1 | 1779 | 1 | 1 | 0 | 0 | NA |
| 2 | 0 | 9 | 0 | 0 | 1 | 0 |
| 2 | 16 | 9 | 4 | 0 | 0.556 | 0.027 |
| 2 | 191 | 5 | 1 | 0 | 0.444 | 0.027 |
| 2 | 242 | 4 | 1 | 0 | 0.333 | 0.025 |
| 2 | 257 | 3 | 1 | 0 | 0.222 | 0.019 |
| 2 | 285 | 2 | 1 | 0 | 0.111 | 0.011 |
| 2 | 563 | 1 | 1 | 0 | 0 | NA |

A node **Analysis: Time to Event Response:Kaplan-Meier1** is created in the Library. Also a sub-node **Kaplan-Meier Plot1** is created in the Library. Click the **Kaplan-Meier Plot1** node to open the plot.



Note that in this plot, the estimated survivals are plotted for both the drugs on the same time axis, so that comparison of survivals is possible. The Kaplan-Meier Plot indicates that the patients on Drug arm have better survival as compared with those on the control arm.

97 *Analysis-Multiple Comparison Procedures for Survival Data*

It is often the case that multiple objectives are to be addressed in one single trial. These objectives are formulated into a family of hypotheses. Type I error rate is inflated when one considers the inferences together as a family. Failure to compensate for multiplicities can have adverse consequences. For example, a drug could be approved when actually it is not better than placebo. Multiple comparison (MC) procedures provide a guard against inflation of type I error due to multiple testing. Probability of making at least one type I error is known as family wise error rate (FWER). **East** supports several parametric and p-value based MC procedures.

We have seen how to simulate survival data under different MC procedures in chapter 51. This chapter explains how to analyze survival data with different MC procedures available in **East**.

97.1 Available Procedures

The probability of making at least one type I error is known as family wise error rate (FWER). All the MC procedures available in **East** strongly control FWER. Strong control of FWER refers to preserving the probability of incorrectly claiming at least one null hypothesis. To contrast strong control with weak control of FWER, the latter controls the FWER under the assumption that all hypotheses are true.

The following MC procedures are available for survival endpoints in East.

| Category | Procedure | Reference |
|---------------|-------------------------------|------------------------------------|
| P-value Based | Bonferroni | Bonferroni CE (1935, 1936) |
| | Sidak | Sidak Z (1967) |
| | Weighted Bonferroni | Benjamini Y and Hochberg Y (1997) |
| | Holm’s Step Down | Holm S (1979) |
| | Hochberg’s Step Up | Hochberg Y (1988) |
| | Hommel’s Step Up | Hommel G (1988) |
| | Fixed Sequence | Westfall PH, Krishen A (2001) |
| Fallback | Wiens B, Dimitrienko A (2005) | |

East provides three types of test statistics for the analysis of survival data incorporating MC procedures, which include the Logrank, Wilcoxon-Gehan, and the Harrington-Fleming. For illustration purposes, the examples below will only utilize the Logrank test statistic for data analysis.

STAMPEDE Trial

Throughout this chapter we consider the data derived from the design of the STAMPEDE trial discussed in chapter 51 to illustrate the analysis of survival data under different MC procedures. The STAMPEDE study is an ongoing, open-label, 5-stage, 6-arm randomized controlled trial using multi-arm, multi-stage (MAMS) methodology for men with prostate cancer. Started in 2005, it was the first trial of this design to use multiple arms and stages synchronously. The study population consists of men with high-risk localized or metastatic prostate cancer, who are being treated for the first time with long-term androgen deprivation therapy (ADT) or androgen suppression. The study started with 5 treatment groups:

- Standard of care (SOC) = ADT
- SOC + zoledronic acid (IV)
- SOC + docetaxel (IV)
- SOC + celecoxib, an orally administered cox-2 inhibitor
- SOC + zoledronic acid + docetaxel
- SOC + zoledronic acid + celecoxib

We want to control the FWER at 5% level of significance.

Dataset: The data to be used for the examples below arise from the STAMPEDE design described in chapter 51. The resulting **SubjectData** was generated during a

97 Analysis-Multiple Comparison Procedures for Survival Data

design simulation that captured subject level data for every simulation run:

Design: Survival Endpoint: Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - L

Number of Arms: Include Options

Test Parameters | Response Generation | **Accrual / Dropouts** | Randomization | Simulation Controls

Number of Simulations:
 Refresh Frequency:
 Random Number Seed
 Clock
 Fixed
 Suppress All Intermediate Output
 Pause after Refresh
 Pause at End

Output Options
 Output Type:
 Save summary statistics for every simulation run
 Save subject level data for simulation runs
 Note: Max. 100,000 records will be saved.
 Output Time Unit:

Simulate

SurvivalTime_Years: 3390 Value: 0.691949600935616

| | SimulationID | SubjectID | ArrivalTime_ | TreatmentID | SurvivalTim | CensorInd |
|----|--------------|-----------|--------------|-------------|-------------|-----------|
| 1 | 2 | 1 | 000583965331 | 5 | 33.1929124 | 0 |
| 2 | 2 | 2 | .00135622639 | 5 | 6.59749359 | 1 |
| 3 | 2 | 3 | .00147017278 | 0 | 5.75658262 | 1 |
| 4 | 2 | 4 | .00765839722 | 0 | 9.06586646 | 0 |
| 5 | 2 | 5 | .00914330846 | 1 | 25.7907973 | 0 |
| 6 | 2 | 6 | 0.01389929 | 0 | 5.45830969 | 1 |
| 7 | 2 | 7 | 0.0141818587 | 4 | 1.03824813 | 1 |
| 8 | 2 | 8 | 0.0142699918 | 2 | 10.431085 | 0 |
| 9 | 2 | 9 | 0.0145352175 | 3 | 19.4443065 | 0 |
| 10 | 2 | 10 | 0.0170551711 | 4 | 6.88579751 | 1 |
| 11 | 2 | 11 | 0.0186827979 | 1 | 1.65168043 | 1 |
| 12 | 2 | 12 | 0.0228972737 | 0 | 0.195713353 | 1 |
| 13 | 2 | 13 | 0.0229346178 | 1 | 15.0375844 | 0 |
| 14 | 2 | 14 | 0.0231156575 | 0 | 2.16785289 | 1 |
| 15 | 2 | 15 | 0.0249825242 | 1 | 9.47944529 | 0 |
| 16 | 2 | 16 | 0.0287036645 | 5 | 7.96413251 | 0 |
| 17 | 2 | 17 | 0.0307711842 | 1 | 9.79368567 | 0 |
| 18 | 2 | 18 | 0.0356241793 | 5 | 8.60429448 | 0 |
| 19 | 2 | 19 | 0.038857447 | 0 | 4.92766474 | 1 |

97.2 Single step MC procedures

East supports three p-value based single step MC procedures:

- Bonferroni procedure
- Sidak procedure and
- Weighted Bonferroni procedure.

For the Bonferroni procedure, H_i is rejected if $p_i < \frac{\alpha}{k-1}$ and the adjusted p-value is given as $\min(1, (k-1)p_i)$.

For the Sidak procedure, H_i is rejected if $p_i < 1 - (1 - \alpha)^{\frac{1}{k-1}}$ and the adjusted p-value is given as $1 - (1 - p_i)^{k-1}$.

For the weighted Bonferroni procedure, H_i is rejected if $p_i < w_i\alpha$ and the adjusted p-value is given as $\min(1, \frac{p_i}{w_i})$. Here w_i denotes the proportion of α allocated to the H_i such that $\sum_{i=1}^{k-1} w_i = 1$. Note that, if $w_i = \frac{1}{k-1}$, then the Bonferroni procedure is reduced to the regular Bonferroni procedure.

Example: Bonferroni procedure

Select the **SubjectData** node under the appropriate **Simulation** node in the **Library**. Next, under the **Analysis** tab in the **Events** group, select **Many Samples - Pairwise Comparisons to Control - Logrank**. The following screen is displayed:

The screenshot displays the configuration window for a survival analysis test. The title bar reads "Analysis: Survival Endpoint: Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Logrank Test". The "Data Set" is "SubjectData". The "Advanced" tab is selected. Under "Input", "Raw Subject Data" is chosen. The "Test Statistic" is set to "Logrank". Under "Multiple Comparison Procedure", "Select MCP:" is set to "Bonferroni". Other fields include "Treatment Variable:", "Control Arm:", "Response Variable:", and "Censor:", all of which are currently empty.

97 Analysis-Multiple Comparison Procedures for Survival Data

Select the following values for the **Main** tab:

Analysis: Survival Endpoint: Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Logrank Test

Data Set: SubjectData

Main Advanced

Input
 Raw Subject Data Raw p-values

Treatment Variable: TreatmentID

Control Arm: 0

Response Variable: SurvivalTime_Years

Censor: CensorInd

Censor Value: 0

Complete: 1

Frequency:

Test Statistic: Logrank

Multiple Comparison Procedure
Select MCP: Bonferroni

No. of Treatment Arms: 5

OK

Keep the following default values for the **Advanced** tab:

Analysis: Survival Endpoint Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Logrank Test

Data Set: SubjectData

Main Advanced

By Variable 1: Confidence Level:

By Variable 2: Rejection Region:

OK

Click **OK** to analyze the data. The output will be displayed in the main window.

Analysis: Multiple Pairwise Comparisons with Control: Logrank Test

$H_i: \lambda_i / \lambda_0 \leq 1$ Vs. $K_i: \lambda_i / \lambda_0 > 1$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Data File: SubjectData
 Treatment Variable: TreatmentID (Control=0)
 Response Variable: SurvivalTime_Years
 Censor Indicator Variable: CensorInd (Censor Value=0, Complete=1)
 Type of Test Statistic: Logrank
 Confidence Level: 0.95
 Rejection Region: Right-Tail
 Multiple Comparison Procedure: Bonferroni

Output

Summary of Observed Data:

| Arm | No. of Subjects | Events | | Censored | |
|------|-----------------|--------|---------|----------|---------|
| | | Count | % | Count | % |
| Ctrl | 971 | 392 | 40.371% | 579 | 59.629% |
| 1 | 486 | 159 | 32.716% | 327 | 67.284% |
| 2 | 486 | 165 | 33.951% | 321 | 66.049% |
| 3 | 486 | 160 | 32.922% | 326 | 67.078% |
| 4 | 485 | 156 | 32.165% | 329 | 67.835% |
| 5 | 486 | 168 | 34.568% | 318 | 65.432% |

Inference for Each Treatment Level Versus Control:

| Arm | Hazard Rate | Hazard Ratio (HR) | Logrank Score | Std. Error | Standardized Test Statistic | p-value | | 95% Confidence Interval for HR (2-Sided) | |
|------|-------------|-------------------|---------------|------------|-----------------------------|---------|----------|--|-------------|
| | | | | | | Raw | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 0.056 | NA | NA | NA | NA | NA | NA | NA | NA |
| 1 | 0.043 | 0.752 | -33.99 | 11.193 | -3.037 | 0.999 | 1 | 0.626 | 0.904 |
| 2 | 0.045 | 0.792 | -28.217 | 11.23 | -2.513 | 0.994 | 1 | 0.661 | 0.95 |
| 3 | 0.043 | 0.762 | -32.573 | 11.191 | -2.911 | 0.998 | 1 | 0.634 | 0.915 |
| 4 | 0.042 | 0.738 | -35.979 | 11.164 | -3.223 | 0.999 | 1 | 0.613 | 0.888 |
| 5 | 0.046 | 0.814 | -25.675 | 11.25 | -2.282 | 0.989 | 1 | 0.676 | 0.971 |

97.2 Single step MCM procedures

Adjusted Global p-value: 1
 Total Number of Records: 3400
 Number of Records Rejected: 0
 Total Number of Arms (k): 6

97 Analysis-Multiple Comparison Procedures for Survival Data

The adjusted p-values for comparison of Dose1, Dose 2 ... up to Dose 5 vs. Placebo are all essentially 1. Therefore, after multiplicity adjustment according to Bonferroni procedure for this design, we can conclude that no additional treatment in addition to the standard of care at the tested dose levels is significantly different from the current standard treatment (ADT only).

Example: Sidak procedure

Again with the appropriate **SubjectData** node selected, under the **Analysis** tab in the **Events** group, select **Many Samples - Pairwise Comparisons to Control - Logrank**.

Select the following values for the **Main** tab:

Analysis: Survival Endpoint: Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Logrank Test

Data Set: SubjectData

Main Advanced

Input
 Raw Subject Data Raw p-values

Treatment Variable: TreatmentID
 Control Arm: 0

Response Variable: SurvivalTime_Years

Censor: CensorInd
 Censor Value: 0
 Complete: 1

Frequency:

Test Statistic: Logrank

Multiple Comparison Procedure
 Select MCP: Sidak
 No. of Treatment Arms: 5

OK

Keep the default values for the **Advanced** tab and click **OK** to analyze the data. The

output will be displayed in the main window.

Analysis: Multiple Pairwise Comparisons with Control: Logrank Test

$H_i : \lambda_i / \lambda_0 \leq 1$ Vs. $K_i : \lambda_i / \lambda_0 > 1$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Data File: SubjectData
 Treatment Variable: TreatmentID (Control=0)
 Response Variable: SurvivalTime_Years
 Censor Indicator Variable: CensorInd (Censor Value=0, Complete=1)
 Type of Test Statistic: Logrank
 Confidence Level: 0.95
 Rejection Region: Right-Tail
 Multiple Comparison Procedure: Sidak

Output

Summary of Observed Data:

| Arm | No. of Subjects | Events | | Censored | |
|------|-----------------|--------|---------|----------|---------|
| | | Count | % | Count | % |
| Ctrl | 971 | 392 | 40.371% | 579 | 59.629% |
| 1 | 486 | 159 | 32.716% | 327 | 67.284% |
| 2 | 486 | 165 | 33.951% | 321 | 66.049% |
| 3 | 486 | 160 | 32.922% | 326 | 67.078% |
| 4 | 485 | 156 | 32.165% | 329 | 67.835% |
| 5 | 486 | 168 | 34.568% | 318 | 65.432% |

Inference for Each Treatment Level Versus Control:

| Arm | Hazard Rate | Hazard Ratio (HR) | Logrank Score | Std. Error | Standardized Test Statistic | p-value | | 95% Confidence Interval for HR (2-Sided) | |
|------|-------------|-------------------|---------------|------------|-----------------------------|---------|----------|--|-------------|
| | | | | | | Raw | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 0.056 | NA | NA | NA | NA | NA | NA | NA | NA |
| 1 | 0.043 | 0.752 | -33.99 | 11.193 | -3.037 | 0.999 | 1 | 0.626 | 0.904 |
| 2 | 0.045 | 0.792 | -28.217 | 11.23 | -2.513 | 0.994 | 1 | 0.661 | 0.95 |
| 3 | 0.043 | 0.762 | -32.573 | 11.191 | -2.911 | 0.998 | 1 | 0.634 | 0.915 |
| 4 | 0.042 | 0.738 | -35.979 | 11.164 | -3.223 | 0.999 | 1 | 0.613 | 0.888 |
| 5 | 0.046 | 0.81 | -25.673 | 11.25 | -2.282 | 0.989 | 1 | 0.676 | 0.971 |

Adjusted Global p-value: 1
 Total Number of Records: 3400
 Number of Records Rejected: 0
 Total Number of Arms (k): 6

The adjusted p-values for comparison of Dose1, Dose 2 ... up to Dose 5 vs. Placebo are all essentially 1. Therefore, after multiplicity adjustment according to Sidak procedure for this design, we can conclude that no additional treatment in addition to the standard of care at the tested dose levels is significantly different from the current standard treatment (ADT only).

Example: Weighted Bonferroni procedure Dataset:

97 Analysis-Multiple Comparison Procedures for Survival Data

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below.

Analysis: Continuous Endpoint Many-Sample Test - Multiple Comparisons Design - Pairwise Comparisons to Control - Differences... .00 .00

Data Set: Hypertension-trial.cyd

Multiple Comparison Procedure

Select MCP:

No. of Treatment Arms:

Table of Treatment specific parameters:

| Arm | Proportion of... |
|-----|------------------|
| D1 | 0.25 |
| D2 | 0.25 |
| D3 | 0.25 |
| D4 | 0.25 |

4. Upon selection of weighted Bonferroni procedure, a table will appear under the drop-down box. The table has two columns - **Arm** and **Proportion of Alpha**. In the column **Proportion of Alpha**, you have to specify the proportion of total alpha you want to spend in each test. Ideally, the values in this column should add up to 1; if not, then **East** will normalize it to add them up to 1. By default, **East** distributes the total alpha equally among all tests. Here we have 4 tests in total, therefore each of the tests have proportion of alpha as $1/4$ or 0.25 . You can specify other proportions as well. For this example, keep the equal proportion of alpha for each test.
5. Click **OK** to analyze the data. The output will be displayed in the main window

once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis

$H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: UnEqual
 Multiple Comparison Procedure: Weighted Bonferroni
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail
 Table For Proportion of Alpha:

| Arms | Proportion of Alpha |
|------|---------------------|
| D1 | 0.25 |
| D2 | 0.25 |
| D3 | 0.25 |
| D4 | 0.25 |

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 1 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.031 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.045 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 0 | 0.001 | 3.585 | INF |

Adjusted Global p-value: 0.001
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 0.982, 0.031, 0.044 and 0.001, respectively. Therefore, after multiplicity adjustment according to Sidak procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

Notice that the adjusted p-values in weighted Bonferroni MC procedure and the simple Bonferroni procedures are identical. This is because the weighted Bonferroni procedure with equal proportion reduces to the simple Bonferroni procedure.

97 Analysis-Multiple Comparison Procedures for Survival Data

97.3 Step down MC procedure

In the single step MC procedures, the decision to reject any hypothesis does not depend on the decision to reject other hypotheses. On the other hand, in the stepwise procedures decision of one hypothesis test can influence the decisions on the other tests of hypotheses. There are two types of stepwise procedures. One type of procedures proceeds in data-driven order. The other type proceeds in a fixed order set a priori. Stepwise tests in a data-driven order can proceed in step-down or step-up manner. **East** supports Holm step-down MC procedure which start with the most significant comparison and continue as long as tests are significant until the test for certain hypothesis fails. The testing procedure stops at the first time a non-significant comparison occurs and all remaining hypotheses will be retained. In i -th step, $H_{(k-i)}$ is rejected if $p_{(k-i)} \leq \frac{\alpha}{i}$ and go to the next step.

Example: Holm's step-down

Dataset:

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below.

4. Click **OK** to analyze the data. The output will be displayed in the main window

once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis
 $H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: UnEqual
 Multiple Comparison Procedure: Holm's step down
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 0.634 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.023 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.023 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 0 | 0.001 | 3.585 | INF |

Adjusted Global p-value: 0.001
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 0.634, 0.023, 0.023 and 0.001, respectively. Therefore, after multiplicity adjustment according to Holm's step-down procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

97.4 Data-driven step-up MC procedures

newline Step-up tests start with the least significant comparison and continue as long as tests are not significant until the first time when a significant comparison occurs and all remaining hypotheses will be rejected. East supports two such MC procedures - Hochberg step-up and Hommel step-up procedures. In the Hochberg step-up procedure, in i -th step $H_{(k-i)}$ is retained if $p_{(k-i)} > \frac{\alpha}{i}$. In the Hommel step-up procedure, in i -th step $H_{(k-i)}$ is retained if $p_{(k-j)} > \frac{i-j+1}{i} \alpha$ for $j = 1, \dots, i$. Fixed sequence test and fallback test are the types of tests which proceed in a prespecified order.

97 Analysis-Multiple Comparison Procedures for Survival Data

Example: Hochberg's step-up procedure newline **Dataset:**

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below.

The screenshot shows the 'Multiple Comparison Procedure' dialog box with the 'Main' tab selected. The 'Data Set' is 'Hypertension-trial.cyd'. Under 'raw Data', the 'raw Data' radio button is selected. The 'Treatment Variable and its Control Arm' section has 'Dose' selected in the dropdown and 'P' selected with a radio button. The 'Response Variable' section has 'Response' selected in the dropdown. At the bottom, the 'Unequal Variance' radio button is selected. On the right, the 'Multiple Comparison Procedure' section has 'Hochberg's step up' selected in the dropdown.

4. Click **OK** to analyze the data. The output will be displayed in the main window

once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis
 $H_1 : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: UnEqual
 Multiple Comparison Procedure: Hochberg's step Up
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 0.634 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.022 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.022 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 0 | 0.001 | 3.585 | INF |

Adjusted Global p-value: 0.001
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 0.634, 0.022, 0.022 and 0.001, respectively. Therefore, after multiplicity adjustment according to Hochberg's step-up procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

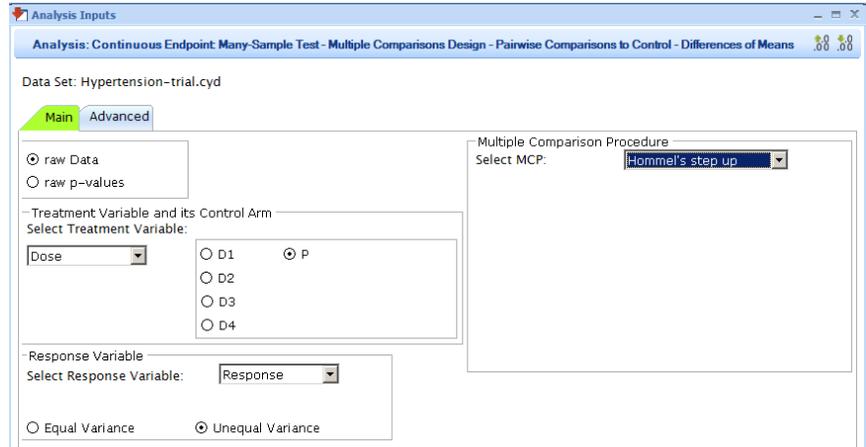
Example: Hommel's step-up procedure newline **Dataset:**

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown

97 Analysis-Multiple Comparison Procedures for Survival Data

below.



4. Click **OK** to analyze the data. The output will be displayed in the main window

once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

$H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: UnEqual
 Multiple Comparison Procedure: Hommel's step up
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|----------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 0.634 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.017 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.022 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 3.196E-4 | 0.001 | 3.585 | INF |

Adjusted Global p-value: 0.001
 Total Number of Records: 130
 Number of Records Rejected: 5
 Total Number of Arms (k): 5

The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 0.634, 0.017, 0.022 and 0.001, respectively. Therefore, after multiplicity adjustment according to Hommel's step-up procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

97.5 Fixed-sequence stepwise MC procedures

In data-driven stepwise procedures, we don't have any control on the order of the hypotheses to be tested. However, sometimes based on our preference or prior knowledge we might want to fix the order of tests a priori. Fixed sequence test and fallback test are the types of tests which proceed in a pre-specified order. East supports both these procedures.

Assume that H_1, H_2, \dots, H_{k-1} are ordered hypotheses and the order is pre-specified so that H_1 is tested first followed by H_2 and so on. Let p_1, p_2, \dots, p_{k-1} be the associated raw marginal p-values. In the fixed sequence testing procedure, for $i = 1, \dots, k - 1$, in i -th step, if $p_i < \alpha$, reject H_i and go to the next step; otherwise

97 Analysis-Multiple Comparison Procedures for Survival Data

retain H_i, \dots, H_{k-1} and stop.

Fixed sequence testing strategy is optimal when early tests in the sequence have largest treatment effect and performs poorly when early hypotheses have small treatment effect or are nearly true (Westfall and Krishen (2001)). The drawback of fixed sequence test is that once a hypothesis is not rejected no further testing is permitted. This will lead to lower power to reject hypotheses tested later in the sequence.

Fallback test alleviates the above undesirable feature for fixed sequence test. Let w_i be the proportion of α for testing H_i such that $\sum_{i=1}^{k-1} w_i = 1$. In the fixed sequence testing procedure, in i -th step ($i = 1, \dots, k - 1$), test H_i at $\alpha_i = \alpha_{i-1} + \alpha w_i$ if H_{i-1} is rejected and at $\alpha_i = \alpha w_i$ if H_{i-1} is retained. If $p_i < \alpha_i$, reject H_i ; otherwise retain it. Unlike the fixed sequence testing approach, the fallback procedure can continue testing even if a non-significant outcome is encountered by utilizing the fallback strategy. If a hypothesis in the sequence is retained, the next hypothesis in the sequence is tested at the level that would have been used by the weighted Bonferroni procedure. With $w_1 = 1$ and $w_2 = \dots = w_{k-1} = 0$, the fallback procedure simplifies to fixed sequence procedure.

Example: Fixed sequence testing procedure Dataset:

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:

Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means

3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below. Upon selection of **Fixed Sequence** procedure, a table will appear under the drop-down box. The table has two columns - **Arm** and **Test Sequence**. In the column **Test Sequence**, you have to specify the order in which the hypotheses will be tested. Specify 1 for the arm that will be compared first with Placebo, 2 for the arm that will be compared next and so on. By default **East** specifies 1 to the first arm, 2 to the second arm and so on. This default order implies that Dose1 will be compared first with Placebo, then Dose2 will be compared followed by comparison of Dose3 vs. Placebo and finally Dose 4 will be compared with Placebo. However, if we believe that efficacy of drug increases with dose, then the dose groups should be compared in descending order of dose. Therefore, specify 4, 3, 2 and 1 in column **Test Sequence** for D1,

D2, D3 and D4, respectively. This order implies that Dose4 will be compared first with Placebo, then Dose3 will be compared followed by comparison of Dose2 vs. Placebo and finally Dose 1 will be compared with Placebo.

Data Set: Hypertension-trial.cyd

Main
Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
 Select Treatment Variable:

D1 P
 D2
 D3
 D4

Response Variable
 Select Response Variable:

Equal Variance Unequal Variance

Multiple Comparison Procedure
 Select MCP:

No. of Treatment Arms:

Table of Treatment specific parameters:

| Arm | Test Sequence |
|-----|---------------|
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 4 |

Click **OK** to analyze the data. The output will be displayed in the main window

97 Analysis-Multiple Comparison Procedures for Survival Data

once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis
 $H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: UnEqual
 Multiple Comparison Procedure: Fixed Sequence
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail
 Table For Proportion of Alpha:

| Arms | Test Sequence |
|------|---------------|
| D1 | 4 |
| D2 | 3 |
| D3 | 2 |
| D4 | 1 |

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 0.634 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.011 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.011 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 0 | 0 | 3.585 | INF |

Adjusted Global p-value: 0
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The input section of the output displays the tests sequence along with the other input values we have provided. The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs Placebo and Dose4 vs. Placebo are 0.634, 0.011, 0.011 and 0.000, respectively. Therefore, after multiplicity adjustment according to fixed sequence procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

Example; Fallback procedure

Dataset:

Analysis Steps:

1. Open the dataset from **Samples** folder.
2. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
3. In the ensuing dialog box, under the **Main** tab choose the variables as shown below. Upon selection of **Fallback** procedure, a table will appear under the drop-down box. The table has three columns - **Arm**, **Proportion of Alpha** and **Test Sequence**. Specify **4, 3, 2** and **1** in column **Test Sequence** for D1, D2, D3 and D4, respectively. For this example, keep the equal proportion of alpha for each test in the column **Proportion of Alpha**.

Data Set: Hypertension-trial.cyd

Main Advanced

raw Data
 raw p-values

Treatment Variable and its Control Arm
 Select Treatment Variable: Dose

D1 P
 D2
 D3
 D4

Response Variable
 Select Response Variable: Response

Equal Variance Unequal Variance

Multiple Comparison Procedure
 Select MCP: Fallback

No. of Treatment Arms: 4

Table of Treatment specific parameters:

| Arm | Proportion of... | Test Sequence |
|-----|------------------|---------------|
| 1 | 0.25 | 4 |
| 2 | 0.25 | 3 |
| 3 | 0.25 | 2 |
| 4 | 0.25 | 1 |

4. Click **OK** to analyze the data. The output will be displayed in the main window

97 Analysis-Multiple Comparison Procedures for Survival Data

once the analysis is over.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis
 $H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: Hypertension-trial.cyd
 Treatment Variable: Dose
 Response Variable: Response
 Control Arm: P
 Variance Type: UnEqual
 Multiple Comparison Procedure: Fallback
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Table For Proportion of Alpha:

| Arms | Proportion of Alpha | Test Sequence |
|------|---------------------|---------------|
| D1 | 0.25 | 4 |
| D2 | 0.25 | 3 |
| D3 | 0.25 | 2 |
| D4 | 0.25 | 1 |

Output

Inference for each Treatment Level versus Control:

| Arm | Sample Size | Mean | Std. Err. of Diff. of Means | Test Statistic | p-value | | 95% Confidence Interval of Diff. of Means(One sided) | |
|------|-------------|-------|-----------------------------|----------------|---------|----------|--|-------------|
| | | | | | Naive | Adjusted | Lower Limit | Upper Limit |
| Ctrl | 25 | 0.704 | NA | NA | NA | NA | NA | NA |
| D1 | 24 | 0.008 | 2.026 | -0.343 | 0.634 | 0.634 | -4.097 | INF |
| D2 | 26 | 5.254 | 1.815 | 2.507 | 0.008 | 0.022 | 1.506 | INF |
| D3 | 24 | 5.629 | 2.081 | 2.367 | 0.011 | 0.022 | 1.43 | INF |
| D4 | 26 | 7.331 | 1.814 | 3.654 | 0 | 0.001 | 3.585 | INF |

Adjusted Global p-value: 0.001
 Total no. of records: 130
 No. of records rejected: 5
 Total Number of Arms (k): 5

The input section of the output displays the tests sequence along with the other input values we have provided. The adjusted p-values for comparison of Dose1 vs. Placebo, Dose2 vs. Placebo, Dose3 vs. Placebo and Dose4 vs. Placebo are 0.634, 0.022, 0.022 and 0.001, respectively. Therefore, after multiplicity adjustment according to fallback procedure, we can conclude that Dose2, Dose3 and Dose4 are significantly different from Placebo, but Dose1 is not significantly different from Placebo at 5% level of significance.

97.6 Example: Raw p-values as input

Suppose we don't have the dataset containing all the observations, rather we have the raw p-values and we want to adjust these using Bonferroni procedure. Here we will consider the 4 raw p-values returned by East using the example STAMPEDE data in all the above output. These p-values are 0.634, 0.008, 0.011 and 0.000. We will use

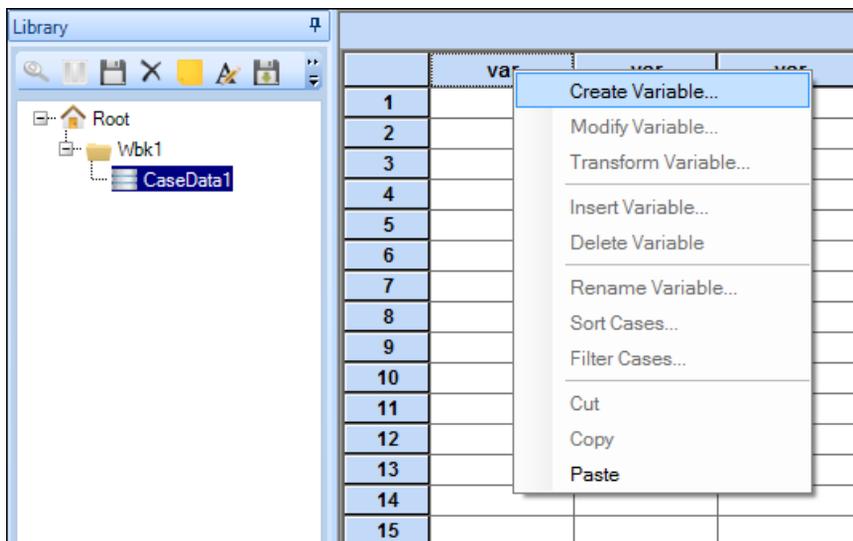
these raw p-values to obtain adjusted p-values. In order to do this, first, we need to create a dataset containing these p-values.

Dataset: New Dataset to be created.

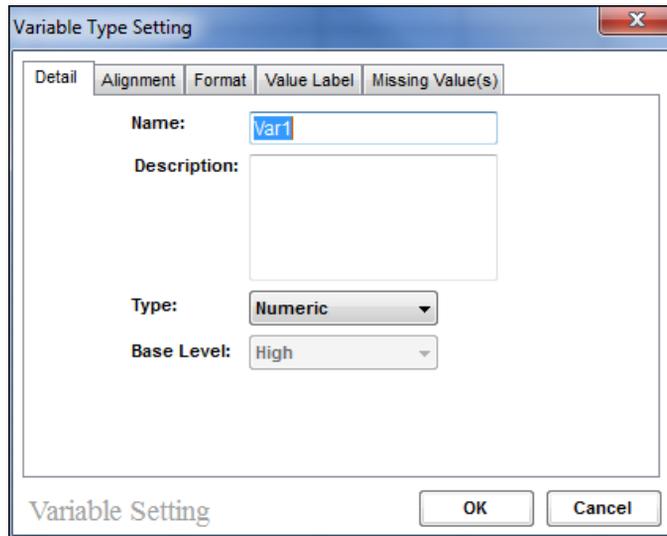
97 Analysis-Multiple Comparison Procedures for Survival Data

Analysis Steps:

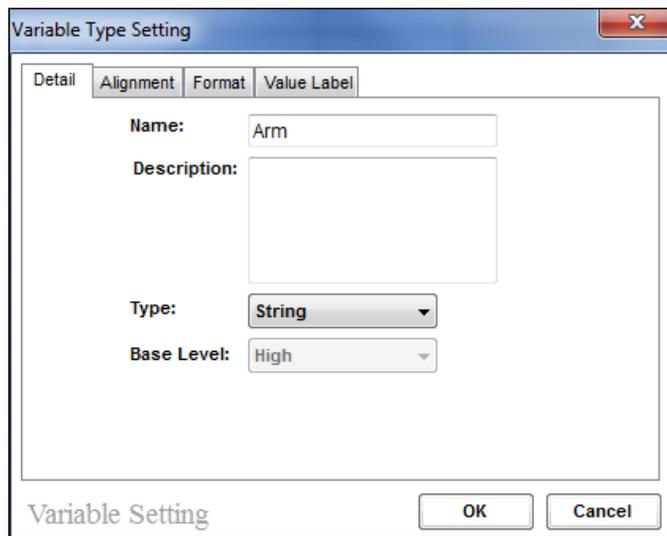
1. In the **Home** tab, choose textbf New > Case Data.
This will open an empty dataset in the main window. Now right click on the column header and click **Create Variable** as shown below.



2. This will bring up the following **Variable Type Setting** dialog box.

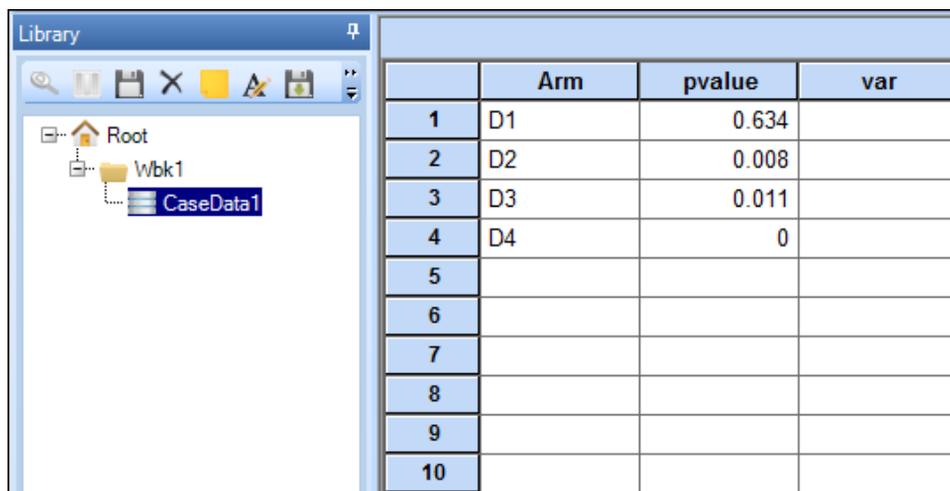


3. Type in **Arm** for **Name** and choose the type of variable as **String**.



97 Analysis-Multiple Comparison Procedures for Survival Data

4. Click **OK** and this will add a column with name **Arm** in the dataset. Similarly, create a numeric column with label **pvalue**. Now, enter the values in the table as follows:



| | Arm | pvalue | var |
|----|-----|--------|-----|
| 1 | D1 | 0.634 | |
| 2 | D2 | 0.008 | |
| 3 | D3 | 0.011 | |
| 4 | D4 | 0 | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |

5. East assigns a default name **CaseData1** to this dataset.
6. Choose the menu item:
Analysis > (Continuous) Many Samples > (Multiple Comparisons) Pairwise Comparisons to Controls - Difference of Means
7. This will display several input fields associated with multiple comparison test in the main window. In the **Main tab**, select the radio-button corresponding to **raw p-values**. In the ensuing two boxes, select **Arm** as **Treatment variable** and select **pvalue** for **Select raw p-values**. Choose **Bonferroni** from the drop-down list in **Select MCP**.

Analysis: Continuous Endpoint: Many-Sample Test - Multi-Arm Design - Pairwise Comparisons to Control - Differences of Means

Data Set: CaseData1

The screenshot displays the software's configuration window for a multiple comparison test. It features two tabs: 'Main' and 'Advanced'. Under the 'Main' tab, there are two radio buttons: 'raw Data' (unselected) and 'raw p-values' (selected). Below these are two input sections. The first, 'Treatment Variable', has a dropdown menu set to 'Arm' and a list of options: D1, D2, D3, and D4. The second, 'Raw p-value Variable', has a dropdown menu set to 'pvalue'. To the right, the 'Multiple Comparisons Procedures' section contains a dropdown menu labeled 'Select MCP:' which is set to 'Bonferroni'.

97 Analysis-Multiple Comparison Procedures for Survival Data

8. Click **OK**. The output will be displayed in the main window.

Analysis: Multiple Pairwise Comparisons with Control: Difference of Means

Hypothesis
 $H_i : \mu_i - \mu_0 \leq 0$ Vs. $K_i : \mu_i - \mu_0 > 0$ for $i = 1, 2, \dots, k-1$ where k is the total number of arms.

Input Parameters

Data File: CaseData1
 Treatment Variable: Arm
 Raw p-values variable: pvalue
 Multiple Comparison Procedure: Bonferroni
 Confidence Level: 0.95
 Test Type: t-test
 Rejection Region: Right-Tail

Output

Inference for each Treatment Level versus Control:

| Arm | p-value | |
|-----|---------|----------|
| | Raw | Adjusted |
| D1 | 0.634 | 1 |
| D2 | 0.008 | 0.032 |
| D3 | 0.011 | 0.044 |
| D4 | 0 | 0 |

Adjusted Global p-value: 0
 Total no. of records: 4
 No. of records rejected: 0
 Total Number of Arms (k): 4

The adjusted p-values for D1, D2, D3 and D4 are 1, 0.032, 0.044 and 0.000, respectively. Note that these adjusted p-values are very close to what we have obtained with Bonferroni procedure using the dataset [Hypertension-trial.cyd](#). Ideally, both set of p-values should exactly match. The difference in p-values is only due to rounding error.

Volume 10

Appendices

| | |
|---|------|
| <i>A Introduction to Volume 10</i> | 2269 |
| <i>B Group Sequential Design in East 6</i> | 2271 |
| <i>C Interim Monitoring in East 6</i> | 2313 |
| <i>D Computing the Expected Number of Events</i> | 2334 |
| <i>E Generating Survival Simulations in EastSurv</i> | 2345 |
| <i>F Spending Functions Derived from Power Boundaries</i> | 2347 |
| <i>G The Recursive Integration Algorithm</i> | 2352 |
| <i>H Theory - Multiple Comparison Procedures</i> | 2353 |
| <i>I Theory - Multiple Endpoint Procedures</i> | 2368 |
| <i>J Theory-Multi-arm Multi-stage Group Sequential Design</i> | 2374 |
| <i>K Theory - MultiArm Two Stage Designs Combining p-values</i> | 2394 |
| <i>L Technical Details - Predicted Interval Plots</i> | 2404 |

| | |
|--|-------------|
| <i>M Enrollment/Events Prediction - Theory</i> | <i>2409</i> |
| <i>N Dose Escalation - Theory</i> | <i>2412</i> |
| <i>O R Functions</i> | <i>2427</i> |
| <i>P East 5.x to East 6.4 Import Utility</i> | <i>2478</i> |
| <i>Q Technical Reference and Formulas: Single Look Designs</i> | <i>2484</i> |
| <i>R Technical Reference and Formulas: Analysis</i> | <i>2542</i> |
| <i>S Theory - Design - Binomial One-Sample Exact Test</i> | <i>2605</i> |
| <i>T Theory - Design - Binomial Paired-Sample Exact Test</i> | <i>2611</i> |
| <i>U Theory - Design - Simon's Two-Stage Design</i> | <i>2614</i> |
| <i>V Theory-Design - Binomial Two-Sample Exact Tests</i> | <i>2617</i> |
| <i>W Classification Table</i> | <i>2638</i> |
| <i>X Glossary</i> | <i>2639</i> |
| <i>Y On validating the East Software</i> | <i>2657</i> |
| <i>Z List of East Beta Testers</i> | <i>2686</i> |

A *Introduction to Volume 10*

This volume contains all the Appendices for East 6 manual.

Appendix **B** provides the technical details of the design phase.

Appendix **C** deals with the technical explanation of interim monitoring phase.

Appendix **D** deals with the formulas used for the expected number of events in one treatment arm in various situations. The situations we consider vary from simple ones where the hazard rate is constant, the accrual rate is constant, there are no dropouts and each patient is followed until the end of the study, to complex ones where the survival curve is modeled as a piecewise exponential function with K pieces of variable hazard rates, variable accrual rates, constant non-zero dropout rates and where patients are followed for a fixed duration.

Appendix **E** gives the details of the powerful simulation tools available in East for trials with time-to-event endpoints. The simulations may be used to actually design for non-standard problems where power and sample size calculations are analytically intractable. For instance, East allows the user to simulate trials in which the hazard rates for each treatment arm are non-proportional. By trial and error, running simulations under various parameter choices, the user may find an appropriate design for this kind of trial.

Appendix **F** discusses the technical aspects involved in using spending functions boundaries and Wang-Tsiatis or Pampallona-Tsiatis family boundaries in design and monitoring of trials.

Appendix **G** explains the efficiency achieved by employing the Recursive Integration Algorithm in the computations for the various procedures in East.

Appendix **H** lays out the theory behind multiple comparison procedures like Step-up and Step-down Dunnett's test and other p-value based procedures like Bonferroni, Sidak and some more.

Appendix **I** lays out the theory behind multiple endpoint procedures like Serial Gatekeeping and Parallel Gatekeeping.

Appendix **S** lays out the theory behind East's power and sample size computations in the case of the exact fixed sample test and the exact group sequential test of a proportion π being equal to a constant π_0 .

A Introduction to Volume 10

Appendix **T** lays out the theory behind East's power and sample size computations in the case of the exact McNemar's test for the difference of proportions arising from paired binomial populations.

Appendix **U** lays out the theory behind the two-stage optimal design for phase 2 clinical trials developed by Simon (1989).

Appendix **V** lays out the theory behind exact power and sample size computations for comparing two independent binomials.

Appendix **N** lays out the theory behind the dose escalation procedures like 3+3, CRM, mTPI and BLRM introduced in East 6.3

Appendix **M** lays out the theory behind the subject enrollment and event prediction introduced in East 6.3.

Appendix **Q** lays out the theory behind the designs in **East** and formulas used for calculations. For each test we provide its null hypothesis, test statistic, distribution of the test statistic under null hypothesis.

Appendix **R** lays out the theory and formulas used in **East** for analyzing data under the **Analysis** menu.

Appendix **W** lists down the formulas used in computing classification errors.

Appendix **O** discusses the R Integration feature in simulation module which provides the user the opportunity to perform various tasks using R. In this appendix, we list all tasks for which R functions can be used. We will provide syntax and suggested format for various functions.

Appendix **X** provides a glossary of terms and quantities used in East6.

Appendix **Y** describes the extensive validating procedures carried out on all the features incorporated in East 6 and some earlier versions of East.

Appendix **Z** lists down all the beta testers of East who have given their valuable inputs while developing this software.

B Group Sequential Design in East 6

East provides the software support for a repeated significance testing strategy whereby the accumulating data in a phase-III randomized clinical trial are monitored, and the trial is terminated with early rejection of either the null or the alternative hypothesis if a given test statistic crosses a given stopping boundary. This strategy is executed in two phases – the design phase and the interim monitoring phase. Appendix B provides the technical details of the design phase. Appendix C deals with the interim monitoring phase. A thorough coverage of group-sequential methods for clinical trials is offered at an expository level in the textbook by Jennison and Turnbull (2000). This textbook is an excellent complement to the methods discussed in these appendix chapters and implemented in the East software.

At the design phase the user specifies the statistical process generating the data, the null and alternative hypotheses being tested, the desired type-I error, the power of the sequential testing procedure, the shape parameters for the spending functions or stopping boundaries, the planned number of interim looks, and the timing of the interim looks. East uses these input parameters to generate the appropriate stopping boundaries and to compute the maximum statistical information that would be needed to achieve the desired operating characteristics of the sequential testing procedure. Depending on the end point of the clinical trial, the maximum statistical information might be expressed in terms of the patient accrual, the number of events such as failures or deaths, or an abstract dimensionless quantity termed Fisher information.

We lay the ground work for designing group sequential studies in Section B.1 where we define the test statistic to be monitored and specify its distributional properties. This distribution theory is presented first in terms of a general framework which is then applied to studies with normal, binomial, time to failure and general end points. In Section B.2 we derive the stopping boundaries for various group sequential designs. In Section B.3 we introduce the notion of an inflation factor and show how it can be applied in the General and Information Based designs available in East. In Section B.4 we compute the expected sample size and expected study duration for these group sequential designs.

Although the methodology in this appendix has been developed with reference to two-arm clinical trials, it applies with obvious modifications to the one-sample setting as well. For multi-arm trials in which two or more treatment arms are compared to a common control arm, the two-arm approach can still be applied if supplemented by multiple testing procedures such as Bonferroni or Hochberg. More general situations are handled as special cases of the regression problem discussed in Section B.1.4. In effect one unified approach is adopted for all the group sequential procedures in East.

B Group Sequential Design in East 6

However, since the various cases considered utilize different test statistics for interim monitoring we have provided the formula for each test statistic in Appendix Q.

B.1 Distribution Theory

B.1.1 Normal Data

B.1.2 Binomial Data

B.1.3 Time to Event Data

B.1.4 General Regression Models

Consider a two arm randomized clinical trial comparing an experimental treatment with a control treatment. Let the treatment difference of primary interest be denoted by a single scalar parameter δ . The choice of parameter δ will depend on the model generating the patient response. For normal response, δ might represent the difference of means. For binomial response, δ might represent a difference of proportions, a ratio of proportions, an odds ratio, or a log odds ratio. For time-to-event response, δ might represent a difference of medians, a difference of survival rates at a given time-point, a hazard ratio, or a log hazard ratio. More generally, δ might be the coefficient of the treatment effect in a regression model. Suppose we intend to monitor the accumulating data sequentially up to a maximum of K times thereby gathering, in succession, I_1, I_2, \dots, I_K units of statistical information about δ . In a parametric model I is called the Fisher information. In a semiparametric model, it is called the semiparametric information bound. Since I_K represents the maximum information we could obtain, we will also denote it by I_{\max} . It is convenient to define the information fraction $t_j = I_j/I_{\max}$. For trials with normal or binomial response, I_j is proportional to n_j , the total sample size attained by the j th monitoring time-point, and $t_j = n_j/n_{\max}$. For trials with time-to-event response, I_j is approximately proportional to d_j , the total number of events observed by the j th monitoring time-point. In that case $t_j = d_j/d_{\max}$. One may regard the information fraction $t \in [0, 1]$ as the internal time of the clinical trial.

We assume that at each interim monitoring time-point, t_j , we can obtain an efficient estimate, $\hat{\delta}(t_j)$ for δ , a consistent estimate, $\text{var}[\hat{\delta}(t_j)]$ for the variance of $\hat{\delta}(t_j)$, and the sample size (or number of events) is large enough that

$$I_j^{-1} \approx \text{var}[\hat{\delta}(t_j)].$$

Formally an estimate is efficient if it achieves the Cramer-Rao lower bound for parametric models and the information bound as defined by Bickel et. al. (1993) for semiparametric models. In particular maximum likelihood estimates are efficient. Most estimates produced by standard statistical packages like SAS or S-plus for parametric or semiparametric models are efficient. Scharfstein, Tsiatis and Robins (1997) have shown that, under the above conditions, the joint distribution of the Wald statistics

$$Z(t_j) = \frac{\hat{\delta}(t_j) - \delta_0}{\sqrt{\text{var}[\hat{\delta}(t_j)]}} \tag{B.1}$$

for testing

$$H_0: \delta = \delta_0, \tag{B.2}$$

computed sequentially at information fractions t_1, t_2, \dots, t_K , is asymptotically multivariate normal with

$$E[Z(t_j)] = \eta\sqrt{t_j}, \tag{B.3}$$

$$\text{var}[Z(t_j)] = 1, \tag{B.4}$$

and for any $t_{j_1} < t_{j_2}$,

$$\text{covar}[Z(t_{j_1}), Z(t_{j_2})] = \sqrt{\frac{I_{j_1}}{I_{j_2}}}, \tag{B.5}$$

where

$$\eta = (\delta - \delta_0)\sqrt{I_{\max}} \tag{B.6}$$

is known as the drift parameter. Usually, $\delta_0 = 0$ for superiority trials and $\delta_0 > 0$ for non-inferiority trials.

An alternative way to express this result is in terms of a process of independent increments. Define

$$W(t_j) = \sqrt{t_j}Z(t_j). \tag{B.7}$$

Then the joint distribution of $\{W(t_1), W(t_2), \dots, W(t_K)\}$ is asymptotically multivariate normal with

$$E[W(t_j)] = \eta t_j, \tag{B.8}$$

$$\text{var}[W(t_j)] = t_j \tag{B.9}$$

and

$$\text{covar}[W(t_{j_1}), W(t_{j_2})] = t_{j_1}. \tag{B.10}$$

From this it follows that, for any $t_{j_2} > t_{j_1}$, the random variables $W(t_{j_1})$ and $W(t_{j_2}) - W(t_{j_1})$ are independent. A parallel result has been obtained by Jennison and Turnbull (1997). This important result has three implications.

1. Most clinical trials, including trials with normal, binomial and survival endpoints, utilize test statistics of the form (B.1). Therefore, by the above theorem, the distributional structure of these test statistics after applying the transformation (B.7), is asymptotically the same as that of the $W(t_j)$'s. Thus one may construct group sequential stopping boundaries for the $W(t_j)$ stochastic process, having the property that under $H_0: \eta = 0$ the probability of crossing a boundary is limited to α , the desired type-1 error. These same

B Group Sequential Design in East 6

boundaries will then be applicable to the test statistics developed to monitor trials with normal, binomial or survival endpoints, or even more general endpoints like those available through the information based design module of East. Thereby we can construct a common set of boundaries that are applicable to any type of trial provided the test statistics used for monitoring the trial have the same asymptotic distributional structure as the $W(t_j)$ stochastic process. The details of boundary construction are provided in Section B.2.

2. Having generated the appropriate boundaries one may compute boundary crossing probabilities for the stochastic process $W(t_j)$ under alternative hypotheses of the form $H_1: \eta = \eta_1$. One can thereby search for the value of η_1 at which the boundary crossing probability equals the desired power, $1 - \beta$. By substituting this value of η into equation (B.6) one can estimate I_{\max} , the maximum information needed to attain the desired power $1 - \beta$, at any pre-specified clinically meaningful treatment difference $\delta = \delta_1$. The details of these computations are provided in Section B.2.
3. Because of the independent increments structure of the $W(t_j)$'s it is possible to perform the actual computations that lead to these group sequential stopping boundaries and their crossing probabilities very efficiently by the recursive integration techniques of Armitage, McPherson and Rowe (1969).

The distribution theory developed above is applicable to data generated from any arbitrary probability model in which a single scalar parameter δ characterizes the relationship under investigation. In the remainder of Section B.1 we demonstrate that many different statistical models for generating the data provide us with a test statistic whose distributional structure is asymptotically the same as that of the $W(t_j)$ stochastic process. We first consider two-arm randomized clinical trials with normal, binomial and survival endpoints. We then show how the approach may be generalized to any data generating process in which inference is required for a single scalar parameter estimated by an efficient estimator.

B.1.1 Normal Data

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of a normally distributed outcome variable, X , with means μ_t and μ_c , respectively, and with a common variance σ^2 . We intend to monitor the data up to K times after accruing $n_1, n_2, \dots, n_K \equiv n_{\max}$ patients. The information fraction at the j th look is given by $t_j = n_j/n_{\max}$. Let r denote the fraction randomized to treatment T.

Efficacy Trials

Define the treatment difference to be

$$\delta = \mu_t - \mu_c .$$

The null hypothesis of interest is

$$H_0 : \delta = 0 .$$

We wish to construct a K -look group sequential level- α test of H_0 having $1 - \beta$ power at the alternative hypothesis

$$H_1 : \delta = \delta_1 .$$

Let $\bar{X}_t(t_j)$ and $\bar{X}_c(t_j)$ be the mean responses of the experimental and control groups, respectively, at time t_j . Then

$$\hat{\delta}(t_j) = \bar{X}_t(t_j) - \bar{X}_c(t_j) \tag{B.11}$$

and

$$\text{var}[\hat{\delta}(t_j)] = \frac{\sigma^2}{n_j(r)(1-r)} . \tag{B.12}$$

Therefore, by the Scharfstein, Tsiatis and Robins (1997), Jennison and Turnbull (1997) theorem the stochastic process

$$W(t_j) = \sqrt{t_j} \frac{\bar{X}_t(t_j) - \bar{X}_c(t_j)}{\sqrt{\frac{\sigma^2}{n_j(r)(1-r)}}}, \quad j = 1, 2, \dots, K, \tag{B.13}$$

is $N(\eta t_j, t_j)$ with independent increments, where $\eta = 0$ under H_0 and $\eta = \delta_1 \sqrt{I_{\max}}$ under H_1 . We refer to η as the drift parameter.

Non-Inferiority Trials

Define the treatment difference to be

$$\delta = \mu_t - \mu_c .$$

Let δ_0 be the non-inferiority margin. The null hypothesis of interest is

$$H_0 : \delta = \delta_0 .$$

We wish to construct a K -look group sequential level- α test of H_0 having $1 - \beta$ power at the alternative hypothesis

$$H_1 : \delta = \delta_1 .$$

B Group Sequential Design in East 6

Then, by the Scharfstein, Tsiatis and Robins (1997), Jennison and Turnbull (1997) theorem, the stochastic process

$$W(t_j) = \sqrt{t_j} \frac{\bar{X}_t(t_j) - \bar{X}_c(t_j) - \delta_0}{\sqrt{\frac{\sigma^2}{n_j(r)(1-r)}}, \quad j = 1, 2, \dots, K, \quad (\text{B.14})$$

is $N(\eta t_j, t_j)$ with independent increments, where $\eta = 0$ under H_0 and $\eta = (\delta_1 - \delta_0)\sqrt{I_{\max}}$ under H_1 . We refer to η as the drift parameter.

Note that equation (B.12) implies that

$$I_{\max} = \left[\frac{\sigma^2}{n_{\max}(r)(1-r)} \right]^{-1}, \quad (\text{B.15})$$

for both the efficacy and non-inferiority trials. We shall show in Section B.2 how to estimate the value of I_{\max} needed in order to achieve a desired amount of power. Equation (B.15) is required for converting maximum information, an abstract dimensionless quantity, into maximum sample size, a physical resource that one usually has to specify at the planning stages of the clinical trial. The equation shows that in order to make the translation from I_{\max} to n_{\max} one must know the value of σ^2 , a nuisance parameter.

Test Statistics Used for the Interim Monitoring

The test statistics (B.13) and (B.14) both contain σ^2 , a nuisance parameter whose value is typically unknown. Thus we cannot track the path traced by these statistics in the course of a clinical trial, and cannot know for sure if they have crossed a stopping boundary. In practice therefore we replace σ^2 by its estimate $s^2(t_j)$, at each interim monitoring time-point t_j , when monitoring a clinical trial with normal endpoints. The modified statistics also have the same large sample behavior and independent increment structure as the $W(t_j)$'s. Therefore the operating characteristics of hypothesis tests and confidence intervals derived by tracking the modified statistics will resemble those that would have been obtained by tracking the original statistics.

B.1.2 Binomial Data

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of a binary response variable, X , with response probabilities π_t and π_c for the experimental and control arms, respectively. We intend to monitor the data up to K times after accruing $n_1, n_2, \dots, n_K \equiv n_{\max}$ patients. The information fraction at the j th look is given by $t_j = n_j/n_{\max}$. Let r denote the fraction randomized to treatment T.

Efficacy Trials

Define the treatment difference to be

$$\delta = \pi_t - \pi_c$$

The null hypothesis of interest is

$$H_0 : \delta = 0 .$$

We wish to construct a K -look group sequential level- α test of H_0 having $1 - \beta$ power at the alternative hypothesis

$$H_1 : \delta = \delta_1 .$$

Let $\hat{\pi}_t(t_j)$ and $\hat{\pi}_c(t_j)$ be the maximum likelihood estimates of π_t and π_c , respectively, at time t_j . Then

$$\hat{\delta}(t_j) = \hat{\pi}_t(t_j) - \hat{\pi}_c(t_j) \tag{B.16}$$

and

$$\text{var}[\hat{\delta}(t_j)] = \frac{\pi_t(1 - \pi_t)}{rn_j} + \frac{\pi_c(1 - \pi_c)}{(1 - r)n_j} . \tag{B.17}$$

Therefore, by the Scharfstein, Tsiatis and Robins, Jennison and Turnbull (1997) theorem, the stochastic process

$$W_0(t_j) = \sqrt{t_j} \frac{\hat{\pi}_t(t_j) - \hat{\pi}_c(t_j)}{\sqrt{\frac{\pi_c(1 - \pi_c)}{n_j(r)(1 - r)}}} \tag{B.18}$$

is $N(0, t_j)$ with independent increments, under H_0 . Under H_1 , the stochastic process

$$W_1(t_j) = \sqrt{t_j} \frac{\hat{\pi}_t(t_j) - \hat{\pi}_c(t_j)}{\sqrt{\frac{(\pi_c + \delta_1)(1 - \pi_c - \delta_1)}{rn_j} + \frac{\pi_c(1 - \pi_c)}{(1 - r)n_j}}} \tag{B.19}$$

is $N(\eta t_j, t_j)$ with independent increments, where $\eta = \delta_1 \sqrt{I_{\max}}$ is known as the drift parameter.

Note that equation (B.17) and H_1 together imply that

$$I_{\max} = \left[\frac{(\pi_c + \delta_1)(1 - \pi_c - \delta_1)}{rn_{\max}} + \frac{\pi_c(1 - \pi_c)}{(1 - r)n_{\max}} \right]^{-1} . \tag{B.20}$$

We shall show in Section B.2 how to estimate the value of I_{\max} needed in order to achieve a desired amount of power. Equation (B.20) is required for converting maximum information, an abstract dimensionless quantity, into maximum sample size,

B Group Sequential Design in East 6

a physical resource that one usually has to specify at the planning stages of the clinical trial. The equation shows that in order to make the translation from I_{\max} to n_{\max} one must know the value of π_c , a nuisance parameter.

Non-Inferiority Trials

Define the treatment difference to be

$$\delta = \pi_t - \pi_c .$$

Let the non-inferiority margin be δ_0 . The null hypothesis of interest is

$$H_0 : \delta = \delta_0 .$$

We wish to construct a K -look group sequential level- α test of H_0 having $1 - \beta$ power at the alternative hypothesis

$$H_1 : \delta = \delta_1 .$$

Then, by the Scharfstein, Tsiatis and Robins (1997), Jennison and Turnbull (1997) theorem, the stochastic process

$$W_0(t_j) = \sqrt{t_j} \frac{\hat{\pi}_t(t_j) - \hat{\pi}_c(t_j) - \delta_0}{\sqrt{\frac{(\pi_c - \delta_0)(1 - \pi_c + \delta_0)}{rn_j} + \frac{\pi_c(1 - \pi_c)}{(1 - r)n_j}}} \tag{B.21}$$

is $N(0, t_j)$ with independent increments under H_0 . Under H_1 , the stochastic process

$$W_1(t_j) = \sqrt{t_j} \frac{\hat{\pi}_t(t_j) - \hat{\pi}_c(t_j) - \delta_0}{\sqrt{\frac{(\pi_c - \delta_1)(1 - \pi_c + \delta_1)}{rn_j} + \frac{\pi_c(1 - \pi_c)}{(1 - r)n_j}}} \tag{B.22}$$

is $N(\eta t_j, t_j)$ with independent increments, where $\eta = (\delta_1 - \delta_0)\sqrt{I_{\max}}$ is known as the drift parameter.

Note that equation (B.17) and H_1 together imply that

$$I_{\max} = \left[\frac{(\pi_c - \delta_1)(1 - \pi_c + \delta_1)}{rn_{\max}} + \frac{\pi_c(1 - \pi_c)}{(1 - r)n_{\max}} \right]^{-1} . \tag{B.23}$$

We shall show in Section B.2 how to estimate the value of I_{\max} needed in order to achieve a desired amount of power. Equation (B.23) is required for converting maximum information, an abstract dimensionless quantity, into maximum sample size, a physical resource that one usually has to specify at the planning stages of the clinical trial. The equation shows that in order to make the translation from I_{\max} to n_{\max} one must know the value of π_c , a nuisance parameter.

Test Statistics Used for the Interim Monitoring

The test statistics (B.18), (B.19), (B.21) and (B.22) all contain π_c , an unknown nuisance parameter. Therefore, in practice, modified test statistics, whose values can be computed from the interim data, are used to track the progress of the trial and determine if a stopping boundary has been crossed.

1. For superiority trials East provides two options in the choice of test statistic to be used during interim monitoring. The default assumption is that the test statistic

$$\tilde{W}_s(t_j) = \sqrt{t_j} \frac{\hat{\pi}_t(t_j) - \hat{\pi}_c(t_j)}{\sqrt{\frac{\hat{\pi}_t(t_j)(1-\hat{\pi}_t(t_j))}{n_{tj}} + \frac{\hat{\pi}_c(t_j)(1-\hat{\pi}_c(t_j))}{n_{cj}}}}, \quad (\text{B.24})$$

will be used for the interim monitoring, where n_{tj} and n_{cj} are the sample sizes on the treatment and control arms, respectively, at monitoring time-point t_j .

Asymptotically, $\tilde{W}_s(t_j)$ behaves like (B.18) under H_0 and behaves like (B.19) under H_1 . Thus in either case $\tilde{W}_s(t_j)$ has the same asymptotic behavior as the $N(\eta t_j, t_j)$ stochastic process with independent increments. Therefore the operating characteristics of hypothesis tests and confidence intervals derived by tracking $\tilde{W}_s(t_j)$ will resemble those that would have been obtained by tracking the (B.18) under H_0 and tracking (B.19) under H_1 .

An alternative choice for the test statistic to be used during the interim monitoring phase is

$$\tilde{W}_{0s}(t_j) = \sqrt{t_j} \frac{\hat{\pi}_t(t_j) - \hat{\pi}_c(t_j)}{\sqrt{\frac{\hat{\pi}(t_j)(1-\hat{\pi}(t_j))}{n_j(r)(1-r)}}}, \quad (\text{B.25})$$

where $\hat{\pi}(t_j)$, the **pooled** estimate of the binomial response probability at time t_j , is given by

$$\hat{\pi}(t_j) = \frac{n_{tj}\hat{\pi}_t(t_j) + n_{cj}\hat{\pi}_c(t_j)}{n_j}. \quad (\text{B.26})$$

The denominator of (B.25) is an estimate of the standard error of $\hat{\pi}_t(t_j) - \hat{\pi}_c(t_j)$ under the null hypothesis H_0 : $\delta = 0$. Therefore $\tilde{W}_{0s}(t_j)$ behaves asymptotically like (B.18) under H_0 . However, unlike $\tilde{W}_s(t_j)$, it does not behave like (B.19) under H_1 . For this reason, as we shall show in Section B.2.5, the maximum information I_{\max} , required to attain any given power $1 - \beta$, differs by whether the unpooled statistic $\tilde{W}_s(t_j)$ or the pooled statistic $\tilde{W}_{0s}(t_j)$ is used for interim monitoring.

2. For non-inferiority trials we use the test statistic

$$\tilde{W}_{ni}(t_j) = \sqrt{t_j} \frac{\hat{\pi}_t(t_j) - \hat{\pi}_c(t_j) - \delta_0}{\sqrt{\frac{\hat{\pi}_t(t_j)(1-\hat{\pi}_t(t_j))}{n_{tj}} + \frac{\hat{\pi}_c(t_j)(1-\hat{\pi}_c(t_j))}{n_{cj}}}}, \quad (\text{B.27})$$

B Group Sequential Design in East 6

where n_{tj} and n_{cj} are the sample sizes on the treatment and control arms, respectively, at monitoring time-point t_j . Asymptotically, $\tilde{W}_{ni}(t_j)$ behaves like (B.21) under H_0 and behaves like (B.22) under H_1 . Thus in either case $\tilde{W}_{ni}(t_j)$ has the same asymptotic behavior as the $N(\eta t_j, t_j)$ stochastic process with independent increments. Therefore the operating characteristics of hypothesis tests and confidence intervals derived by tracking $\tilde{W}_{ni}(t_j)$ will resemble those that would have been obtained by tracking the (B.21) under H_0 and tracking (B.22) under H_1 .

B.1.3 Time to Event Data

Consider a randomized clinical trial comparing two treatments, T and C, on the basis of time to event data. Let the fraction of patients randomized to treatment T be r . We intend to monitor the data up to K times at calendar times l_1, l_2, \dots, l_K . At calendar time l_j let there be q_j distinct failures, with corresponding failure times denoted by $\tau_1(l_j), \tau_2(l_j), \dots, \tau_{q_j}(l_j)$ (on the patient follow-up time-scale, not the calendar time-scale). At the i th of these q_j failure times let $d_t(\tau_i(l_j))$ be the number of failures on treatment T, $n_t(\tau_i(l_j))$ be the number of subjects on treatment T at risk of failure, $d_c(\tau_i(l_j))$ be the number of failures on treatment C, and $n_c(\tau_i(l_j))$ be the number of subjects on treatment C at risk of failure. The data at calendar time l_j may thus be represented as $q_j \times 2$ contingency tables, where the i th table is of the form

| Status | Treatment T | Treatment C | Total |
|------------|---------------------------------------|---------------------------------------|-----------------------------------|
| Failed | $d_t(\tau_i(l_j))$ | $d_c(\tau_i(l_j))$ | $d(\tau_i(l_j))$ |
| Not Failed | $n_t(\tau_i(l_j)) - d_t(\tau_i(l_j))$ | $n_c(\tau_i(l_j)) - d_c(\tau_i(l_j))$ | $n(\tau_i(l_j)) - d(\tau_i(l_j))$ |
| Total | $n_t(\tau_i(l_j))$ | $n_c(\tau_i(l_j))$ | $n(\tau_i(l_j))$ |

Efficacy Trials

The logrank score statistic $S(l_j)$, at calendar time l_j , is obtained by summing the observed minus the expected values in cell (1, 1) of the above collection of $q_j \times 2$ tables:

$$S(l_j) = - \sum_{i=1}^{q_j} \left\{ d_t(\tau_i(l_j)) - \frac{n_t(\tau_i(l_j)) \times d(\tau_i(l_j))}{n(\tau_i(l_j))} \right\}. \quad (\text{B.28})$$

If treatments T and C have the same underlying distribution, it is well known (see for example, Mantel, 1966) that the marginal distribution of $S(l_j)$ is asymptotically normal with a mean of zero and with variance equal to the sum of hypergeometric

variances across all the tables:

$$\text{var}[S(l_j)] = \sum_{i=1}^{q_j} \frac{n_t(\tau_i(l_j)) \times n_c(\tau_i(l_j)) \times d(\tau_i(l_j)) \times [n(\tau_i(l_j)) - d(\tau_i(l_j))]}{[n(\tau_i(l_j))]^2 [n(\tau_i(l_j)) - 1]} . \tag{B.29}$$

The above variance cannot be used for designing a time-to-event trial, however, since it depends on quantities that cannot be estimated a priori. However, the asymptotic distribution of $S(l_j)$ under proportional hazard alternatives was reduced to a simpler form suitable for designing time-to-event trials by Schoenfeld (1981). Specifically, let $\lambda_t(\tau)$ and $\lambda_c(\tau)$ be the hazard functions for treatment T and treatment C, respectively. Assume that the ratio of hazard functions is constant for all values of τ and define the treatment difference as

$$\delta = \ln \frac{\lambda_t(\tau)}{\lambda_c(\tau)} .$$

Let the total number of failures observed by calendar time l_j be

$$D(l_j) = \sum_{i=1}^{q_j} d(\tau_i(l_j)) ,$$

and let r denote the proportion randomized to treatment T. Then, for $j = 1, 2, \dots, K$, $S(l_j)$ is asymptotically normal with

$$E[S(l_j)] = \delta D(l_j) r (1 - r) , \tag{B.30}$$

$$\text{var}[S(l_j)] = D(l_j) r (1 - r) . \tag{B.31}$$

Tsiatis (1981) proved that the sequentially computed logrank score statistics $S(l_1), S(l_2), \dots, S(l_K)$ have independent increments. That is, and for any $j_2 > j_1$, $S(l_{j_1})$ and $S(l_{j_2}) - S(l_{j_1})$ are independent. The independent increments property and the asymptotic normality of $S(l_j)$ makes it possible to design group sequential trials by the same methods as are used to design group sequential trials with normal endpoints, as we now show.

We wish to test the null hypothesis

$$H_0: \delta = 0$$

versus the alternative hypothesis

$$H_1: \delta = \delta_1 .$$

B Group Sequential Design in East 6

In performing this hypothesis test it is useful to transform the stochastic process $S(l_j)$, $j = 1, 2, \dots, K$, from a process defined on the calendar time l_j , to a process defined on the information fraction

$$t_j = \frac{D(l_j)}{D(l_K)} \equiv \frac{D(l_j)}{D_{\max}} .$$

Thus we define

$$W(t_j) = \frac{S(l_j)}{\sqrt{r(1-r)D_{\max}}} . \tag{B.32}$$

Now, since the variance of $S(l_j)$ is also the Fisher information for δ at the monitoring time l_j (Jennison and Turnbull, 2000, page 78), it follows that the Fisher information at the final monitoring time l_K is given by

$$I_{\max} = \text{var}[S(l_K)] = r(1-r)D_{\max} . \tag{B.33}$$

Therefore $W(t_j) \sim N(\eta t_j, t_j)$ with independent increments, where $\eta = 0$ under H_0 and $\eta = \delta_1 \sqrt{I_{\max}}$ under H_1 . We refer to η as the drift parameter. We shall show in Section B.2 how to estimate the value of I_{\max} needed in order to achieve a desired amount of power. Equation (B.33) establishes the relationship between maximum information, an abstract dimensionless quantity, and the maximum number of events, a physical resource that one usually has to specify at the planning stages of the clinical trial. Notice that D_{\max} plays the same role in a time-to-event trial that N_{\max} plays in a normal endpoint trial.

As an alternative to computing $W(t_j)$ by equation (B.32) one may compute

$$Z(t_j) = \frac{\hat{\delta}(t_j)}{\sqrt{\text{var}(\hat{\delta}(t_j))}} \tag{B.34}$$

where $\hat{\delta}(t_j)$ and its standard error are obtained by fitting a Cox proportional hazards model to the data. Then $\sqrt{t_j}Z(t_j)$ has the same asymptotic distribution as $W(t_j)$.

Non-Inferiority Trials

For non-inferiority trials we again define the treatment difference as

$$\delta = \ln \frac{\lambda_t(\tau)}{\lambda_c(\tau)} . \tag{B.35}$$

Now, however, we are interested in testing the null hypothesis

$$H_0: \delta = \delta_0 ,$$

against the alternative hypothesis

$$H_1: \delta = \delta_1 .$$

where δ_0 is the non-inferiority margin. Accordingly we derive the logrank statistic $S(l_j)$ from the score equations evaluated at $\delta = \delta_0$ (see, for example, Collett, 1994, page 105) so that

$$S(l_j) = \sum_{i=1}^{q_j} \left\{ d_t(\tau_i(l_j)) - \frac{d(\tau_i(l_j)) \times n_t(\tau_i(l_j)) \exp(-\delta)}{n_t(\tau_i(l_j)) \exp(-\delta) + n_c(\tau_i(l_j))} \right\} \quad (\text{B.36})$$

and the variance of this statistic is

$$\text{var}[S(l_j)] = \sum_{i=1}^{q_j} \left\{ \frac{d(\tau_i(l_j)) \times n_t(\tau_i(l_j)) \exp(-\delta) \times n_c(\tau_i(l_j))}{(n_t(\tau_i(l_j)) \exp(-\delta) + n_c(\tau_i(l_j)))^2} \right\} . \quad (\text{B.37})$$

By extending Schoenfeld's (1981) results to this setting we can show that $S(l_j)$ is asymptotically normal with

$$E(S(l_j)) = (\delta - \delta_0)D(l_j)r(1 - r) , \quad (\text{B.38})$$

$$\text{var}[S(l_j)] = D(l_j)r(1 - r) . \quad (\text{B.39})$$

Also it can be shown by application of Martingale results derived from counting processes (L.J.Wei, 2005; personal communication) that the sequentially computed non-central logrank score statistics $S(l_1), S(l_2), \dots, S(l_K)$ have independent increments. Define

$$W(t_j) = \frac{S(l_j)}{\sqrt{\text{var}[S(l_K)]}} . \quad (\text{B.40})$$

Then, asymptotically, $W(t_j) \sim N(\eta t_j, t_j)$ with independent increments, where $\eta = 0$ under H_0 , $\eta = (\delta_1 - \delta_0)\sqrt{I_{\max}}$ under H_1 , and

$$I_{\max} = \text{var}[S(l_K)] = r(1 - r)D_{\max} . \quad (\text{B.41})$$

We refer to η as the drift parameter. We shall show in Section B.2 how to estimate the value of I_{\max} needed in order to achieve a desired amount of power. Equation (B.41) is required for converting maximum information, an abstract dimensionless quantity, into the maximum number of events, a physical resource that one usually has to specify at the planning stages of the clinical trial.

B Group Sequential Design in East 6

As an alternative to computing $W(t_j)$ by equation (B.40) one may compute

$$Z(t_j) = \frac{\hat{\delta}(t_j) - \delta_0}{\sqrt{\text{var}(\hat{\delta}(t_j))}} \tag{B.42}$$

where $\hat{\delta}(t_j)$ and its standard error are obtained by fitting a Cox proportional hazards model to the data. Then $\sqrt{t_j}Z(t_j)$ has the same asymptotic distribution as $W(t_j)$.

B.1.4 General Regression Models

Consider any general regression model including, for example, the linear regression model, the Cox proportional hazards model, the logistic regression model, and the random effects model for longitudinal data. Let δ be the single scalar coefficient of this model that characterizes the treatment effect of interest. (The case where δ is a vector is not considered in this development.) Let $\tau_1, \tau_2, \dots, \tau_K$ denote K monitoring time-points of calendar time. Let $\hat{\delta}(\tau_j)$ be an efficient estimator of δ , $\text{se}(\hat{\delta}(\tau_j))$ be its standard error and

$$Z(\tau_j) = \frac{\hat{\delta}(\tau_j) - \delta_0}{\text{se}(\hat{\delta}(\tau_j))}$$

be the Wald test statistic, based on all the data available at time τ_j . Let $I(\tau_j)$ be the statistical (or Fisher) information about δ available at time τ_j . The quantity $I(\tau_j)$ is estimated by $[\text{se}(\hat{\delta}(\tau_j))]^{-2}$. At any time τ_j we define the information fraction

$$t_j = \frac{I(\tau_j)}{I(\tau_K)} \approx \frac{[\text{se}(\hat{\delta}(\tau_j))]^{-2}}{[\text{se}(\hat{\delta}(\tau_K))]^{-2}}$$

and compute the test statistic

$$W(t_j) = \sqrt{t_j}Z(\tau_j) .$$

Then, using results derived by Scharfstein, Tsiatis and Robins (1997), Jennison and Turnbull (1997), we can show that

$$W(t_j) \sim N(\eta t_j, t_j) , \tag{B.43}$$

where

$$\eta = (\delta - \delta_0)\sqrt{I(\tau_K)} , \tag{B.44}$$

and for any $t_{j'} > t_j$,

$$\text{covar}\{W(t_j), W(t_{j'})\} = t_j . \tag{B.45}$$

This general result encompasses all situations in which group sequential inference is desired for a single scalar parameter δ and where an efficient estimator for δ exists. East provides the option to design and monitor studies within this general framework through its information based approach discussed in Chapter 59.

B.2 Stopping Boundaries and Maximum Information

B.2.1 Haybittle-Peto Boundaries

B.2.2 W-T Power Boundaries

B.2.3 P-T Power Boundaries

B.2.4 Spending Function Boundary

B.2.5 Special Considerations

Suppose we plan to monitor the study a maximum of K times at information fractions t_1, t_2, \dots, t_K . Let the desired type-1 error be α . In this section we show how to compute the following quantities:

1. One- and two-sided boundaries for early stopping to reject the null hypothesis $H_0: \delta = \delta_0$;
2. One- and two-sided boundaries for early stopping to reject either the null hypothesis $H_0: \delta = \delta_0$ or the alternative hypothesis $H_1: \delta = \delta_1$;
3. One- and two-sided boundaries for early stopping to reject only the alternative hypothesis $H_1: \delta = \delta_1$ (also known as futility boundaries);
4. I_{\max} , the maximum information needed to achieve a power of $1 - \beta$ at the alternative hypothesis $H_1: \delta = \delta_1$.

All computations will be performed for the process of independent increments $W(t_j) \sim N(\eta t_j, t_j)$. We have seen in Section B.1 that a very large class of group sequential tests, including all those available in East, are represented by this stochastic process. Hypotheses about δ , the primary parameter of interest, can be converted into corresponding hypotheses about η by the relationship (B.6). Once stopping boundaries have been obtained for the $W(t_j)$ statistic they can readily be transformed into corresponding stopping boundaries for the Wald statistic $Z(t_j)$ because of the relationship $W(t_j) = \sqrt{t_j} Z(t_j)$ implied by equation (B.1). Boundaries in East are represented primarily in terms of the Wald statistic.

Three classes of stopping boundaries are available in East: p-value boundaries – also known as Haybittle-Peto boundaries; power boundaries – also known as Wang-Tsiatis or Pampallona-Tsiatis boundaries; spending function boundaries. Each class is discussed separately below.

B.2.1 P-Value or Haybittle-Peto Boundaries

P-value or Haybittle-Peto boundaries are available for early rejection of the null hypothesis. As first proposed by Haybittle (1971), these boundaries are derived by pre-specifying a small p-value, p_1 say, as the stopping criterion for the first $K - 1$ interim analyses and then computing a final p-value, p_2 say, for declaring statistical significance at the last look in such a way that the overall type-1 error is α . Let z_p denote the upper p th quantile of the standard normal distribution; i.e., $1 - \Phi(z_p) = p$. The trial stops at the first interim look that the p-value is less than or equal to p_1 . If this event does not occur, the trial proceeds to the K th look and statistical significance is declared if the final p-value is less than or equal to p_2 . For one-sided tests, the value of

B Group Sequential Design in East 6

p_2 needed to preserve the type-1 error is obtained by solving the equation

$$1 - P_0(W(t_1) < \sqrt{t_1}z_{p_1}, \dots, W(t_{K-1}) < \sqrt{t_{K-1}}z_{p_1}, W(t_K) < z_{p_2}) = \alpha, \quad (\text{B.46})$$

where $P_0(\cdot)$ denotes probability under the assumption that $\eta = 0$. The solution is obtained by numerical search using the recursive integration method of Armitage, McPherson and Rowe (1969) (the AMR algorithm) discussed in Appendix G. Once the value of p_2 has been determined, the maximum information is obtained by invoking the AMR algorithm repeatedly and searching for the value of η at which

$$P_\eta(W(t_1) < \sqrt{t_1}z_{p_1}, \dots, W(t_{K-1}) < \sqrt{t_{K-1}}z_{p_1}, W(t_K) < z_{p_2}) = \beta. \quad (\text{B.47})$$

Denote the solution by η_1 . Then, by (B.6) the desired maximum information I_{\max} is

$$I_{\max} = \left[\frac{\eta_1}{\delta_1 - \delta_0} \right]^2$$

We can convert maximum information into maximum sample size or maximum events, depending on the model being used, by selecting the appropriate translation equation from (B.15) for normal endpoints, (B.20) or (B.23) for binomial endpoints, and (B.33) or (B.41) for time to event endpoints.

To obtain Haybittle-Peto stopping boundaries for two-sided tests, replace $W(t_j)$ by $|W(t_j)|$ throughout equations (B.46) and (B.47).

In East 6, we have generalized the Haybittle-Peto stopping boundaries to accommodate unequal p-values at each look. Consider a K -look design where we pre-specify small p-values p_1, \dots, p_{K-1} as stopping criteria for each of the first $K - 1$ interim looks at the data. We would now like to compute a final p-value p_K for declaring statistical significance in such a way as to preserve the overall type-1 error α . This is achieved by solving the equation

$$1 - P_0(W(t_1) < \sqrt{t_1}z_{p_1}, \dots, W(t_{K-1}) < \sqrt{t_{K-1}}z_{p_{K-1}}, W(t_K) < z_{p_K}) = \alpha, \quad (\text{B.48})$$

Where $P_0(\cdot)$ denotes probability under the assumption that $\eta = 0$. The solution is obtained by numerical search using the recursive integration method of Armitage,

McPherson and Rowe (1969) (the AMR algorithm) discussed in Appendix G. Once the value of p_K has been determined, the maximum information is obtained by invoking the AMR algorithm repeatedly and searching for the value of η at which

$$P_\eta(W(t_1) < \sqrt{t_1}z_{p_1}, \dots, W(t_{K-1}) < \sqrt{t_{K-1}}z_{p_{K-1}}, W(t_K) < z_{p_K}) = \beta. \quad (\text{B.49})$$

Denote the solution by η_1 . Then, by (B.6) the desired maximum information I_{\max} is

$$I_{\max} = \left[\frac{\eta_1}{\delta_1 - \delta_0} \right]^2$$

We can convert maximum information into maximum sample size or maximum events, depending on the model being used, by selecting the appropriate translation equation from (B.15) for normal endpoints, (B.20) or (B.23) for binomial endpoints, and (B.33) or (B.41) for time to event endpoints.

B.2.2 Wang-Tsiatis Power Boundaries

The power boundaries of Wang and Tsiatis (1987) are available for early rejection of the null hypothesis. These boundaries are of the form

$$c(t_j) = C(\Delta, \alpha, K)t_j^\Delta \quad (\text{B.50})$$

for $j = 1, 2, \dots, K$, where Δ is a shape parameter that characterizes the boundary shape and $C(\Delta, \alpha, K)$ is a positive constant indexed by Δ, α and K . The choice $\Delta = 0$ yields the O'Brien-Fleming (1979) stopping boundaries while $\Delta = 0.5$ yields the Pocock stopping boundaries. More generally Wang and Tsiatis (1987) explore this family to find the value of Δ that minimizes the expected sample size for various design specifications. For one-sided tests, the trial stops at the first interim look that $W(t_j) \geq c(t_j)$. Therefore, in order to preserve the type-1 error the boundaries must satisfy

$$1 - P_0\left\{ \bigcap_{j=1}^K W(t_j) < c(t_j) \right\} = \alpha. \quad (\text{B.51})$$

Since, by equation (B.50), the boundary values $c(t_1), c(t_2), \dots, c(t_K)$ are completely specified by $C(\Delta, \alpha, K)$, this constant can be evaluated by numerical search for any

B Group Sequential Design in East 6

choice of Δ , α and K using the AMR algorithm. Once the boundaries have been determined the maximum information is obtained by again invoking the AMR algorithm, this time to find the value of η that satisfies the type-2 error equation

$$P_{\eta}\left\{\bigcap_{j=1}^K W(t_j) < c(t_j)\right\} = \beta. \tag{B.52}$$

Denote the solution by η_1 . Then, by (B.6) the desired maximum information I_{\max} is

$$I_{\max} = \left[\frac{\eta_1}{\delta_1 - \delta_0}\right]^2.$$

We can convert maximum information into maximum sample size or maximum events, depending on the model being used, by selecting the appropriate translation equation from (B.15) for normal endpoints, (B.20) or (B.23) for binomial endpoints, and (B.33) or (B.41) for time to event endpoints.

To obtain Wang-Tsiatis stopping boundaries for two-sided tests, replace $W(t_j)$ by $|W(t_j)|$ throughout equations (B.51) and (B.52).

B.2.3 Pampallona-Tsiatis Power Boundaries

The power boundaries of Pampallona and Tsiatis (1994) are available for early rejection of either H_0 or H_1 . It is convenient to discuss the one-sided and two-sided tests separately for these boundaries.

One-Sided Tests

There are two stopping boundaries for these designs; an “upper” stopping boundary for early rejection of H_0 and a “lower” stopping boundary for early rejection of H_1 . We reject H_0 in favor of H_1 the first time we encounter an information fraction t_j such that

$$W(t_j) \geq C_1(\Delta_1, \alpha, \beta, K)t_j^{\Delta_1} \text{ (upper boundary) ,}$$

and reject H_1 in favor of H_0 the first time we encounter an information fraction t_j such that

$$W(t_j) < \eta t_j - C_2(\Delta_2, \alpha, \beta, K)t_j^{\Delta_2} \text{ (lower boundary) ,}$$

where $C_1(\Delta_1, \alpha, \beta, K)$ and $C_2(\Delta_2, \alpha, \beta, K)$ are positive and indexed by shape parameters, Δ_1 and Δ_2 , that might take different values. We impose the additional constraint

$$C_1(\Delta_1, \alpha, \beta, K) = \eta - C_2(\Delta_2, \alpha, \beta, K)$$

so as to force the boundaries to meet at the last look, thereby ensuring that a decision to reject either of the two hypotheses will indeed be made. The upper and lower stopping boundaries thus form a triangular continuation region.

We wish these stopping boundaries to have the property that at the null hypothesis, $\delta = 0$, we will cross the upper stopping boundary with probability α , but at the specific alternative hypothesis of interest, say $\delta = \delta_1$, we will cross the upper stopping boundary with probability $1 - \beta$ and the lower stopping boundary with probability β . The coefficients $C_1(\Delta_1, \alpha, \beta, K)$ and $C_2(\Delta_2, \alpha, \beta, K)$ are found using a two-dimensional search to simultaneously solve the two equations corresponding to the desired type-1 and type-2 errors of the test:

$$P_0(W(t_1) \geq u_1) + P_0(l_1 < W(t_1) < u_1, W(t_2) \geq u_2) + \dots \\ \dots + P_0(l_1 < W(t_1) < u_1, \dots, l_{K-1} < W(t_{K-1}) < u_{K-1}, W(t_K) \geq u_K) = \alpha$$

and

$$P_\eta(W(t_1) \leq l_1) + P_\eta(l_1 < W(t_1) < u_1, W(t_2) \leq l_2) + \dots \\ \dots + P_\eta(l_1 < W(t_1) < u_1, \dots, l_{K-1} < W(t_{K-1}) < u_{K-1}, W(t_K) \leq l_K) = \beta$$

where

$$u_j = C_1(\Delta_1, \alpha, \beta, K)t_j^{\Delta_1}$$

and

$$l_j = \eta t_j - C_2(\Delta_2, \alpha, \beta, K)t_j^{\Delta_2}$$

for $j = 1, 2, \dots, K$. The parameter η is determined simultaneously along with $C_1(\Delta_1, \alpha, \beta, K)$ and $C_2(\Delta_2, \alpha, \beta, K)$ through the relationship

$$\eta = C_1(\Delta_1, \alpha, \beta, K) + C_2(\Delta_2, \alpha, \beta, K) . \tag{B.53}$$

Denote the solution by η_1 . Then, by (B.6) the desired maximum information I_{\max} is

$$I_{\max} = \left[\frac{\eta_1}{\delta_1 - \delta_0} \right]^2$$

We can convert maximum information into maximum sample size or maximum events, depending on the model being used, by selecting the appropriate translation equation from (B.15) for normal endpoints, (B.20) or (B.23) for binomial endpoints, and (B.33) or (B.41) for time to event endpoints.

Two-Sided Tests

The two-sided test is based on a pair of outer boundaries for early rejection of H_0 plus an inner wedge for early rejection of H_1 . These tests reject H_0 in favor of H_1^+ : $\delta > 0$ if

$$W(t_j) \geq C_1(\Delta_1, \alpha, \beta, K)t_j^{\Delta_1} \quad (\text{top outer boundary}) ,$$

reject H_0 in favor of H_1^- : $\delta < 0$ if

$$W(t_j) \leq -C_1(\Delta_1, \alpha, \beta, K)t_j^{\Delta_1} \quad (\text{bottom outer boundary}) ,$$

B Group Sequential Design in East 6

and reject $H_1: \delta \neq 0$ if

$$C_2(\Delta_2, \alpha, \beta, K)t_j^{\Delta_2} - \eta t_j \leq W(t_j) \leq \eta t_j - C_2(\Delta_2, \alpha, \beta, K)t_j^{\Delta_2} \quad \text{.(inner wedge)}$$

The boundaries for these tests jointly form two symmetric triangular continuation regions with outer regions for stopping to reject H_0 and an inner wedge for stopping to reject H_1 . The boundaries are required to have the property that, under $H_0: \delta = 0$, the overall probability of crossing either of the two outer boundaries is α , while for the specific alternative of interest, $\delta = \delta_1$ say, the overall probability of crossing either outer boundaries is $1 - \beta$ and the probability of entering the inner wedge is β . Again we will impose the constraint

$$C_1(\Delta_1, \alpha, \beta, K) = \eta - C_2(\Delta_2, \alpha, \beta, K)$$

so that in the end a decision to reject one of the two hypotheses is reached. Notice that the inner wedge is undefined at information fractions t_j such that

$$C_2(\Delta_2, \alpha, \beta, K)t_j^{\Delta_2} - \eta t_j > \eta t_j - C_2(\Delta_2, \alpha, \beta, K)t_j^{\Delta_2} .$$

Therefore it will not be possible to stop the trial with rejection of H_1 at the j th information fraction unless the trial has progressed sufficiently far so that

$$t_j \geq \left[\frac{C_2(\Delta_2, \alpha, \beta, K)}{\eta} \right]^{\frac{1}{1-\Delta_2}} . \tag{B.54}$$

With this in mind we will find it convenient to set $t_j = -\infty$ whenever t_j fails to satisfy the condition (B.54).

Computing Maximum Information The above computations show that the Wang-Tsiatis and Pampallona-Tsiatis boundaries are generated simultaneously with the drift parameter η needed to achieve $1 - \beta$ power. Denote the solution by η_1 . Then, by (B.6) the desired maximum information I_{\max} is

$$I_{\max} = \left[\frac{\eta_1}{\delta_1 - \delta_0} \right]^2$$

We can convert maximum information into maximum sample size or maximum events, depending on the model being used, by selecting the appropriate translation equation from (B.15) for normal endpoints, (B.20) or (B.23) for binomial endpoints, and (B.33) or (B.41) for time to event endpoints.

B.2.4 Spending Function Boundaries

The most general way to generate stopping boundaries is through α and β spending

functions where α and β are, respectively, the type-1 and type-2 errors pre-specified for the trial. An α spending function is any monotone function defined on the unit interval with $\alpha(0) = 0$ and $\alpha(1) = \alpha$. Similarly a β spending function is any monotone function defined on the unit interval with $\beta(0) = 0$ and $\beta(1) = \beta$. The idea of using an α spending function to derive stopping boundaries for early rejection of H_0 was first introduced by Lan and DeMets (1983). Subsequently Pampallona, Tsiatis and Kim (1995) (2001) developed the notion of a β spending function to derive stopping boundaries for early rejection of H_1 . One may either use the α - and β - spending functions singly, or combine both α - and β -spending in a single trial, with one-sided or two-sided, symmetric or asymmetric boundaries.

Below we list and briefly describe all the α - and β -spending functions available in East. We give all the functional forms in terms of $\alpha(t)$ but it is understood that these functional forms also apply to $\beta(t)$.

LD (OF) Lan-DeMets spending function with O'Brien-Fleming flavor. Published by Lan and DeMets (Biometrika, 1983). Functional form:

$$\alpha(t) = \begin{cases} 2 - 2\Phi\left(\frac{z_{\alpha/2}}{\sqrt{t}}\right) & \text{for one-sided tests} \\ 4 - 4\Phi\left(\frac{z_{\alpha/4}}{\sqrt{t}}\right) & \text{for two-sided tests} \end{cases}$$

This function generates stopping boundaries that closely resemble the O'Brien-Fleming (1979) stopping boundaries.

LD (PK) Lan-DeMets spending function with Pocock flavor. Published by Lan and DeMets (Biometrika, 1983). Functional form:

$$\alpha(t) = \alpha \ln\{1 + (e - 1)t\} .$$

This function generates stopping boundaries that closely resemble the Pocock (1977) stopping boundaries.

Gm (γ) Gamma spending function. Published by Hwang, Shih and DeCani (Statistics in Medicine, 1990). Functional Form:

$$\alpha(t) = \begin{cases} \alpha \frac{(1 - e^{-\gamma t})}{(1 - e^{-\gamma})}, & \text{if } \gamma \neq 0 \\ \alpha t & \text{if } \gamma = 0 . \end{cases}$$

Negative values of γ yield convex spending functions that increase in conservatism as γ decreases, while positive values of γ yield concave spending functions that increase in aggressiveness as γ increases. The choice $\gamma = 0$ spends the error linearly. The choice $\gamma = -4$ produces stopping boundaries that resemble the O'Brien-Fleming boundaries. The choice $\gamma = 1$ produces stopping boundaries that resemble the Pocock boundaries.

B Group Sequential Design in East 6

Rho (ρ) Rho spending function. First published by Kim and DeMets (Biometrika, 1987) and generalized by Jennison and Turnbull (2000). Functional form:

$$\alpha(t) = at^\rho, \rho > 0.$$

When $\rho = 1$, the corresponding stopping boundaries resemble the Pocock stopping boundaries. When $\rho = 3$, the boundaries resemble the O'Brien-Fleming boundaries. Larger values of ρ yield increasingly conservative boundaries.

Power Documented in the East 3 User Manual, Appendix B and C (Cytel Software Corporation, 2000). Obtained by inverting 10-look Wang-Tsiatis (1987) stopping boundaries at ten equally spaced intervals and fitting a smooth curve through the ten points.

In the following paragraphs, we provide the technical details for generating a stopping boundary from a spending function. We assume throughout that the study is designed for a total of K looks at the information fractions t_1, t_2, \dots, t_K . A one sided test is assumed for simplicity. The extension to two-sided tests follows readily by replacing $W(t_j)$ throughout by $|W(t_j)|$.

Boundaries and Maximum Information for Early Rejection of H_0 Only

The boundaries are computed recursively, with $c(t_j)$ being based on the values of $c(t_l)$, $l = 1, 2, \dots, j - 1$. For the first look, at information fraction t_1 , find the upper boundary $c(t_1)$ such that

$$P_0(W(t_1) \geq c(t_1)) = \alpha(t_1) . \tag{B.55}$$

For subsequent looks $j = 2, 3, \dots, K$, having already computed the upper boundaries $c(t_1), c(t_2), \dots, c(t_{j-1})$, find the upper boundary $c(t_j)$ such that

$$\alpha(t_{j-1}) + P_0(W(t_1) < c(t_1), \dots, W(t_{j-1}) < c(t_{j-1}), W(t_j) \geq c(t_j)) = \alpha(t_j) . \tag{B.56}$$

These computations are performed by the AMR algorithm. Once the boundaries have been determined the maximum information is obtained by again invoking the AMR algorithm, this time to find the value of η that satisfies the type-2 error equation

$$P_\eta\left\{\bigcap_{j=1}^K W(t_j) < c(t_j)\right\} = \beta . \tag{B.57}$$

Denote the solution by η_1 . Then, by (B.6) the desired maximum information I_{\max} is

$$I_{\max} = \left[\frac{\eta_1}{\delta_1 - \delta_0}\right]^2$$

We can convert maximum information into maximum sample size or maximum events, depending on the model being used, by selecting the appropriate translation equation from (B.15) for normal endpoints, (B.20) or (B.23) for binomial endpoints, and (B.33) or (B.41) for time to event endpoints.

To obtain spending function boundaries for symmetric two-sided tests, replace $W(t_j)$ by $|W(t_j)|$ throughout equations (B.56) and (B.57).

Two-Sided Asymmetric Boundaries for Early Rejection of H_0 Only

Suppose one wishes to split the total type-1 error, α , of a two-sided test into two components α_l and α_u , with $\alpha_l + \alpha_u = \alpha$ in such a way that the probability, under the null hypothesis, of crossing the upper (lower) boundary is α_u (α_l). Denote the critical values of the two-sided boundary at interim monitoring time t_j by $(a(t_j), b(t_j))$, $j = 1, 2, \dots, K$. These boundary values are obtained by inverting corresponding spending function values $(\alpha_l(t_j), \alpha_u(t_j))$, $j = 1, 2, \dots, K$, as follows. For the first look, at information fraction t_1 , find the lower boundary $a(t_1)$ such that

$$P_0(W(t_1) \leq a(t_1)) = \alpha_l(t_1) , \tag{B.58}$$

B Group Sequential Design in East 6

and the upper boundary $b(t_1)$ such that

$$P_0(W(t_1) \geq b(t_1)) = \alpha_u(t_1) . \tag{B.59}$$

For subsequent looks $j = 2, 3, \dots, K$, having already computed the boundaries $\{(a(t_1), b(t_1)), (a(t_2), b(t_2)), \dots, (a(t_{j-1}), b(t_{j-1}))\}$ compute $(a(t_j), b(t_j))$ such that $\alpha_l(t_{j-1}) + P_0(a(t_1) < W(t_1) < b(t_1), \dots, a(t_{j-1}) < W(t_{j-1}) < b(t_{j-1}), W(t_j) \leq a(t_j)) = \alpha_l(t_j)$ (B.60)

and

$$\alpha_u(t_{j-1}) + P_0(a(t_1) < W(t_1) < b(t_1), \dots, a(t_{j-1}) < W(t_{j-1}) < b(t_{j-1}), W(t_j) \geq b(t_j)) = \alpha_u(t_j) . \tag{B.61}$$

We wish to point out that spending functions used to obtain the upper and lower boundaries in the above procedure can belong to different families if desired.

Boundaries and Maximum Information for Early Rejection of either H_0 or H_1

There are two stopping boundaries for these designs; an “upper” stopping boundary for early rejection of H_0 and a “lower” stopping boundary for early rejection of H_1 . We reject H_0 in favor of H_1 the first time we encounter an information fraction t_j such that a boundary is crossed and it is an upper boundary. We reject H_1 in favor of H_0 the first time we encounter an information fraction t_j such that a boundary is crossed and it is a lower boundary. We impose the constraint that the upper and lower boundaries must meet at t_K , thereby ensuring that a decision to reject either of the two hypotheses will indeed be made. The upper and lower stopping boundaries thus form a triangular continuation region.

We wish these stopping boundaries to have the property that at the null hypothesis, $\delta = 0$, we will cross the upper stopping boundary with probability α , but at the specific alternative hypothesis of interest, say $\delta = \delta_1$, we will cross the upper stopping boundary with probability $1 - \beta$ and the lower stopping boundary with probability β . The upper boundaries, u_j and the lower boundaries l_j , $j = 1, 2, \dots, K$, are found using a two-dimensional search to simultaneously solve two equations corresponding to the desired type-1 and type-2 errors of the test. The drift parameter η is determined simultaneously along with the boundaries. The procedure is specified below:

1. Set the drift parameter η to some arbitrary initial value $\eta = \eta_1$.
2. At the first look, at information fraction t_1 , search for the upper boundary u_1 such that

$$P_0(W(t_1) \geq u_1) = \alpha(t_1) , \tag{B.62}$$

and for the lower boundary l_1 such that

$$P_\eta(W(t_1) \leq l_1) = \beta(t_1) . \tag{B.63}$$

3. For subsequent looks $j = 2, 3, \dots, K - 1$, having already computed the pairs of boundaries up to and including (l_{j-1}, u_{j-1}) , find the upper boundary u_j such that

$$\begin{aligned} & P_0(W(t_1) \geq u_1) + P_0(l_1 < W(t_1) < u_1, W(t_2) \geq u_2) + \dots \\ & \dots + P_0(l_1 < W(t_1) < u_1, \dots, l_{j-1} < W(t_{j-1}) < u_{j-1}, W(t_j) \geq u_j) = \alpha(t_j) \end{aligned} \quad (\text{B.64})$$

and find the lower boundary l_j such that

$$\begin{aligned} & P_\eta(W(t_1) \leq l_1) + P_\eta(l_1 < W(t_1) < u_1, W(t_2) \leq l_2) + \dots \\ & \dots + P_\eta(l_1 < W(t_1) < u_1, \dots, l_{j-1} < W(t_{j-1}) < u_{j-1}, W(t_j) \leq l_j) = \beta(t_j). \end{aligned} \quad (\text{B.65})$$

4. At the K th and final look the upper boundary u_K satisfies

$$\begin{aligned} & P_0(W(t_1) \geq u_1) + P_0(l_1 < W(t_1) < u_1, W(t_2) \geq u_2) + \dots \\ & \dots + P_0(l_1 < W(t_1) < u_1, \dots, l_{K-1} < W(t_{K-1}) < u_{K-1}, W(t_K) \geq u_K) = \alpha. \end{aligned} \quad (\text{B.66})$$

Since we want to reach a decision at the last look (either in favor of the null or the alternative) we have to set the lower boundary l_K equal to the upper boundary u_K . Thus set $l_K = u_K$ and find the value of β^* by calculating

$$\begin{aligned} & \beta^* = P_\eta(W(t_1) \leq l_1) + P_\eta(l_1 < W(t_1) < u_1, W(t_2) \leq l_2) + \dots \\ & \dots + P_\eta(l_1 < W(t_1) < u_1, \dots, l_{K-1} < W(t_{K-1}) < u_{K-1}, W(t_K) \leq l_K). \end{aligned} \quad (\text{B.67})$$

- (a) If $\beta^* = \beta$, then the set of boundaries just computed satisfy the required operating characteristics at the alternative $\eta = \eta_1$.
- (b) If $\beta^* > \beta$ select a new value $\eta_{\text{new}} < \eta_1$. Set $\eta_1 = \eta_{\text{new}}$ and repeat the steps from Step 2 onward.
- (c) If $\beta^* < \beta$ select a new value $\eta_{\text{new}} > \eta_1$. Set $\eta_1 = \eta_{\text{new}}$ and repeat the steps from Step 2 onward

The above iterative procedure ends with simultaneous computation of the final stopping boundaries and the final drift parameter η_1 . Then, by (B.6) the desired maximum information I_{max} is

$$I_{\text{max}} = \left[\frac{\eta_1}{\delta_1 - \delta_0} \right]^2$$

We can convert maximum information into maximum sample size or maximum events, depending on the model being used, by selecting the appropriate translation equation from (B.15) for normal endpoints, (B.20) or (B.23) for binomial endpoints, and (B.33) or (B.41) for time to event endpoints.

Two-sided boundaries for early rejection of H_0 or H_1 are obtained by replacing $W(t_j)$ with $|W(t_j)|$. The boundary for early rejection of H_0 in the one-sided case is now

B Group Sequential Design in East 6

replaced by two boundaries for early rejection of H_0 , symmetrically placed on either side of the X-axis. Similarly the boundary for early rejection of H_1 is now replaced by two boundaries for early rejection of H_1 , symmetrically placed on either side of the X-axis, and constructed so as to meet their corresponding H_0 -rejection boundaries at the final look. This results in two triangular continuation regions and an inner wedge.

Making the H_0 - H_1 Boundaries Non-Binding

In the discussion that follows we will refer to the boundary for early rejection of H_0 as the “efficacy” boundary and the boundary for early rejection of H_1 as the “futility” boundary. Equations (B.62) through (B.67) were used to generate the efficacy and futility boundaries simultaneously. One practical drawback of this simultaneous computation is that the futility boundary cannot be overruled. In other words, if the test statistic crosses the futility boundary the trial **must** be terminated, or else the type-1 error might be inflated. This is so because the interaction between the two boundaries during their construction causes the efficacy boundary to be shifted relative to the position it would have occupied if there were no futility boundary. It could happen, for example, that the presence of the futility boundary “pulls down” the efficacy boundary, making it easier to cross under the null hypothesis, if the futility boundary can be arbitrarily overruled. If the efficacy boundary is disturbed in this manner, the only way to prevent the possibility of inflating the type-1 error is to make the futility boundary strictly binding. This is usually not acceptable to the sponsor of a clinical trial or to the data monitoring committee assigned to the trial. This is the primary motivation for constructing non-binding futility boundaries.

We now show how to simultaneously compute the efficacy and futility boundaries in such a way that the early rejection criteria of the efficacy boundary remain the same as the corresponding criteria in a H_0 -only design. In that case there is no danger of inflating the type-1 error even if the futility boundary is overruled. The only cost of this added flexibility is an increase in the maximum information. For ease of exposition we will only describe the one-sided H_0 - H_1 case

1. Generate the one-sided level- α efficacy boundary as specified by equations (B.55) and (B.56). Denote this boundary by $\{u_1, u_2, \dots, u_K\}$.
2. For this boundary find the value of η that will satisfy the type-2 error equation (B.57).
3. Keeping this value of η and the previously obtained efficacy boundary values $\{u_1, u_2, \dots, u_K\}$ fixed, compute the futility boundary $\{l_1, l_2, \dots, l_K\}$ as follows:

$$P_\eta(W(t_1 \leq l_1) = \beta(t_1)) \quad (\text{B.68})$$

and for $j = 2, 3, \dots, K - 1$,

$$P_\eta(W(t_1) \leq l_1) + P_\eta(l_1 < W(t_1) < u_1, W(t_2) \leq l_2) + \dots \\ \dots + P_\eta(l_1 < W(t_1) < u_1, \dots, l_{j-1} < W(t_{j-1}) < u_{j-1}, W(t_j) \leq l_j) = \beta(t_j). \quad (\text{B.69})$$

Since the efficacy and futility boundaries are required to meet at look K simply set $l_K = u_K$.

4. Compute the power of a K -look design utilizing these boundaries with drift parameter η by evaluating

$$P_\eta(W(t_1) \geq u_1) + P_\eta(l_1 < W(t_1) < u_1, W(t_2) \geq u_2) + \dots \\ \dots + P_\eta(l_1 < W(t_1) < u_1, \dots, l_{K-1} < W(t_{K-1}) < u_{K-1}, W(t_K) \geq u_K) \quad (\text{B.70})$$

Denote this power by $1 - \beta_{i-1}^*$.

5. Repeat Step 4 with progressively increasing values of η until $1 - \beta^*$ is equal to the desired power $1 - \beta$. At that point denote final the drift parameter by η_1 . Then, by (B.6) the desired maximum information I_{\max} is

$$I_{\max} = \left[\frac{\eta_1}{\delta_1 - \delta_0} \right]^2.$$

We can convert maximum information into maximum sample size or maximum events, depending on the model being used, by selecting the appropriate translation equation from (B.15) for normal endpoints, (B.20) or (B.23) for binomial endpoints, and (B.33) or (B.41) for time to event endpoints.

The above iterative procedure produces efficacy and futility boundaries having the property that the probability of crossing the efficacy boundary under the alternative hypothesis $\delta = \delta_1$ is $1 - \beta$. Thus the desired power is obtained. Next, since the efficacy boundary was computed at Step 1 in the absence of a futility boundary, and was never altered in any subsequent step, the probability of crossing it under the null hypothesis is at most α . This probability is exactly equal to α if the futility boundary is always overruled and can only decrease if the futility boundary is respected at one or more looks. Thus, in either case the type-1 error cannot exceed α . This shows that boundaries constructed as described above produce the desired power and preserve the type-1 error with the added flexibility that the futility boundary can be overruled.

Boundaries and Maximum Information for Early Rejection of H_1 Only

Boundaries for early rejection of H_1 only are also known as futility boundaries. They are obtained by only spending the β error at the interim looks according to a β -spending function. The α error is spent in its entirety at the last look. These boundaries and the associated maximum information can therefore be obtained by setting $\alpha(t_j) = 0$ for $j = 1, 2, \dots, K - 1$ in equations (B.62), (B.64) and (B.66). Equations (B.63), (B.65), (B.66) and (B.67) are unchanged. The computations then proceed as before.

B Group Sequential Design in East 6

B.2.5 Special Considerations for Binomial Designs

Maximum information and maximum sample size computations for binomial designs are complicated by the dependence of the variance of a binomial random variable on its mean. Therefore, even if we keep all other design parameters the same, the required maximum sample size for a binomial trial may differ, depending on how we intend to estimate the variance of the treatment difference at the interim monitoring stage. Although this special consideration applies both to superiority trials as well as to non-inferiority trials, the present discussion will be restricted to superiority trials only, where East provides two options at the design stage.

The issue is, how will the observed treatment difference

$$\hat{\delta}(t_j) = \hat{\pi}_t(t_j) - \hat{\pi}_c(t_j)$$

be standardized at the interim monitoring stage of the trial? The standardization method one intends to use at the interim monitoring stage must be reflected in the computation of sample size at the design stage. In East we offer two options.

Unpooled Estimate the variance without pooling the data from the two treatment arms. Thus

$$\text{var}[\hat{\delta}(t_j)] = \frac{\hat{\pi}_t(t_j)(1 - \hat{\pi}_t(t_j))}{n_{tj}} + \frac{\hat{\pi}_c(t_j)(1 - \hat{\pi}_c(t_j))}{n_{cj}}, \quad (\text{B.71})$$

which implies that the statistic $\tilde{W}_s(t_j)$ given by equation (B.24) will be used to monitor the data. We have already seen in Section B.1.2 that this statistic is asymptotically $N(0, t_j)$ under the null hypothesis, asymptotically $N(\eta t_j, t_j)$ under the alternative hypothesis, and has independent increments. Therefore all the computations discussed in Section B.2 for obtaining stopping boundaries, estimating the maximum information I_{\max} , and converting maximum information into maximum sample size N_{\max} , remain valid without any modifications. In the East software, the unpooled estimate of variance is the default for the design of binomial endpoint trials.

Pooled Estimate the variance after pooling the data from the two treatments. Thus

$$\text{var}[\hat{\delta}(t_j)] = \frac{\hat{\pi}(t_j)(1 - \hat{\pi}(t_j))}{n_j(r)(1 - r)} \quad (\text{B.72})$$

where $\hat{\pi}(t_j)$, the **pooled** estimate of the binomial response probability at time t_j , is given by equation (B.26). This implies that the statistic $\tilde{W}_{0s}(t_j)$ given by equation (B.25) will be used to monitor the data. As already stated in Section B.1.2, $\tilde{W}_{0s}(t_j)$ is $N(0, t_j)$ with independent increments under

$H_0: \delta = 0$. However, since the variance term (B.72) is based on a pooled estimate of response, the distribution of $\tilde{W}_{0s}(t_j)$ is no longer $N(\eta t_j, t_j)$ under the alternative hypothesis. Therefore if we intend to use the pooled estimate of variance at the interim monitoring stage, the computation of I_{\max} and N_{\max} must be modified for H_0 -only boundaries, and the computation of stopping boundaries, I_{\max} and N_{\max} must be altered for H_0 - H_1 boundaries. These modifications are described below.

First consider the case of H_0 -only boundaries. For expository purposes we will only consider boundaries derived from α -spending functions for one-sided tests. The same approach also works for the Haybittle-Peto and Wang-Tsiatis boundaries, and for two-sided tests. Since $\tilde{W}_{0s}(t_j) \sim N(0, t_j)$ with independent increments, the boundaries $\{c(t_1), c(t_2), \dots, c(t_K)\}$ generated by equation (B.56) will preserve the type-1 error without any modification. These boundaries cannot, however, be directly utilized by equation (B.57) because $\tilde{W}_{0s}(t_j)$ is not $N(\eta t_j, t_j)$ under the alternative hypothesis. It is easy to show, however, that asymptotically

$$P_{\eta} \left\{ \bigcap_{j=1}^K \tilde{W}_{0s}(t_j) < c(t_j) \right\} \approx P_{\eta} \left\{ \bigcap_{j=1}^K \tilde{W}_s(t_j) < hc(t_j) \right\} \tag{B.73}$$

where

$$h = \sqrt{\frac{\bar{\pi}(1 - \bar{\pi})(r^{-1} + (1 - r)^{-1})}{\pi_e(1 - \pi_e)r^{-1} + \pi_c(1 - \pi_c)(1 - r)^{-1}}} \tag{B.74}$$

and

$$\bar{\pi} = r\pi_e + (1 - r)\pi_c. \tag{B.75}$$

Since $\tilde{W}_s(t_j)$ is asymptotically distributed as $N(\eta t_j, t_j)$ with independent increments, we can estimate the maximum information, I_{\max} , by finding the value of η that satisfies the following modification of equation B.57:

$$P_{\eta} \left\{ \bigcap_{j=1}^K W(t_j) < hc(t_j) \right\} = \beta. \tag{B.76}$$

For H_0 - H_1 boundaries the modification is slightly more complex since in this case the stopping boundaries are not computed independently of I_{\max} . The procedure is identical to the four-step procedure outlined on page 2294 with the following modification: for any equation involving $\beta(t_j)$ on the right hand side, replace (l_j, u_j) by (hl_j, hu_j) .

B Group Sequential Design in East 6

The test statistic $W_{0s}(t_j)$ based on the pooled variance can be transformed into

$$X_{0s}^2(t_j) = \left[\frac{W_{0s}(t_j)}{\sqrt{(t_j)}} \right]^2 \tag{B.77}$$

which reduces to the familiar Pearson chi-square statistic.

The option to base the design on the pooled estimate of variance is being offered because the chi-square test is a popular method for comparing two binomial populations. For a fixed sample study ($K = 1$) and the sample size obtained by the pooled approach specializes to the following formula given by Lachin (1981):

$$N = \left[\frac{z_\alpha \sqrt{\bar{\pi}(1 - \bar{\pi})(r^{-1} + (1 - r)^{-1})} + z_\beta \sqrt{\pi_e(1 - \pi_e)r^{-1} + \pi_c(1 - \pi_c)(1 - r)^{-1}}}{\delta_1} \right]^2 \tag{B.78}$$

In contrast the sample size for a fixed sample design based on the unpooled estimate of variance is

$$N = [\pi_e(1 - \pi_e)r^{-1} + \pi_c(1 - \pi_c)(1 - r)^{-1}] \times \left[\frac{z_\alpha + z_\beta}{\delta_1} \right]^2 . \tag{B.79}$$

We shall show in the next section that when $K > 1$ the above sample sizes are multiplied by an appropriate inflation factor that takes into account the number of looks, K , as well as the type of stopping boundary. For balanced designs ($r \approx 0.5$) the maximum sample sizes for the pooled and unpooled methods are very similar. If, however, the design is severely unbalanced, there can be substantial differences in the maximum sample sizes required to attain the desired power. It follows from equations (B.73) and (B.74) that if $h < 1$, the pooled variance will produce a more powerful test than the unpooled variance, whereas if $h > 1$, the unpooled variance will produce a more powerful test than the pooled variance. We have illustrated these points with examples in Chapter 23.

B.3 The Inflation Factor

B.3.1 General Designs

B.3.2 Information Based Designs

B.3.3 G versus I Designs

It should be clear from the manner in which the drift parameter was computed in the previous section that its value depends on K, α, β and the stopping boundary or spending function selected for the design. Therefore, in this section we will recognize explicitly that drift parameter is a function of these items by denoting it as $\eta(\alpha, \beta, K, \text{boundaries})$.

The relationship

$$I_{\max} = \left[\frac{\eta_1(\alpha, \beta, K, \text{boundaries})}{\delta_1 - \delta_0} \right]^2$$

implied by equation (B.6) is equivalent to

$$I_{\max} = \left[\frac{z_{\alpha} + z_{\beta}}{\delta_1 - \delta_0} \right]^2 \times \left[\frac{\eta_1(\alpha, \beta, K, \text{boundaries})}{z_{\alpha} + z_{\beta}} \right]^2. \tag{B.80}$$

Observe that the first term in equation (B.80) is the information needed to achieve $1 - \beta$ power at an effect size of δ_1 for a single-look, one-sided, level- α study of the null hypothesis $\delta = \delta_0$. We denote this term by

$$I_1 = \left[\frac{z_{\alpha} + z_{\beta}}{\delta_1 - \delta_0} \right]^2. \tag{B.81}$$

The second term is a multiplier for inflating the information required by the single-look study so that it will preserve the desired power of $1 - \beta$ if $K > 1$ looks are taken. We refer to the second term as the inflation factor and denote it by

$$\text{IF}(\alpha, \beta, K, \text{boundaries}) = \left[\frac{\eta_1(\alpha, \beta, K, \text{boundaries})}{z_{\alpha} + z_{\beta}} \right]^2 \tag{B.82}$$

If we denote I_{\max} by I_K for a K -look group sequential study, we have the relationship

$$I_K = I_1 \times \text{IF}(\alpha, \beta, K, \text{boundaries}). \tag{B.83}$$

In Table B.1 we have tabulated the inflation factors for some common choices of α, β, K and the shape parameter, Δ for the Wang-Tsiatis boundaries.

Table B.1: Inflation Factors for Pocock ($\Delta = 0.5$) and O’Brien-Fleming ($\Delta = 0$) Stopping Boundaries

| $\alpha = 0.05$ (two-sided) | | | | | $\alpha = 0.01$ (two-sided) | | | | |
|-----------------------------|-------------------|-----------------------|------|------|-----------------------------|-------------------|-----------------------|------|------|
| K | Stopping Boundary | Power ($1 - \beta$) | | | K | Stopping Boundary | Power ($1 - \beta$) | | |
| | | 0.80 | 0.90 | 0.95 | | | 0.80 | 0.90 | 0.95 |
| 2 | Pocock | 1.11 | 1.10 | 1.09 | 2 | Pocock | 1.09 | 1.08 | 1.08 |
| 2 | O-F | 1.01 | 1.01 | 1.01 | 2 | O-F | 1.00 | 1.00 | 1.00 |
| 3 | Pocock | 1.17 | 1.15 | 1.14 | 3 | Pocock | 1.14 | 1.12 | 1.12 |
| 3 | O-F | 1.02 | 1.02 | 1.02 | 3 | O-F | 1.01 | 1.01 | 1.01 |
| 4 | Pocock | 1.20 | 1.18 | 1.17 | 4 | Pocock | 1.17 | 1.15 | 1.14 |
| 4 | O-F | 1.02 | 1.02 | 1.02 | 4 | O-F | 1.01 | 1.01 | 1.01 |
| 5 | Pocock | 1.23 | 1.21 | 1.19 | 5 | Pocock | 1.19 | 1.17 | 1.16 |
| 5 | O-F | 1.03 | 1.03 | 1.02 | 5 | O-F | 1.02 | 1.01 | 1.01 |

B.3.1 Role of Inflation Factors in General Designs

The inflation factor is a convenient device for converting fixed sample studies into

B Group Sequential Design in East 6

corresponding group sequential studies. This is the basis of the General design module. In this module East accepts the sample size (or information) required for a single-look study with a given power and type-1 error. East then uses the appropriate inflation factor to convert the single-look study into a K -look group sequential study. This is useful when we are required to design a group sequential study to evaluate some end point that is not currently available directly in East. (For example, the end point might be the comparison of two Poisson rates, or it might be a covariate in a logistic regression model). The first step is to obtain the sample-size or statistical information that would be required if this were a fixed-sample study. This can be done with the help of any convenient non-sequential design package. The sample size so obtained is then inflated by the appropriate inflation factor based on the desired number of looks, significance level, power and stopping boundary desired for the group sequential trial. See Chapter 60 for examples where East designs and monitors general studies of this type.

B.3.2 Role of Inflation Factor in Information Based Designs

Suppose one wishes to test $H_0: \delta = \delta_0$ versus $H_1: \delta = \delta_1$ where δ is a scalar parameter of interest in some mathematical model of the data generating process. In the Information Based Design module of East one specifies $\delta_1 - \delta_0$. East then computes the required fixed sample information through equation (B.81) and inflates it appropriately for a K look group sequential study through equation (B.82). The information is expressed in the dimensionless units of $[\text{se}(\hat{\delta}(\tau))]^{-2}$. The study is then monitored on this information scale.

Designing a study so that the information will be in the dimensionless units of $[\text{se}(\hat{\delta}(\tau))]^{-2}$ has both advantages and disadvantages. The disadvantage is that, prior to activating the study, one needs to interpret the desired information in terms of a physical resource like sample size or number of failures. The formula for making this conversion depends on the mathematical model of the data generating process. Sometimes a closed-form formula exists, but for more complex models one must resort to simulation. (See, for example, Scharfstein, Tsiatis and Robins, 1997, or Scharfstein and Tsiatis 1998.) Additionally, the conversion usually depends on initial estimates of nuisance parameters like the baseline response rate or the other covariates in the mathematical model of the data generating process. If we estimate the values of these nuisance parameters incorrectly, the sample size (or other physical resource) too will be incorrect and the study will not have the operating characteristics it was intended to have. For this reason it is often preferable to design the study and implement the interim monitoring on the dimensionless information scale where we do not require any knowledge about the nuisance parameters. Provided we continue to monitor the data until either full information I_K is achieved (in terms of the desired $[\text{se}(\hat{\delta}(\tau))]^{-2}$),

or else a stopping boundary is crossed, we are assured of preserving the operating characteristics of the design. We might of course wish to update the conversion of the desired statistical information into a physical resource like sample size at each interim monitoring time point, as revised estimates of the nuisance parameters become available. Illustrative examples in which the value of I_K remains constant on the dimensionless information scale but changes on the sample size scale, as more accurate estimates of nuisance parameters are obtained, are given in Chapter 59.

B.3.3 Selecting the General versus the Information Based Option

The General (**G**) and the Information Based (**I**) modules in East both rely on the same general distribution theory developed by Scharfstein, Tsiatis and Robins (1997) and by Jennison and Turnbull (1997). In both cases an inflation factor is applied to a corresponding fixed sample design as discussed at the beginning of this section. The question then arises as to which module to adopt for a given problem. Here are some guidelines.

1. If software to design the corresponding single-look study is available, the **G** module is easier to use than **I** module since the information is measured in terms of a physical resource like sample size or number of events.
2. If software to design the corresponding single-look study is not available, the **I** module can still be used since it only requires one to input the size of the treatment effect δ_1 under the alternative hypothesis. The **I** module, however, specifies the maximum required information in terms of a dimensionless quantity representing the inverse square of the standard error of the parameter being tested. It is usually necessary to translate this dimensionless information into a physical resource, either through simulation or analytically.
3. The **I** module is preferable to the **G** module in situations where the model for generating the data contains unknown nuisance parameters like the variance, the baseline response, or the coefficients of covariates in a regression model. To use the **G** module one would have to make assumptions about these unknown nuisance parameters. But the **I** module only requires you to specify the magnitude of the treatment effect you are interested in detecting.
4. The **I** module facilitates sample-size re-estimation since the maximum information is specified in dimensionless units that remain constant while the translation of maximum information into the corresponding sample size can be made more accurate at each interim look as increasingly accurate estimates of nuisance parameters become available.

B Group Sequential Design in East 6

B.4 Computation of Expected Information

- B.4.1 Exit Probabilities
- B.4.2 Expected Sample Size
- B.4.3 Expected Events

In Section B.1 we defined the maximum information, $I_{\max} \equiv I_K$, to be committed up-front for a K look group sequential clinical trial, and in Section B.2 we showed how to compute this quantity for various stopping boundaries. In practice of course a group sequential study might be terminated earlier than the K th look because of the sequential monitoring. Thus the actual information is a random variable. In this section we show how to compute the probability of crossing the stopping boundaries at each interim look. We then derive from these exit probabilities, the expected value of the information that will be obtained in a group sequential clinical trial. For normal and binomial end points, information will be represented by sample size. For time to failure end points, information will be represented by the number of failures.

B.4.1 Boundary Crossing Probability at Each Look

Let u_1, u_2, \dots, u_K be the upper stopping boundaries for a one-sided group sequential test with possible early rejection of H_0 . The probability of boundary crossing for the first time at look j is

$$P_{bc,j} = P_{\eta}[W(t_1) < u_1, W(t_2) < u_2, \dots, W(t_j) > u_j]$$

East computes and displays this boundary crossing probabilities under both H_0, H_1 and $H_{1/2}$ (half way between the null and alternative hypotheses) for all $j = 1, 2, \dots, K$.

Similarly for one-sided tests allowing for early to reject either H_0 or H_1 , East computes and displays

$$P_{bc,j} = P_{\eta}[l_1 < W(t_1) < u_1, \dots, l_{j-1} < W(t_{j-1}) < u_{j-1}, W(t_j) > u_j] + P_{\eta}[l_1 < W(t_1) < u_1, \dots, l_{j-1} < W(t_{j-1}) < u_{j-1}, W(t_j) < l_j]$$

under H_0, H_1 and $H_{1/2}$.

Similar displays are also available for the two-sided tests. For two-sided tests we simply replace $W(t_j)$ with $|W(t_j)|$ in the above boundary crossing probability equations.

B.4.2 Expected Sample Sizes for Normal and Binomial Studies

In general, for a study with K interim analyses performed at information fractions t_1, t_2, \dots, t_K , the expected stopping time, E_t , can be computed under various hypotheses on the basis of the boundary crossing probabilities as follows:

$$E_t = \sum_{j=1}^{K-1} t_j \times P_{bc,j} + 1 - \sum_{j=1}^{K-1} P_{bc,j} .$$

The expected sample size is then computed as

$$E_N = N_{\max} E_t, \quad (\text{B.84})$$

where N_{\max} is the projected maximum sample size, evaluated as discussed at the end of each subsection of Section B.2. The expected value E_N is referred to as ASN or average sample number in some of the East charts.

B.4.3 Expected Number of Events for Survival Studies

We have shown in each subsection of Section B.2 that the value of I_{\max} needed to detect $\delta = \delta_1$ with $1 - \beta$ power is given by

$$I_{\max} = \left[\frac{\eta_1}{\delta_1 - \delta_0} \right]^2. \quad (\text{B.85})$$

Furthermore, equations (B.33) and (B.41) show that I_{\max} is directly proportional to D_{\max} , the maximum number of events. It follows that

$$D_{\max} = \frac{1}{(r)(1-r)} \left[\frac{\eta_1}{\delta_1 - \delta_0} \right]^2, \quad (\text{B.86})$$

where δ_0 and δ_1 are specified by the null and alternative hypotheses, r is the proportion randomized to treatment T under the alternative hypothesis, and η_1 is computed along with the stopping boundaries as discussed in Section B.2. The expected number of events is thus

$$E_D = D_{\max} E_t, \quad (\text{B.87})$$

B.5 Sample Size and Expected Study Duration for Survival Studies

Equation (B.86) shows that the power of a time-to-event trial is determined by the maximum number of events, D_{\max} , rather than the maximum sample size, N_{\max} . However, the total time one must wait for the D_{\max} events to arrive can be controlled through sample size. The larger the sample size, the faster the required number of events are expected to arrive. A typical time-to-event trial is characterized by an accrual phase during which new subjects are enrolled, and a follow-up phase during which there is no further enrollment but subjects continue to be followed until the required number of events have been observed. A longer accrual phase implies a shorter follow-up phase, and usually also implies a shorter total study duration. Kim and Tsatis (1990) analyzed this trade off for the simplest possible case in which subjects enroll at a constant rate a for a fixed period S_a , there are no drop-outs, the survival distributions for the two treatment arms are exponential, and all subjects who have not yet experienced the event are followed until the trial is terminated. For this

B Group Sequential Design in East 6

special case they calculated that the expected number of events at calendar time l given a constant hazard rate of λ is

$$E(l|a, S_a, \lambda) = \begin{cases} a[l - \frac{1-e^{-\lambda l}}{\lambda}] & \text{if } l \leq S_a \\ a[S_a - \frac{e^{-\lambda l}}{\lambda}(e^{\lambda S_a} - 1)] & \text{if } l > S_a \end{cases} \quad (\text{B.88})$$

In Appendix D we have generalized (B.88) to handle variable enrollment rates, drop-outs, and piece-wise exponential survival, for both the **variable follow-up setting**, where the each subject still on-study is followed until the trial ends, and the **fixed follow-up setting**, where each subject still on-study is followed for a fixed amount of time, m . The generalized expression is denoted here by

$$E(l|\mathbf{a}, S_{\mathbf{a}}, \boldsymbol{\lambda}, \gamma, m) = \text{expected number of events at calendar time } l \text{ given}$$

- \mathbf{a} : a vector of enrollment rates for different intervals in the enrollment phase;
- $S_{\mathbf{a}}$: a vector of enrollment durations corresponding to the components of \mathbf{a} ;
- $\boldsymbol{\lambda}$: a vector of hazard rates for piece-wise exponential survival;
- γ : a drop-out rate for subjects lost to follow-up;
- m : a fixed follow-up time for each subject ($m = \infty$ denotes variable follow-up).

Thus, if the fraction randomized to the treatment arm is r , the expected number of events from both arms together at calendar time l is

$$E_{(1)}(l|\mathbf{a}, S_{\mathbf{a}}, \boldsymbol{\lambda}, \gamma, m) = rE(l|\mathbf{a}, S_{\mathbf{a}}, \boldsymbol{\lambda}_T, \gamma_T, m) + (1 - r)E(l|\mathbf{a}, S_{\mathbf{a}}, \boldsymbol{\lambda}_C, \gamma_C, m), \quad (\text{B.89})$$

where $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_E, \boldsymbol{\lambda}_C)$ and $\boldsymbol{\gamma} = (\gamma_E, \gamma_C)$. A chart displaying $E(l|\mathbf{a}, S_{\mathbf{a}}, \boldsymbol{\lambda}_C, \gamma, m)$, $E(l|\mathbf{a}, S_{\mathbf{a}}, \boldsymbol{\lambda}_T, \gamma, m)$ and $E_{(1)}(l|\mathbf{a}, S_{\mathbf{a}}, \boldsymbol{\lambda}, \gamma, m)$ versus calendar time l can be displayed by clicking on the  icon in the **Plots** menu of East's Library.

B.5.1 Estimating Maximum Expected Study Duration

In a K -look group sequential trial we are committed to keeping the trial open until D_{\max} events are observed, unless a stopping boundary is crossed earlier. Although the actual calendar time at which these D_{\max} events will occur is a random variable it is nevertheless useful, for design purposes, to compute the calendar time, l_{\max} say, at which we would expect to observe D_{\max} events under various assumptions about accrual, drop-outs and survival distributions. Therefore we solve for l_{\max} from the

equation

$$E_{(1)}(l_{\max}|a, S_a, \lambda, \gamma, m) = D_{\max} . \tag{B.90}$$

The solution to l_{\max} in the above equation is obtained by iteration and represents the maximum length of time that we would expect the study to remain open if no early stopping boundary was crossed.

B.5.2 Trading-Off Maximum Study Duration Against Sample Size

We now establish a trade-off between the maximum expected study duration l_{\max} and sample size. In order to present the essential features of this trade-off, we will only discuss the special case where enrollment is constant at a rate a per unit time and the duration of the enrollment phase is S_a . In Appendix D we show that East does indeed handle the more general case in which there are Q distinct enrollment rates $\mathbf{a} = (a_1, a_2, \dots, a_Q)$ with corresponding enrollment durations denoted by $\mathbf{S}_a = (S_{a_1}, S_{a_2}, \dots, S_{a_Q})$. However a detailed discussion of the general case in the present section would be a distraction. It involves more complex notation and would needlessly prolong the discussion without providing any additional insight about the trade-off involved.

Case (i): Variable Follow-Up Design ($m = \infty$)

In this design subjects are enrolled for S_a units of time. All subjects are followed until the trial ends, unless they drop out or achieve the endpoint before trial termination. Thus the first subject enrolled could potentially be followed for l_{\max} units of time while the last subject enrolled could potentially be followed for $l_{\max} - S_a$ units of time.

We may express the maximum expected study duration in the form

$$l_{\max} = S_a + S_f \tag{B.91}$$

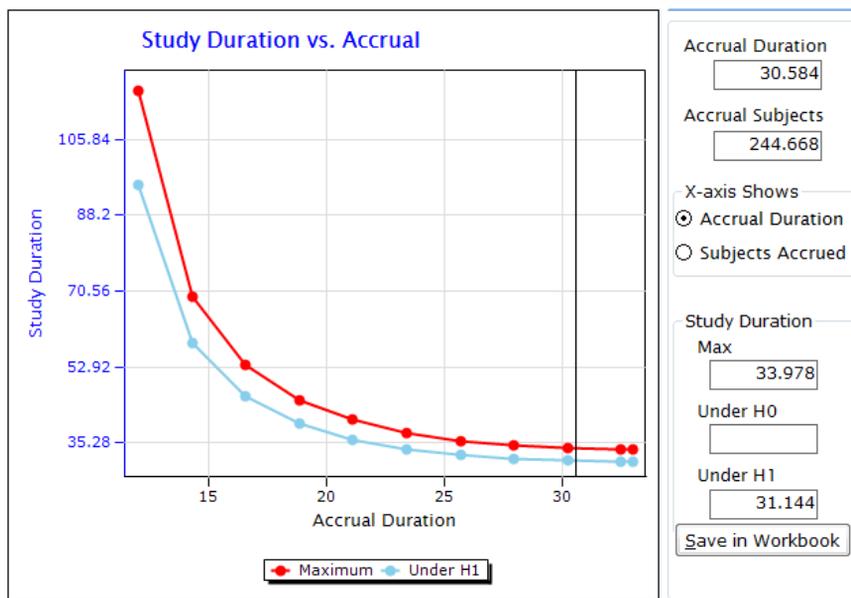
where S_f is the duration of the follow-up phase of the trial. Then, for a fixed value of S_a , East determines the value of S_f such that

$$E_{(1)}(S_a + S_f|a, S_a, \lambda, \gamma, \infty) = D_{\max} . \tag{B.92}$$

(Observe that the symbol $m = \infty$ has been used in the above expression for $E_{(1)}(\cdot|a, S_a, \lambda, \gamma, m)$, thereby indicating that this is a variable follow-up design.) By entering different enrollment durations, S_a , into equation (B.92) one obtains corresponding follow-up durations, S_f , and hence also obtains corresponding maximum study durations, $l_{\max} = S_a + S_f$. Graphs of l_{\max} versus study duration, S_a ,

B Group Sequential Design in East 6

and l_{\max} versus sample size, aS_a , are obtained by clicking on the  icon in the **Plots** menu of East's Library. A typical pair of graphs is shown below:



The red line on the graph displays the maximum expected study duration, l_{\max} , versus enrollment duration S_a . The calculation of l_{\max} is done under the alternative hypothesis and does not take into consideration a possible shortening of the study duration caused by the early stopping. The blue graph on the graph displays the expected study duration under the alternative hypothesis that accounts for the possibility of early stopping. Similar graphs can be obtained for l_{\max} (or its expectation under H_1) versus the sample size aS_a . All these relationships are monotone decreasing, highlighting that the greater the duration of the enrollment phase, or number of patients enrolled, the shorter the follow up phase and hence the shorter the total expected study duration.

We can establish a range of acceptable enrollment durations, $(S_{a,\min}, S_{a,\max})$, as well as a range of corresponding acceptable sample sizes $(aS_{a,\min}, aS_{a,\max})$ within which it is reasonable to make a selection. To determine $S_{a,\max}$ we argue that it is not necessary to prolong the enrollment phase beyond the time required to obtain D_{\max} events. Thus we search iteratively for the value of l at which

$$E_{(1)}(l|a, S_a = l, \lambda, \gamma, \infty) = D_{\max} \tag{B.93}$$

$S_{a,max}$, is the value of l that solves equation (B.93).

To determine $S_{a,min}$ we start with the smallest possible enrollment duration $S_a^* = D_{max}/a$ and see if it is feasible. To determine feasibility we progressively increase the follow-up time, starting with $S_f = 0$, and compute $E_1(S_a^* + S_f|a, S_a^*, \lambda, \gamma, \infty)$. If it should turn out that no matter how large we make S_f we always have

$$E_{(1)}(S_a^* + S_f|a, S_a^*, \lambda, \gamma, \infty) < D_{max} ,$$

then the current value of S_a^* is not feasible. In that case we increase the enrollment duration by a small amount ϵ . After setting $S_a^* \leftarrow S_a^* + \epsilon$, we once again test for feasibility by computing $E_1(S_a^* + S_f|a, S_a^*, \lambda, \gamma, \infty)$ with progressively increasing values of S_f . We iterate in this manner until we finally obtain the smallest possible S_a^* , denoted by $S_{a,min}$, such that there exists a value of S_f at which

$$E_1(S_a^* + S_f|a, S_a^*, \lambda, \gamma, \infty) = D_{max} \tag{B.94}$$

The solution, $S_{a,min}$, is the smallest that we can make the enrollment period and still hope to obtain D_{max} events. If there are no drop-outs, $S_{a,min} = \frac{D_{max}}{a}$, but in the presence of drop-outs, $S_{a,min} > \frac{D_{max}}{a}$.

East displays the enrollment duration range ($S_{a,min}, S_{a,max}$) (as well as corresponding sample size range ($aS_{a,min}, aS_{a,max}$)), and selects the mid-point of this range as the default. The user can change this default value and thereby trade-off sample size against total study duration.

| Accrual | | | |
|--|------|--------|------------|
| | Min. | Comtd. | Sugg. Max. |
| <input type="radio"/> Duration: | 11 | 22.25 | 33.5 |
| <input checked="" type="radio"/> Subjects: | 88 | 178 | 268 |

Case (ii): Designs with Fixed Follow-Up m

In many trials the clinical endpoint is of interest only if it is obtained within a fixed time period m . For example, in trials involving acute coronary syndrome, the primary question is whether the clinical endpoint (e.g., death, MI or refractory ischemia) has occurred within $m = 30$ days of entry into the study. In such trials each subject is only followed for a maximum of m units of time, after which the subject goes off-study. Therefore the maximum study duration is actually fixed at m units after the

B Group Sequential Design in East 6

last subject has been enrolled; i.e., at time $S_a + m$. The design question is to determine the value of S_a that will ensure that

$$E_{(1)}(S_a + m | a, S_a, \lambda, \gamma, m) = D_{\max} . \tag{B.95}$$

Here m is fixed and we iterate on S_a until equation (B.95) is satisfied. Denote this solution by $S_{a,\min}$. In this case if we enroll $aS_{a,\min}$ subjects and follow each subject for a maximum of m units of time we expect to obtain the desired D_{\max} events at time

$$l_{\max} = S_{a,\min} + m .$$

If we enroll for longer than $S_{a,\min}$ units of time then the desired D_{\max} events are expected to arrive before $S_{a,\min} + m$. In particular if the duration of enrollment extends up to $S_{a,\max}$ units of time, where

$$E_{(1)}(S_{a,\max} | a, S_{a,\max}, \lambda, \gamma, m) = D_{\max} , \tag{B.96}$$

then the desired D_{\max} events will have arrived by the end of the enrollment phase itself. Therefore East specifies that $(S_{a,\min}, S_{a,\max})$ is an acceptable range within which to select the enrollment duration and $(aS_{a,\min}, aS_{a,\max})$ is an acceptable range within which to select the corresponding sample size. Unlike the variable follow-up case where the mid-point of the range is selected as the default, East selects $S_{a,\min}$ ($aS_{a,\min}$) as default choice for the enrollment duration (sample size). With this choice, the trial is expected to be fully powered precisely when the last subject enrolled has been followed for m units of time.

| Accrual | | | |
|--|--------|--------|------------|
| | Min. | Comtd. | Sugg. Max. |
| <input checked="" type="radio"/> Duration: | 428.25 | 428.25 | 428.5 |
| <input type="radio"/> Subjects: | 3426 | 3426 | 3428 |

B.5.3 Choice of Variance for Survival Studies

The maximum amount of Fisher information needed to achieve the desired power is shown in Section B.2 to be

$$I_{\max} = \left[\frac{\eta_1}{\delta_1 - \delta_0} \right]^2 .$$

Equation (B.33) relates the required Fisher information, I_{\max} , to the required number of events, D_{\max} by noting that

$$I_{\max} \approx \text{var}[S(l_K)] = r(1 - r)D_{\max} . \tag{B.97}$$

In the above expression the $\text{var}[S(l_K)]$ has been evaluated under the null hypothesis, leading to the result

$$D_{\max} = \frac{1}{(r)(1-r)} \left[\frac{\eta_1}{\delta_1 - \delta_0} \right]^2, \tag{B.98}$$

An alternative approach would be to estimate of the variance of $S(l_K)$ under the alternative hypothesis. In that case

$$I_{\max} \approx \text{var}[S(l_K)] = (p_K)(1-p_K)D_{\max} \tag{B.99}$$

where p_K is the proportion of of the D_{\max} events that occur in the experimental group. This leads to the result

$$D_{\max} = \frac{1}{(p_K)(1-p_K)} \left[\frac{\eta_1}{\delta_1 - \delta_0} \right]^2, \tag{B.100}$$

Since under local alternatives p_K converges to r the above two expressions for $\text{var}[S(l_K)]$, and hence for D_{\max} , are asymptotically equivalent. In small samples, however, the two ways of computing D_{\max} can lead to different results, especially if the randomization fraction is not equal to 0.5. East provides the user with the option to use either the null variance or the alternative variance for evaluating D_{\max} on the **Design Parameters** tab.

Variance of Log Hazard Ratio

Null
 Alternative

The evaluation of p_K for use in equation (B.100) is an iterative process. For any given a and S_a , we proceed through the following steps:

1. Initialize $p_K = r$
2. Compute

$$D_{\max} = \frac{1}{(p_K)(1-p_K)} \left[\frac{\eta_1}{\delta_1 - \delta_0} \right]^2$$

3. Find the value of l_K such that

$$E_{(1)}(l_K | a, S_a, \lambda, \gamma, m) = D_{\max}$$

B Group Sequential Design in East 6

4. With this value of l_K compute $E(l_K|a, S_a, \lambda_T, \gamma_T, m)$, $E(l_K|a, S_a, \lambda_C, \gamma_C, m)$ and

$$D_{\max}(\text{new}) = E(l_K|a, S_a, \lambda_T, \gamma_T, m) + E(l_K|a, S_a, \lambda_C, \gamma_C, m)$$

5. Compute

$$p_K = \frac{E(l_K|a, S_a, \lambda_T, \gamma_T, m)}{D_{\max}(\text{new})}$$

6. Return to step 2

We iterate steps 2 through 5 until the value of D_{\max} stabilizes.

Note: This refinement for computing D_{\max} is only available for superiority trials. For non-inferiority trials $p(l_K) = r$ under the alternative hypothesis, and equation (B.98) may be used with no modification.

C

Interim Monitoring in East 6

The primary characteristic of the interim monitoring phase in East is flexibility. At the design phase East obtains the stopping boundaries and the maximum statistical information by assuming that the study will be monitored a total of K times, after pre-specified increments of information. Provided the study adheres strictly to its planned schedule of interim monitoring, it is assured of maintaining the desired type-I error and power. It is, however, administratively inconvenient to fix in advance the number and timing of the interim looks. For instance, it might be necessary to set the dates of the interim monitoring looks so as to accommodate the schedule of a data and safety monitoring board (DSMB). Typically, the DSMB would plan to meet after equal increments of calendar time, which would not necessarily coincide with the information fractions specified at the design stage. Again, it might be necessary to alter K , the planned number of looks at the data, either for safety reasons, because the accrual assumptions were not met, or for some other administrative reason. These alterations to the original plan could change the operating characteristics of the study unless suitable adjustments were made in the interim monitoring phase. East makes the necessary adjustments by implementing the error spending function methodology first proposed by Lan and DeMets (1983) for studies that stop early to reject H_0 , and subsequently generalized by Pampallona, Tsiatis and Kim (1995), (2000), to studies that stop early to reject either H_0 or H_1 . This appendix chapter covers all the key components of the interim monitoring module in East. The following topics are discussed:

- Flexibility to alter the number and timing of the interim monitoring time-points through the error spending function methodology while preserving the type-1 and type-2 errors (Section C.1).
- Measuring the impact that deviations from the number and timing of the interim monitoring time-points specified at the design phase have on the post-hoc power of the study (Section C.2).
- Conditional power calculations aimed at assisting in the decision to stop early due to futility (Section C.3).
- Repeated confidence intervals that provide the desired coverage for the primary parameter of interest despite the multiple looks (Section C.5).
- Inference at the end of a group sequential trial (Section C.6).
- Sequential monitoring from any general data generating process, not necessarily the normal, binomial or time to failure models that are supported directly by East (Section C.7).
- The ability to monitor on a dimensionless information scale and thereby facilitate sample size recalculation (Section C.8).

C Interim Monitoring in East 6

C.1 Flexible Interim Monitoring

C.1.1 Monitoring with Alpha Functions

C.1.2 Monitoring with Alpha and Beta Functions

The boundary and maximum information computations at the design phase were performed under the assumption that the number and spacing of the interim looks are known in advance. In practice this assumption is unrealistic. A major goal of a practical interim monitoring strategy is to give the user flexibility to monitor the data at arbitrary time points at the interim monitoring stage, possibly perform one or more unplanned analyses, possibly drop one or more planned analyses, and still preserve the type-1 error of the study design. This flexibility is achieved through the spending function approach as originally introduced by Lan and DeMets (1983). If the boundaries at the design stage were themselves derived from spending functions (as discussed in Section B.2.4), one simply uses the same spending functions to re-compute the boundaries at any arbitrary interim monitoring time point. If, however the boundaries constructed at the design stage belong to the Wang-Tsiatis family (Section B.2.2) or the Pampallona-Tsiatis family (Section B.2.3) they are re-computed by inverting special ten-look error spending functions that capture the spirit of these boundaries. (The construction of these ten-look error spending functions is described in detail in Appendix F.)

C.1.1 Monitoring with α -Spending Functions

Suppose the clinical trial was designed for early stopping to reject H_0 . Let $\alpha(t)$ denote its α -spending function. Suppose that the study was originally planned for up to K looks at the accumulating data, at the interim monitoring fractions t_1, t_2, \dots, t_K . Stopping boundaries c_1, c_2, \dots, c_K have already been generated on this basis using the methods discussed in Section B.2.4. If the study is monitored strictly according to plan these same stopping boundaries may be used to make early stopping decisions. If, however, one deviates from the plan, the original stopping boundaries are no longer valid and new boundaries have to be computed on the fly at each interim monitoring time point to reflect the amount of type-1 error that has actually been spent.

Suppose, for example, that the first time we monitor the data, the information fraction is $t'_1 \neq t_1$. We then re-compute the first boundary value c'_1 such that, under the null hypothesis of no treatment difference (H_0),

$$P_0(W(t'_1) \geq c'_1) = \alpha(t'_1).$$

If we do not stop the study at the first interim test, then the data are monitored a second time. Suppose the second monitoring takes place at information fraction $t'_2 \neq t_2$. At this stage, we are allowed to use up a total of $\alpha(t'_2)$ of the significance level. Since we already used $\alpha(t'_1)$ at the first look, we then compute the next boundary value c'_2 so that

$$\alpha(t'_1) + P_0(W(t'_1) < c'_1, W(t'_2) \geq c'_2) = \alpha(t'_2).$$

This guarantees that the probability of stopping and rejecting at the first or second monitoring, under H_0 , will be $\alpha(t'_j)$. In general we compute the boundary c'_j at

information fraction $t'_j \leq 1$ by solving equation

$$\alpha(t'_{j-1}) + P_0(W(t'_1) < c'_1, \dots, W(t'_{j-1}) < c'_{j-1}, W(t'_j) \geq c'_j) = \alpha(t'_j). \quad (C.1)$$

If it should happen that I_j , the information accrued by look j , exceeds I_{\max} , the maximum information stipulated at the design stage, so that $t'_j = I_j/I_{\max} > 1$, East will set $\alpha(t'_j) = \alpha$ and force the j th look to be the last one. Thus, since $\alpha(t'_j) \leq \alpha$ for any j and any information fraction t'_j , this procedure guarantees that the probability under H_0 of ever crossing the upper boundary can never exceed α . Therefore this flexible interim monitoring procedure always preserves the overall type-1 error.

We should note that the α -spending procedure does not guarantee that the type-2 error will be preserved. However, Lan and DeMets (1983) have shown that these procedures, even with few monitoring times, will yield statistical properties similar to those expected with continual monitoring. The operating characteristics and early stopping properties of sequential tests would not be very different whether you monitored the data 5 times, 10 times, or continually. For this reason, once the spending function is specified, we are free to either monitor the accumulating data continually, monitor after equal increments of calendar time, monitor after equal information fractions, or monitor sporadically, without any significant change in the type-2 error. The post-hoc power chart displayed by East (see Section C.2) shows that as long as a study reaches its accrual goals its power is affected minimally even if the interim monitoring schedule differs from what was planned at the design stage.

C.1.2 Monitoring with α - and β -Spending Functions

The spending function approach of Lan and DeMets (1983) was developed in the context of designs that do not allow for early stopping with rejection of the alternative hypothesis. Rejection of the alternative hypothesis could only occur at the last look. Thus, in the initial approach of Lan and DeMets (1983), whereas the type-1 error was spent in accordance with a spending function $\alpha(t)$, the type-2 error had probability exactly equal to zero at all looks except the last, where it had the desired probability, β . However, when sequential designs are constructed in terms of an upper boundary for early rejection of the null hypothesis and a lower boundary for early rejection of the alternative hypothesis, then the total probability of the type-2 error, β , can also be distributed over successive looks. The rate at which the error probability is to be spent can be described by an appropriate strictly increasing function of the information time. Let $\beta(t)$ denote this function such that $\beta(0) = 0$ and $\beta(1) = \beta$. The design of trials that spend α and β simultaneously and stop early to reject either H_0 or H_1 has already been described in Section B.2.4. Suppose we have designed such a trial for up to K monitoring time points at the information fractions t_1, t_2, \dots, t_K . For the one-sided test, let l_j and u_j be the values of the lower and upper boundaries, respectively, at the j^{th} look, $j = 1, 2, \dots, K$.

C Interim Monitoring in East 6

Now suppose we are about to monitor the trial and no longer wish to adhere to either the number or timing of the interim looks specified at the design stage. Pampallona, Tsiatis and Kim (1995) have suggested the following adaptation of the Lan and DeMets (1983) procedure for flexible interim monitoring while simultaneously preserving both the type-1 error and type-2 errors of the study. Suppose that we monitor the data for the first time at information fraction $t'_1 \neq t_1$. Then we would compute the first pair of boundary values, (l'_1, u'_1) , so as to satisfy

$$P_0(W(t'_1) \geq u'_1) = \alpha(t'_1)$$

and

$$P_\eta(W(t'_1) \leq l'_1) = \beta(t'_1)$$

where η , the drift parameter, has been computed at the design stage along with the upper and lower stopping boundaries as described in Section B.2.4. Similarly the boundary values, (l'_j, u'_j) , at subsequent looks, $j \geq 2$, will have to satisfy

$$\alpha(t'_{j-1}) + P_0(l'_1 < W(t'_1) < u'_1, \dots, l'_{j-1} < W(t'_{j-1}) < u'_{j-1}, W(t'_j) \geq u'_j) = \alpha(t'_j) \quad (\text{C.2})$$

and

$$\beta(t'_{j-1}) + P_\eta(l'_1 < W(t'_1) < u'_1, \dots, l'_{j-1} < W(t'_{j-1}) < u'_{j-1}, W(t'_j) \leq l'_j) = \beta(t'_j). \quad (\text{C.3})$$

If it should happen at some look, j^* say, that $I_{j^*} > I_{\max}$, so that $t'_{j^*} = I_{j^*}/I_{\max} > 1$, East will set $\alpha(t'_{j^*}) = \alpha$ and force the j^* th look to be the last one. The upper boundary, u_{j^*} , will then be computed as the solution to

$$\alpha(t'_{j-1}) + P_0(l'_1 < W(t'_1) < u'_1, \dots, l'_{j-1} < W(t'_{j-1}) < u'_{j-1}, W(t'_{j^*}) \geq u'_{j^*}) = \alpha. \quad (\text{C.4})$$

Since we require the stopping boundaries to meet at the last look, it will not be necessary to compute l_{j^*} , the lower boundary at the last look. Instead we will simply set $l_{j^*} = u_{j^*}$. In that case the probability of crossing the lower boundary at the last look or earlier is evaluated by computing

$$\beta^* = \beta(t'_{j-1}) + P_\eta(l'_1 < W(t'_1) < u'_1, \dots, l'_{j^*-1} < W(t'_{j^*-1}) < u'_{j^*-1}, W(t'_{j^*}) \leq u'_{j^*}). \quad (\text{C.5})$$

Since the right hand sides of equations (C.2) and (C.4) can never exceed α this procedure guarantees that the probability under H_0 of ever crossing the upper

boundary can never exceed α . Therefore this flexible interim monitoring procedure always preserves the overall type-1 error. In its present form, however, this procedure is not guaranteed to preserve the type-2 error because β^* , evaluated by equation (C.5), could in principle exceed β . In order to ensure that the type-2 error is always preserved we need to position the last look in such a way that $\beta^* \leq \beta$. The optimal positioning of the last look, to be discussed in Section C.2.3, will ensure this.

C.2 Post-Hoc Power and Preservation of Error

C.2.1 Last-Look Boundary

C.2.2 Computation

C.2.3 Optimal Last Look

C.2.4 Post-Hoc Power Chart

In Section C.1 we developed the error spending function methodology for preserving the type-1 error, despite deviations from the number and timing of the interim looks specified at the design phase of the study. While the type-1 error is indeed preserved by this methodology, it is possible for the alterations in the interim monitoring schedule to affect the type-2 error (hence the power) of the study. Thus it is helpful to compute the post-hoc power at the end of the study, taking into account the actual number and timing of the interim looks. For instance, we would not be too concerned about the impact of the alterations in the interim monitoring schedule if the study was designed for 90% power and the post-hoc power turned out to be 89.5%. This section shows how such post-hoc calculations can be performed. As a by-product we generate a power chart in which, under the assumption that the next look will be the last one, the relationship of post-hoc power to the final statistical information is plotted. The optimal placement of the last look (on the statistical information scale) so as to achieve the power specified at the design phase, is thus obtained. This provides us with a strategy for preserving power by altering the information horizon. Although all the calculations in this section are derived for one-sided tests, they can be readily extended to the two-sided setting by replacing $W(t_j)$ with $|W(t_j)|$.

Note that the post-hoc power calculations in this section differ from the conditional power calculations in Section C.3. Post-hoc power calculations utilize the placement on the information scale of the interim looks already taken, while conditional power calculations utilize, in addition, the current value of the test statistic. Also, the post-hoc power chart is plotted as a function of statistical information whereas the conditional power chart is plotted as a function of the standardized treatment difference.

C.2.1 Boundary Derivation if the Next Look is the Last

Suppose a study has been active for a while, accruing information without the test statistic crossing the stopping boundary at any of the interim monitoring time-points. Eventually, however, the decision must be taken to make the next analysis the last one regardless of the value of the test statistic. As a practical matter it is very unlikely that this last analysis can be performed at the precise time-point that the planned maximum information is attained. In some cases the actual information will exceed the planned maximum and in other cases it will fall short. Some studies may even need to be

C Interim Monitoring in East 6

closed prematurely for administrative reasons, like poor accrual or withdrawal of the drugs under investigation. In all such cases the information fraction $t_L \neq 1$, where L indexes the last analysis. This situation brings up two issues:

1. The boundary for the last look should be computed by spending the balance of the type-I error probability, namely $\alpha - \alpha(t_{L-1})$, in order for the group sequential test to have the desired size α .
2. The power of the adopted sequential procedure usually won't equal the desired, $1 - \beta$, due to the probable departure of the sequence of analyses actually performed from the analyses assumed at the design stage.

For designs allowing for early stopping only to reject the null hypothesis we compute u_L , the boundary for the L -th look, by satisfying the following equation (here given for one-sided tests):

$$\alpha(t_{L-1}) + P_0(W(t_1) < u_1, W(t_2) < u_2, \dots, W(t_{L-1}) < u_{L-1}, W(t_L) \geq u_L) = \alpha. \quad (\text{C.6})$$

For designs allowing for early stopping in favor of either the null or the alternative the last-look upper stopping boundary, u_L , (which must equal the last-look lower stopping boundary, l_L) is obtained by satisfying the following equation (here given for one-sided tests):

$$\alpha(t_{L-1}) + P_0(l_1 < W(t_1) < u_1), \dots, l_{L-1} < W(t_{L-1}) < u_{L-1}, W(t_L) \geq u_L) = \alpha. \quad (\text{C.7})$$

In either case, however, the achieved overall power of the procedure probably won't be what was specified at design time because of deviations from the planned number and timing of the interim analyses. Therefore East computes "post-hoc power" to quantify the power actually achieved by the adopted analysis strategy. This is discussed next.

C.2.2 Calculating Post-Hoc Power

As stated previously, it is highly unlikely that the actual number and timings of the interim analyses will match the K equally spaced analyses assumed at the design stage, and this discrepancy affects the power of the sequential testing procedure. It might be of interest to know what the real power of the study was, based on the actual interim monitoring time-points rather than the assumed ones, even though we can only perform this power calculation post-hoc. If the post-hoc power is reasonably close to the planned power despite deviations in the interim monitoring schedule, one can feel satisfied that the study preserved its original operating characteristics.

If the study is designed for a one-sided test with early stopping to only reject H_0 , East

computes the post-hoc power (PHP) from the following equation.

$$\text{PHP} = 1 - P_{\eta}[W(t_1) < u_1, W(t_2) < u_2, \dots, W(t_{L-1}) < u_{L-1}, W(t_L) < u_L] \quad (\text{C.8})$$

where u_L is the boundary used at the last look to satisfy the type-I error probability, as specified by equation (C.6). Similarly, in the case of a one-sided test allowing for early stopping to reject either H_0 or H_1 , the post-hoc power becomes

$$\text{PHP} = 1 - \left[P_{\eta}[W(t_1) < l_1] + P_{\eta}[l_1 < W(t_1) < u_1, W(t_2) < l_2] + \dots \right] \quad (\text{C.9})$$

where $u_L = l_L$ is the boundary used at the last look to satisfy the type-I error probability, as computed by equation (C.7).

C.2.3 Optimal Placement of Last Look

Suppose that, in the course of the interim monitoring, it was decided to make the next look the last one, regardless of the current interim monitoring time-point. Where should that last look be positioned? To answer this question consider that we designed the sequential test for a type-I error of α and a power of $1 - \beta$. The discussion in Section C.2.1, ensures that the overall type-I error will indeed be α no matter where we position the last look. The deviations from the planned schedule of interim monitoring imply, however, that if we take the last look at the time point specified in the original design, the power of test may no longer be $1 - \beta$.

Pampallona, Tsiatis and Kim (1995) have proposed the following strategy in order to match as closely as possible the desired power, $1 - \beta$. Suppose we have completed look j at information fraction $t_j < 1$ and have not yet crossed a stopping boundary. Let the next look be the last one and suppose that it will be taken at information fraction t_{L^*} , selected in such a way that the power of the test will be $1 - \beta$. For a one-sided test allowing for early stopping to either reject H_0 or H_1 , we jointly solve the following equations for $u_L^* = l_L^*$ and $t_L^* > t_j$, the latter being referred to as the optimal last look position:

$$\alpha(t_j) + P_0[l_1 < W(t_1) < u_1, \dots, l_j < W(t_j) < u_j, W(t_L^*) \geq u_L^*] = \alpha$$

$$\beta(t_j) + P_{\eta}[l_1 < W(t_1) < u_1, \dots, l_j < W(t_j) < u_j, W(t_L^*) \leq u_L^*] = \beta.$$

For a one-sided test allowing for early stopping to only reject H_0 the entire type-2 error can only be spent at the last look. In that case we jointly solve the following

C Interim Monitoring in East 6

equations for $u_L^* = l_L^*$ and $t_L^* > t_j$:

$$\alpha(t_j) + P_0[W(t_1) < u_1, W(t_2) < u_2, \dots, W(t_j) < u_j, W(t_L^*) \geq u_L^*] = \alpha$$

$$P_\eta[W(t_L^*) < u_L^*] = \beta$$

These equations provide the information fraction, t_L^* , and the boundary value, u_L^* , for the next analysis, assuming it to be the last, such that the type-I error and the power are **both** preserved under the adopted schedule of analyses.

Let t_L be the actual information fraction at which the next and last analysis occurs. Then, the position of t_L^* is optimal in the sense that $t_L < t_L^*$ would entail a loss of power, while $t_L > t_L^*$ would make the study unnecessarily overpowered, while only $t_L = t_L^*$ would match the desired power exactly. East computes t_L^* before and after every interim analysis, converts it into units of relevance to the current application (for example, total number of patients or total number of events) and displays this quantity in the box labeled “Ideal Next Look Position”. In the course of the study, this information can guide the investigator to position the next look optimally.

It should be noted that t_1^* corresponds to the information fraction required for a study without interim monitoring (fixed sample size study) relative to the group sequential study under consideration. That is, given that no analyses have yet been performed, the optimum position of the last (and in this case also the first) look, would be the one corresponding to the fixed sample size. East displays this value when the Interim Monitoring module is entered for the first time. If the actual first analysis is performed at $t_1 < t_1^*$ and the stopping boundary has not been crossed then clearly $t_2^* > t_1^*$ and the process continues. In this context it should be pointed out that since the error spending functions are defined only for $t_j \leq 1$, any analysis performed at $t_j > 1$ must necessarily be the last. East is capable of detecting this situation, and will accordingly compute the boundaries, spend the balance of the type I error, and display the post-hoc power.

C.2.4 Post-Hoc Power Chart

We have seen in Section C.2.3 that East is able to adjust the maximum information through the optimal last look methodology so as to satisfy the desired power and significance level of the design, despite departures from the chosen number of equally spaced analyses specified at the design stage. It may however be of interest to know what loss or gain in power would derive from the last analysis being performed at a

time point different from the suggested optimal last look position. The post-hoc power chart answers this question by providing a graph of the post-hoc power (on the Y-axis) versus the total information accumulated by the last look (on the X-axis). The point on the X-axis that matches the optimal last look position will correspond to full power on the Y-axis. Information is expressed in the post-hoc power chart in terms of units of relevance to the outcome being considered (e.g., patient accrual for normally distributed data or events for time to failure data). Towards the end of the study the post-hoc power chart tends to flatten out so that relatively small increases in power occur for relatively large increases in information. The post-hoc power chart is updated after each look and allows the user to decide whether the adjustment to the maximum information suggested by East is worth accepting, should the next look be the last. The chart is not displayed after a stopping boundary is crossed.

C.3 Conditional Power at Ideal Next Look Position (East 5.4)

The concept of conditional power at ideal next look position is borrowed from the setting of fixed sample size studies. It was first proposed in this setting by Lan and Wittes (1988). If the test statistic is computed when only part of the required total information has been collected, then the conditional power quantifies the probability of rejecting the null hypothesis should the total information be eventually available, conditional on the current information. Such a probability, when computed over a range of alternatives, can be of guidance in deciding whether to continue the study given the available evidence. In East this idea is extended to group sequential studies.

Let us initially consider a one-sided group sequential test of size α , designed for early rejection of H_0 . Suppose at the j th analysis the information fraction is t_j and the test statistic, $W(t)$, has value $w(t_j)$. In the notation of Section C.2.3 let t_L^* be the optimal placement of the next look, assuming it to be the last one, with corresponding boundary value u_L^* . In East we define the conditional power at ideal next look position, CP at INLP, as the following probability :

$$\text{CP at INLP} = P_\eta[W(t_L^*) \geq u_L^* \mid w(t_j)] \quad (\text{C.10})$$

Recall from Section B.1 that the statistic $W(t_j)$ is defined as a sum of independent increments. This implies that the decomposition

$$W(t_L^*) = W(t_j) + [W(t_L^*) - W(t_j)]$$

has the following three properties:

1. The random variables $W(t_j)$ and $W(t_L^*) - W(t_j)$ are normal and independent.
2. The means of these random variables are $E[W(t_j)] = \eta t_j$ and $E[W(t_L^*) - W(t_j)] = \eta(t_L^* - t_j)$.

C Interim Monitoring in East 6

3. The variances of these random variables are $\text{Var}[W(t_j)] = t_j$ and $\text{Var}[W(t_L^*) - W(t_j)] = (t_L^* - t_j)$.

Once we have reached the information fraction t_j we know that the random variable $W(t_j)$ has assumed the value $w(t_j)$. Therefore

$$\begin{aligned} \text{CP at INLP} &= P_\eta[W(t_L^*) - w(t_j) \geq u_L^* - w(t_j)] \\ &= P_\eta \left[\frac{W(t_L^*) - w(t_j) - \eta(t_L^* - t_j)}{\sqrt{t_L^* - t_j}} \geq \frac{u_L^* - w(t_j) - \eta(t_L^* - t_j)}{\sqrt{t_L^* - t_j}} \right] \\ &= 1 - \Phi \left(\frac{u_L^* - w(t_j) - \eta(t_L^* - t_j)}{\sqrt{t_L^* - t_j}} \right). \end{aligned} \tag{C.11}$$

where $\Phi(x)$ is the cumulative distribution function for a standard normal random variable.

For two-sided tests the conditional power is expressed as follows:

$$\begin{aligned} \text{CP at INLP} &= P_\eta[|W(t_L^*)| \geq u_L^* \mid w(t_j)] \\ &= 1 - \Phi \left(\frac{u_L^* - w(t_j) - \eta(t_L^* - t_j)}{\sqrt{t_L^* - t_j}} \right) \\ &\quad + \Phi \left(\frac{-u_L^* + w(t_j) - \eta(t_L^* - t_j)}{\sqrt{t_L^* - t_j}} \right). \end{aligned} \tag{C.12}$$

Analogous expressions can be derived for designs with boundaries for early rejection of either H_0 or H_1 .

In East the conditional power is presented as a graph plotted against a wide range of alternatives for δ , including the one specified at the design stage. Now equations (C.11) and (C.12) express conditional power as a function of the drift parameter η rather than as a function of δ . At the design stage, the relationship between η and δ is captured by the equation

$$\eta = (\delta - \delta_0) \sqrt{I_{\max}} \tag{C.13}$$

introduced in Section B.1 of Appendix B.

Finally, it should be noted that, given the described approach, the conditional power curve computed before the very first look is the usual power curve. In particular, at that stage the optimal placement of the next and last look corresponds to a fixed sample size study so that under the alternative specified at design the conditional power is actually equivalent to the a priori unconditional power.

C.4 Conditional and predictive power (East 6)

The concept of conditional power is borrowed from the setting of fixed sample size studies. It was first proposed in this setting by Lan and Wittes (1988). If the test statistic is computed when only part of the required total information has been collected, then the conditional power quantifies the probability of rejecting the null hypothesis should the total information be eventually available, conditional on the current information. Such a probability, when computed over a range of alternatives, can be of guidance in deciding whether to continue the study given the available evidence. In East this idea is extended to group sequential studies.

Suppose at the j th analysis the information fraction is t_j and the test statistic, $W(t)$, has value $w(t_j)$. We define the conditional power at look j as the probability of attaining statistical significance in the direction of the alternative hypothesis at any future look, given $w(t_j)$. Thus, if we are testing the null hypothesis that $\delta = \delta_0$ against the alternative that $\delta > \delta_0$, the conditional power is defined as

$$CP_{\eta}(w(t_j)) = Pr_{\eta}(\cup_{k=j+1}^K W(t_k) > u_k | w(t_j))$$

Here $\eta = (\delta - \delta_0) \sqrt{I_{max}}$ is the trend parameter under the alternative hypothesis. If the alternative hypothesis is that $\delta < \delta_0$, then the conditional power is defined as

$$CP_{\eta}(w(t_j)) = Pr_{\eta}(\cup_{k=j+1}^K W(t_k) < l_k | w(t_j))$$

Analogous expressions can be written for designs with boundaries for early rejection of either H_0 or H_1 and designs for early rejections of the two sided tests. The corresponding probabilities are obtained by the recursive integration.

The reference values of the conditional power are often based on the design or estimated value of the trend parameter η

$$\left[\begin{array}{l} \eta_d = (\delta_d - \delta_0) \sqrt{I_{max}} \\ \hat{\eta}_j = \frac{w(t_j)}{t_j} \end{array} \right]$$

The predictive power $PP(w(t_j))$ provides a weighted average of the conditional power values for a range of values of η

$$PP(w(t_j)) = \int CP_{\eta}(w(t_j)) f(\eta) d\eta$$

C Interim Monitoring in East 6

We follow suggestion of Lan, Hu and Proschan (2009) and use the weighting function

$$f(\eta) = \phi\left(\mu = \hat{\eta}_j, \sigma^2 = \frac{1}{t_j}\right)$$

where ϕ denotes the probability density function of standard normal distribution.

C.5 Repeated Confidence Intervals

C.5.1 RCI's Derived from Boundaries that Reject H_0

C.5.2 RCI's for Boundaries that Reject either H_0 or H_1

C.5.3 East Inputs

In this section we discuss the computation of repeated confidence intervals (RCI's), each interval being computed as part of an interim analysis. These RCI's were first proposed by Jennison and Turnbull (1989) and are discussed in detail in Chapter 9 of their text book (Jennison and Turnbull, 2000). The naive confidence interval one would ordinarily compute from the data gathered at the end of a clinical trial is inappropriate if the confidence interval is computed repeatedly in a group sequential setting. In this setting the naive confidence interval will fail to provide the desired coverage for the parameter of interest due to the problem of multiplicity. In contrast the RCI's provide simultaneous coverage for the parameter of interest at any desired confidence level despite the multiple looks at the data.

C.5.1 RCI's Derived from Boundaries that Reject H_0

For ease of exposition let us consider RCI's for two-sided group sequential trials of efficacy endpoints. The extension to one-sided efficacy or non-inferiority trials is straightforward. Let the primary parameter of interest be δ and suppose we perform K interim analyses at the information fractions t_1, t_2, \dots, t_K . At information fraction t_j we compute the Wald statistic

$$Z(t_j) = \frac{\hat{\delta}(t_j)}{\text{se}(\hat{\delta}(t_j))}. \tag{C.14}$$

Recognizing that in large samples $[\text{se}(\hat{\delta}(t_j))]^{-2} \approx I_j$, where I_j is the information about δ at time t_j , we may also write the Wald statistic as

$$Z(t_j) = \hat{\delta}(t_j)\sqrt{I_j}. \tag{C.15}$$

By the Scharfstein, Tsiatis and Robins (1997) theorem introduced in Section B.1 of Appendix B, $Z(t_j) \sim N(\delta\sqrt{I_j}, 1)$ and $\text{cov}[Z(t_{j_1}), Z(t_{j_2})] = \sqrt{I_{j_1}/I_{j_2}}$.

Let b_1, b_2, \dots, b_K be any two-sided level- α stopping boundaries for the Wald statistic for testing the null hypothesis that $\delta = 0$. That is,

$$P_0\left\{\bigcap_{j=1}^K |Z(t_j)| < b_j\right\} = 1 - \alpha.$$

Now observe that $(Z(t_j) - \delta\sqrt{I_j}) \sim N(0, 1)$ and has the same covariance structure as $Z(t_j)$. Therefore

$$P_\delta\left\{\bigcap_{j=1}^K |Z(t_j) - \delta\sqrt{I_j}| < b_j\right\} = 1 - \alpha \quad (\text{C.16})$$

for any value of δ .

Let H_1, H_2, \dots, H_K denote K two-sided RCI's that maintain simultaneous coverage for δ at level $1 - \alpha$. Therefore we require these confidence intervals to satisfy the probability condition

$$P_\delta\left\{\bigcap_{j=1}^K \delta \in H_j\right\} = 1 - \alpha. \quad (\text{C.17})$$

We can show that the sequence of intervals

$$H_j = \hat{\delta}(t_j) \pm \text{se}(\hat{\delta}(t_j))b_j \text{ for } j = 1, 2, \dots, K, \quad (\text{C.18})$$

satisfy the simultaneous coverage requirement (C.17). To prove this assertion observe that

$$\begin{aligned} P_\delta\left\{\bigcap_{j=1}^K \delta \in H_j\right\} &= P_\delta\left\{\bigcap_{j=1}^K \hat{\delta}(t_j) - \text{se}(\hat{\delta}(t_j))b_j < \delta < \hat{\delta}(t_j) + \text{se}(\hat{\delta}(t_j))b_j\right\} \\ &= P_\delta\left\{\bigcap_{j=1}^K \hat{\delta}(t_j)\sqrt{I_j} - b_j < \delta\sqrt{I_j} < \hat{\delta}(t_j)\sqrt{I_j} + b_j\right\} \\ &= P_\delta\left\{\bigcap_{j=1}^K |\hat{\delta}(t_j) - \delta|\sqrt{I_j} < b_j\right\} \\ &= P_\delta\left\{|Z(t_j) - \delta\sqrt{I_j}| < b_j\right\} \\ &= 1 - \alpha \text{ (by equation (C.16))}. \end{aligned}$$

C.5.2 RCI's for Boundaries that Reject either H_0 or H_1

Consider a K -look, level- α one sided group sequential test of the null hypothesis $H_0: \delta = 0$ having $1 - \beta$ power to detect the alternative hypothesis $H_1: \delta = \delta_1 > 0$. Suppose the interim monitoring takes place at the information fractions t_1, t_2, \dots, t_K .

C Interim Monitoring in East 6

Let $(l_j, u_j), j = 1, 2, \dots, K$ be the futility and efficacy boundaries, respectively, for this test. These boundaries have been derived in Section B.2.4 of Appendix B. Since these boundaries preserve the type-1 error we must have

$$P_0\left\{\bigcap_{j=1}^K Z(t_j) < u_j\right\} = 1 - \alpha . \tag{C.19}$$

Therefore, following the argument made in the previous section,

$$P_\delta\left\{\bigcap_{j=1}^K Z(t_j) - \delta\sqrt{I_j} < u_j\right\} = 1 - \alpha . \tag{C.20}$$

Now the event $Z(t_j) - \delta\sqrt{I_j} < u_j$ if and only if $\delta > \hat{\delta}(t_j) - u_j\text{se}(\hat{\delta}(t_j))$. Thus sequence $\{\hat{\delta}(t_j) - u_j\text{se}(\hat{\delta}(t_j)): j = 1, 2, \dots, K\}$ simultaneously excludes δ from below with probability $1 - \alpha$. It follows that the probability that one or more of these lower confidence bounds fails to cover δ from below is at most α .

Next consider the behaviour of the Wald statistics under $H_1: \delta = \delta_1$. Since the lower stopping boundaries were constructed from a β spending function we must have

$$P_{\delta_1}\left\{\bigcap_{j=1}^K Z(t_j) > l_j\right\} = 1 - \beta . \tag{C.21}$$

Therefore by centralizing the Wald statistic we get

$$P_\delta\{Z(t_j) - \delta\sqrt{I_j} + \delta_1\sqrt{I_j} > l_j\} = 1 - \beta . \tag{C.22}$$

From this we can easily show that the sequence $\{\hat{\delta}(t_j) - \delta_1 - l_j\text{se}(\hat{\delta}(t_j)): j = 1, 2, \dots, K\}$ simultaneously excludes δ from above with probability $1 - \beta$. It follows that the probability that one or more of these upper confidence bounds fails to cover δ from above is at most β . Thus the sequence of intervals

$$\{[\hat{\delta}(t_j) - u_j\text{se}(\hat{\delta}(t_j)), \hat{\delta}(t_j) + \delta_1 - l_j\text{se}(\hat{\delta}(t_j))]: j = 1, 2, \dots, K\} \tag{C.23}$$

simultaneously contains the true value of δ with probability $1 - \alpha - \beta$.

C.5.3 Inputs to East for RCI Computation

Equation (C.18) shows that in order to compute the RCI's East needs to know both the numerator ($\hat{\delta}(t_j)$) and denominator ($\text{se}(\hat{\delta}(t_j))$) of the Wald statistic (C.14). East provides a **Test Statistic Calculator** for entering these two components separately into the appropriate cell of the interim monitoring worksheet. For example, in the case of interim monitoring of a normal design study, if you click the button

Enter Interim Data

on the IM dashboard, the following dialog box appears into which you may enter the observed values for $\hat{\delta}(t_j)$ and $(se(\hat{\delta}(t_j)))$.

The dialog box titled "Test Statistic Calculator" contains the following fields and controls:

- Editing Look #1
 - Set Current Look as Last
- Cumulative Sample Size: 158
- Input for Normal end point: 0.3
- Estimate of δ : $\delta = (\mu_t - \mu_c)$: 0.159
- Standard Error of Estimate of δ : 0.159
- Output
 - Test Statistic: 1.885
- Buttons: Recalc, OK, Cancel

Sometimes, however, the separate components of the Wald statistic may not be known. Then the user has no choice but to directly enter the observed value of the Wald statistic $Z(t_j)$ into the Interim Monitoring worksheet. In such cases, East suppresses the output of repeated confidence intervals, conditional power estimates and the final adjusted inference estimates from the interim monitoring worksheet.

C Interim Monitoring in East 6

C.6 Inference Following Group Sequential Testing

C.6.1 Stage-Wise Ordering

C.6.2 Adjusted P-values

C.6.3 Adjusted Confidence Interval

C.6.4 Point Estimation

C.6.5 Acceptance Boundaries

C.6.6 Drift Parameter and Effect Size

In this section we discuss the computation of p-values and confidence intervals for the parameter of interest at the end of a group sequential clinical trial. The naive approach of computing these quantities in the usual way, ignoring the fact that a sequential monitoring procedure was used to possibly stop early, will fail to preserve the desired type-I error of the significance test or the desired coverage of the confidence interval. Rather, one must first order the sample space to reflect the sequential nature of the test procedure, and then obtain p-values and confidence intervals on the basis of this ordering. Jennison and Turnbull (2000, Chapter 8) discuss four ways to order the sample space of a group sequential experiment and thereby perform an adjusted inference in which the type-I error and the coverage are both preserved. These four ways are, stage-wise ordering, MLE ordering, likelihood ratio ordering, and score test ordering. In East we adopt stage-wise ordering of the sample space. This ordering was first proposed by Armitage (1957) and later used by Fairbanks and Madsen (1982), Tsiatis, Rosner and Mehta (1984), and Kim and DeMets (1987). Of all the four orderings this is the one most favored by Jennison and Turnbull (2000) because it does not require knowledge about the interim monitoring time-points that would have been adopted in the future, had the study not stopped early. The other three orderings of the sample space do require this knowledge and are therefore limited in their practical applicability. In addition, stage-wise ordering ensures consistency between the p-value and the confidence interval. That is, a $100 \times (1 - \alpha)\%$ confidence interval will exclude the parameter value under the null hypothesis if and only if the corresponding p-value does not exceed α . Finally, the p-value based on stage-wise ordering is less than the significance level, α , if and only if H_0 is rejected.

C.6.1 Stage-Wise Ordering of the Sample Space

Suppose that the sequentially computed random variable $W(t) \sim N(\eta t, t)$ crosses a stopping boundary for the first time at the j th look in a group sequential clinical trial where the current information fraction is t_j and the current value of the test statistic is $w^*(t_j)$. Let the information fractions at the earlier looks be $\{t_1, t_2, \dots, t_{j-1}\}$ with corresponding lower and upper stopping boundaries given by (l_i, u_i) , $i = 1, 2, \dots, j - 1$. Define the i th continuation region as $C_i = (l_i, u_i)$. The l_i 's might each be $-\infty$, in which case we have a one-sided sequential test with early stopping to reject H_0 . On the other hand, if $l_i = -u_i$ for all i , we have a two-sided sequential test with early stopping to reject H_0 . More generally the (l_i, u_i) pairs could represent the lower and upper stopping boundaries, respectively, of the Pampallona and Tsiatis (1994) family for a one-sided sequential test with early stopping to reject either H_0 or H_1 . The most complex case, inner-wedge stopping boundaries to reject either H_0 or H_1 with a two-sided test, is not covered in this section but is discussed in Section C.6.5.

The sample space of a sequential experiment which was terminated at the j th interim look with an observed value of $w^*(t_j)$ for the test statistic, consists of the union over all $i = 1, 2, \dots, j$ of all possible trajectories that terminate at the i th look. These trajectories are of the form

$$(t_1, w(t_1)) \rightarrow (t_2, w(t_2)) \rightarrow \dots \rightarrow (t_i, w(t_i))$$

where $w(t_i) \notin \mathcal{C}_i$ but $w(t_g) \in \mathcal{C}_g$, for all $g = 1, 2, \dots, i - 1$. The idea behind stage-wise ordering of this sample space is to associate earlier stopping with larger values of η . Accordingly, in stage-wise ordering, the ordered pair $(t_a, w(t_a))$ is more extreme than the ordered pair $(t_b, w(t_b))$ whenever any one of the following four conditions holds:

- (i) $w(t_a) \geq u_a$ and $w(t_b) \leq l_b$ for $t_a, t_b = 1, 2, \dots, j - 1$,
- (ii) $w(t_a) > w(t_b)$ if $t_a = t_b$ for $t_a, t_b = 1, 2, \dots, j$,
- (iii) $t_a < t_b$ if $w(t_a) \geq u_a$ and $w(t_b) \geq u_b$ for $t_a, t_b = 1, 2, \dots, j - 1$,
- (iv) $t_a > t_b$ if $w(t_a) \leq l_a$ and $w(t_b) \leq l_b$ for $t_a, t_b = 1, 2, \dots, j - 1$.

Figure C.1 is a visual display of the stage-wise ordering of the sample space for a study with three interim looks. For additional discussion of stage-wise ordering refer to Jennison and Turnbull (2000, page 179).

C.6.2 Adjusted P-values

A p-value is defined as the probability, under the null hypothesis, of obtaining an outcome at least as extreme as the one actually observed. The set of points which are at least as extreme as the observed point, $(t_j, w^*(t_j))$, can be identified by applying the stage-wise ordering scheme to each sample point in accordance with the rules set forth in Section C.6.1. Denote this set by \mathcal{E}^* . Then the p-value, adjusted for the sequential testing, is the probability under the null hypothesis of obtaining the event \mathcal{E}^* . That is,

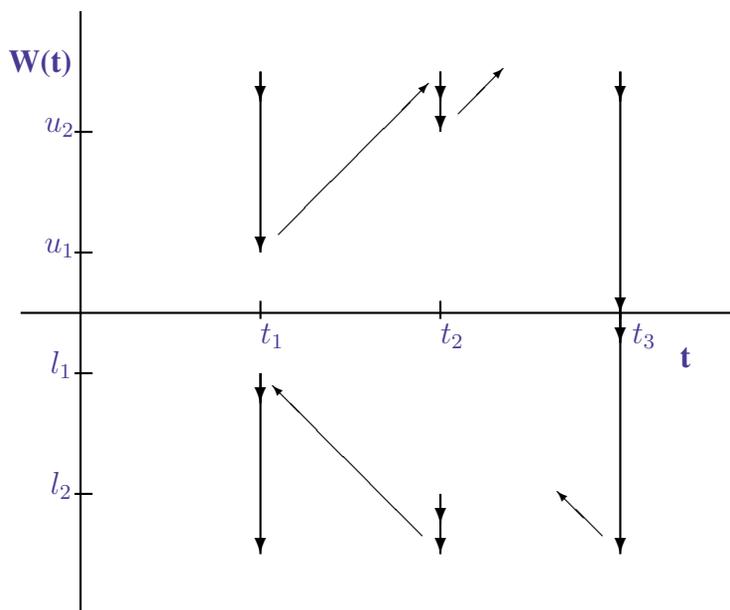
$$p^* = P_0\{\mathcal{E}^*\}. \tag{C.24}$$

C.6.3 Adjusted Confidence Interval

The method applied in East for deriving a confidence interval for η follows the approach proposed by Tsiatis, Rosner and Mehta (1984) and later extended by Kim and DeMets (1987). The basic idea is to search for the upper and lower confidence bounds of η such that the p-value under the alternative hypothesis just becomes statistically significant. Suppose the study was terminated at the observed point $(t_j, w^*(t_j))$ and let \mathcal{E}^* be the set of points at least as extreme as $(t_j, w^*(t_j))$ in accordance with the stage-wise ordering scheme developed in Section C.6.1. Then the

C Interim Monitoring in East 6

Figure C.1: Example of the ordering of the sample space $\{(t_i, W(t_i)); i = 1, 2, 3\}$. (Arrows point from more extreme to less extreme points, where extreme refers to evidence of larger values of the effect size, η .)



$100 \times (1 - 2\nu)$ confidence interval for η is (η^L, η^U) where

$$\eta^U = \sup \{ \eta : P_\eta \{ \mathcal{E}^* \} \leq 1 - \nu \} , \quad (\text{C.25})$$

$$\eta^L = \inf \{ \eta : P_\eta \{ \mathcal{E}^* \} \geq \nu \} . \quad (\text{C.26})$$

C.6.4 Point Estimation

Kim (1989) has proposed the following median unbiased estimator (MUE) for the parameter η . The MUE, denoted by $\tilde{\eta}$ is the value of η that satisfies

$$P_{\tilde{\eta}} \{ \mathcal{E}^* \} = 0.5 . \quad (\text{C.27})$$

C.6.5 Boundaries to Accept H_0

The adjusted p-values, confidence intervals and point estimations discussed in

Sections C.6.2, C.6.3 and C.6.4, respectively, can be extended to the one-sided and two-sided $H_0 - -H_1$ stopping boundaries. We must be careful, however, to exclude from the set \mathcal{E}^* all points that lie within the region where the null hypothesis is accepted. This approach will produce adjusted p-values and confidence intervals with the correct properties so long as the study is not terminated by the test statistic entering the acceptance region. In the latter case, since the null hypothesis is accepted, East will not report a p-value or confidence interval.

C.6.6 Drift Parameter and Effect Size

In Sections C.6.3 and C.6.4 we showed how East computes adjusted confidence intervals and median unbiased point estimates, respectively, for the drift parameter η . These estimates must be transformed into corresponding estimates of the effect size δ in order to be meaningful to the end user. The relationship between η and δ was shown in Section B.1 of Appendix B to be

$$\eta = (\delta - \delta_0) \sqrt{I_{\max}}. \quad (\text{C.28})$$

Thus if we know the value of I_{\max} we can solve the above equation for δ in terms of η . For example,

$$\begin{aligned} \delta^U &= \delta_0 + \frac{\eta^U}{\sqrt{I_{\max}}} \\ \delta^L &= \delta_0 + \frac{\eta^L}{\sqrt{I_{\max}}}. \end{aligned}$$

For each specific application (e.g. normal, binomial or time to failure data) we have derived, in Section B.1 of Appendix B, an expression for I_{\max} in terms of n_{\max} and other parameters specified at the design stage. These relationships are used to transform the confidence interval for η into a corresponding confidence interval for δ . However, these relationships usually contain nuisance parameters that must be estimated from the current data. For example, we would use equation (B.15) to compute I_{\max} for the normal case and would therefore need to estimate σ^2 from the data. We would use equation (B.20) to compute I_{\max} for binomial superiority trials and equation (B.23) to compute I_{\max} for binomial non-inferiority trials. In either case we would need to estimate π_c , the control response rate, from the current data.

In East we use the following unified method to evaluate the maximum information, I_{\max} . Suppose we have just completed the j th interim analysis. Let I_j denote the current information and $t_j = I_j/I_{\max}$ denote the current information fraction. Then we can re-write I_{\max} as

$$I_{\max} = t_j^{-1} I_j = [\sqrt{t_j} \text{se}(\hat{\delta}_j)]^{-2}. \quad (\text{C.29})$$

C Interim Monitoring in East 6

Thus as long as we provide East with the current standard error estimate, $se(\hat{\delta}_j)$, East can estimate I_{\max} from equation (C.29). The value of $se(\hat{\delta}_j)$ is passed to East through the test statistic calculator. If this calculator is by-passed in favor of entering the current value of the test statistic directly into the interim monitoring worksheet, East will not produce adjusted p-values, point estimates or confidence intervals upon study termination.

C.7 Monitoring Data from any General Distribution

The interim monitoring of studies that are designed with the General Design module in East is no different than the procedure used for studies designed by the Normal, Binomial or Survival Design modules. The user supplies East with the maximum information, I_1 needed for a single look study that is designed to investigate some parameter of interest, say δ . Usually I_1 will be translated into a sample size N_1 before it is input to the General Design worksheet of East. Sometimes I_1 will be expressed in terms of the number of events needed for a single look study. The Poisson example in Chapter 60 is one such case. It is also permissible to retain I_1 in terms of Fisher information for a single look study, and to approximate it by $[se(\hat{\delta}(\tau))]^{-2}$. This case is, however, better handled by the **I** module, discussed in Section C.8. Once the value of I_1 is provided to East, it is inflated to I_K by the appropriate K -look inflation factor, as discussed in Section B.3 of Appendix B. Thereafter the inflated information is utilized to determine the information fraction at each interim look in the interim monitoring phase of the study and the entire machinery of flexible monitoring with error spending functions is made available to the study. The study stops when the Wald statistic, given by equation (B.1) crosses a stopping boundary.

This approach is very useful in all those situations that are not currently covered by a specialized module within East. For instance, from any commercial sample size package one might obtain the fixed sample size requirements for the comparison of survival of two groups by means of a stratified log-rank test (expressed in terms of a fixed number of events) or for the comparison of two groups in terms of repeated measures (expressed in terms of a fixed number of subjects). East can then compute the corresponding information needed for group-sequential monitoring.

C.8 Information Based Monitoring

Suppose that the observations are generated from some probability model and a single parameter, δ , from this model characterizes the relationship under investigation while the remainder of the model is characterized by nuisance parameters. Interest focuses on developing a sequential procedure, with possible early stopping, for testing the null hypothesis $H_0 : \delta = \delta_0$ against the alternative $H_1 : \delta = \delta_1$. Suppose the study has been designed for a total of K interim looks. Then the maximum information,

$I_K \equiv I_{\max}$, to be committed up-front is given by equation (B.83) as

$$I_K = \left[\frac{z_\alpha + z_\beta}{\delta_1 - \delta_0} \right]^2 \times \text{IF}(\alpha, \beta, K, \text{boundaries}). \quad (\text{C.30})$$

This information is approximated by $[\text{se}(\hat{\delta}(\tau_K))]^{-2}$, where τ_K is the calendar time at the last look. The information at any intermediate look, taken at calendar time τ_j , is likewise approximated by $[\text{se}(\hat{\delta}(\tau_j))]^{-2}$. Suppose that the interim monitoring takes place at calendar times $\tau_1, \tau_2, \dots, \tau_K$. Then the sequential monitoring procedure at calendar time τ_j requires us to compute the information fraction

$$t_j = \frac{[\text{se}(\hat{\delta}(\tau_j))]^{-2}}{[\text{se}(\hat{\delta}(\tau_K))]^{-2}},$$

read off the values $\alpha(t_j)$ and $\beta(t_j)$ from the appropriate error spending functions, and re-compute the stopping boundaries based on these values, in the manner described in Section C.1 of this appendix. The study is terminated if the Wald statistic

$$Z(t_j) = \frac{\hat{\delta}(t_j) - \delta_0}{\sqrt{\text{var}[\hat{\delta}(\tau_j)]}}$$

crosses a stopping boundary.

The big advantage of monitoring on the above information scale is that the total information, I_K , required in order for the study to achieve the desired $1 - \beta$ power, only depends on $\delta_1 - \delta_0$, the specific parameters of interest under H_0 and H_1 . No nuisance parameters are involved in the computation of maximum information. In contrast, if we were to monitor the study on the scale of a physical resource like sample size or number of events, the maximum information would depend on one or more nuisance parameters. If those nuisance parameters were guessed incorrectly, the study would not have the power it was intended to have at the design phase. This will become much clearer as you work through the example of sample size re-estimation provided in Chapter 59.

D Computing the Expected Number of Events

D.1 General expressions

We consider a single arm of a survival study and derive an expression for the expected number of events $d(l)$ to be observed at the calendar time l . A delay between the calendar time when a subject experiences an event or drops out of a study and the calendar time when this information becomes available to an investigator is assumed to be negligible. Our equations may be viewed as a slight generalization of the expressions presented in Kim and Tsiatis (1990).

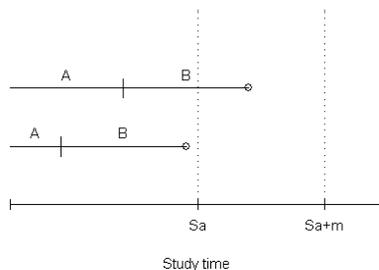


Figure D.1: Geometry of a problem

We are interested in a following general setting:

- A subject is followed no longer than a maximum period of time m . An observation of the event of interest or the subject's drop out from the study terminates a follow-up process.
- An accrual rate $a(u)$, $0 \leq u \leq S_a$ is not uniform.
- The event hazard rate $\lambda(t)$ and the drop-out hazard rate $\gamma(t)$ depend on the subject's follow-up time $t = l - u$.

An important special case arises when a limitation on the maximum follow-up time is removed. It corresponds to $m = \infty$. The accrual rate is often considered known at the time of the design of a study. It may also be calculated based on the known total number of subjects in the study and the known proportion of subjects recruited during the interval $(u, u + du)$.

Figure D.1 illustrates a geometry of the problem. The horizontal axis denotes a calendar time. An accrual period ends at S_a . The follow-up of a subject accrued at $l = S_a$ is completed no later than $l = S_a + m$. Each of the two horizontal lines connects the beginning of an accrual period with a calendar time l which may be positioned within (the lower line) or after (the upper line) an accrual period. At a calendar time l the subjects with the accrual time $0 \leq u \leq l - m$ (group A) are no longer followed because their follow-up windows are closed. Subjects who were accrued later (group B) are continued to be observed unless their follow-up was terminated by the event of interest or a drop out. The value of interest $d(l)$ may be presented as a sum of contributions

$$d(l) = d_A(l) + d_B(l) \tag{D.1}$$

from these groups. We note that in the absence of the restriction on a follow-up time ($m = \infty$) the group A does not exist and the corresponding contribution in (D.1) disappears.

Let us denote by v a time from randomization to an event of interest and by w a time from randomization to the time of subject's drop-out from the study. We assume that random variables v and w are independent and express their probability density functions $f(v)$ and $g(w)$ through the event $\lambda(t)$ and drop-out $\gamma(t)$ hazard functions

$$f(v) = \lambda(v)e^{-\Lambda(v)}, \quad \Lambda(v) = \int_0^v \lambda(t)dt$$

$$g(w) = \gamma(w)e^{-H(w)}, \quad H(w) = \int_0^w \gamma(t)dt,$$

Let us denote by $\Psi(t) = P(v \leq t, w > v)$ a probability that event occurred before the follow-up t and was not censored. We note that

$$\Psi(t) = \int_0^t \left(\int_v^\infty g(w)dw \right) f(v)dv = \int_0^t \kappa(t')dt'$$

with

$$\kappa(t) = \lambda(t) e^{-[\Lambda(t)+H(t)]}$$

D Computing the Expected Number of Events

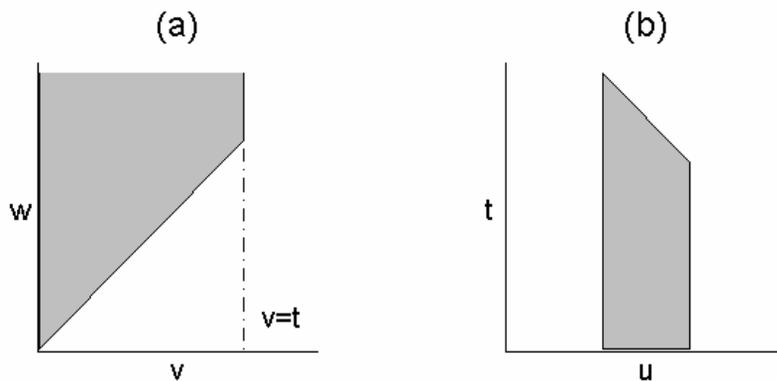


Figure D.2: Geometry of integration

Figure D.2 (a) illustrates a geometry of the integration. A shaded area marks the area of integration in the (v, u) plane.

In a calculation of $d(l)$ we make a distinction between the following cases

Table 1 Special cases.

| Case | l and l^* | $d_A(l)$ | $d_B(l)$ |
|------|--|-------------------------------|------------------------------------|
| 1 | $0 \leq l \leq S_a$ $l^* < 0$ | 0 | $\int_0^l a(u)\Psi(l-u)du$ |
| 2 | $0 \leq l \leq S_a$ $0 \leq l^* \leq S_a - m$ | $\Psi(m) \int_0^{l^*} a(u)du$ | $\int_{l^*}^l a(u)\Psi(l-u)du$ |
| 3 | $S_a < l \leq S_a + m$ $l^* < 0$ | 0 | $\int_0^{S_a} a(u)\Psi(l-u)du$ |
| 4 | $S_a < l \leq S_a + m$ $0 \leq l^* \leq S_a$ | $\Psi(m) \int_0^{l^*} a(u)du$ | $\int_{l^*}^{S_a} a(u)\Psi(l-u)du$ |
| 5 | $S_a + m < l$ $S_a < l^*$ | $\Psi(m) \int_0^{S_a} a(u)du$ | 0 |

We approximate $a(u)$ by a piece-wise constant function splitting an accrual interval

into $i = 1, \dots, n_a$ subintervals with the boundaries $[u_{i-1}, u_i)$ and denoting a constant accrual rate within an interval i by a_i . If the calendar time l^* is located within the interval i^* then

$$\int_0^{l^*} a(u)du = \sum_{i=1}^{i^*-1} a_i (u_i - u_{i-1}) + a_{i^*} (l^* - u_{i^*-1})$$

If $l^* = S_a = u_{n_a}$ then we get

$$\int_0^{S_a} a(u)du = \sum_{i=1}^{n_a} a_i (u_i - u_{i-1})$$

An integral $\int_{u_1}^{u_2} a(u)\Psi(l-u)du$ where u_1 belongs to an interval i_1 and u_2 belongs to an interval i_2 may be written as

$$\begin{aligned} \int_{u_1}^{u_2} a(u)\Psi(l-u)du = & a_{i_1} \varphi(u_1, u_{i_1}, l) + \sum_{i=i_1+1}^{i_2-1} a_i \varphi(u_{i-1}, u_i, l) + \\ & + a_{i_2} \varphi(u_{i_2-1}, u_2, l) \end{aligned}$$

with

$$\varphi(u_{min}, u_{max}, l) = \int_{u_{min}}^{u_{max}} \Psi(l-u)du = \int_{u_{min}}^{u_{max}} du \int_0^{l-u} \kappa(t)dt$$

and $l \geq u_{max}$. An integration region of the two-dimensional integral is shown as a shaded area on Figure D.2 (b). A more convenient expression for $\varphi(u_{min}, u_{max}, l)$ is obtained by changing the order of integration

$$\begin{aligned} \varphi(u_{min}, u_{max}, l) &= \int_0^{l-u_{max}} dt \int_{u_{min}}^{u_{max}} \kappa(t)du + \int_{l-u_{max}}^{l-u_{min}} dt \int_{u_{min}}^{l-t} \kappa(t)du \\ &= (u_{max} - u_{min}) \int_0^{l-u_{max}} \kappa(t)dt \\ &+ \int_{l-u_{max}}^{l-u_{min}} (l - u_{min} - t)\kappa(t)dt \end{aligned} \tag{D.2}$$

The integrals can be calculated numerically for an arbitrary hazard functions $\lambda(t)$ and $\gamma(t)$. In a special case of piece-wise constant hazard functions a calculation of the integrals $\int_a^b \kappa(t)dt$ and $\int_a^b \kappa(t)tdt$ in equation (D.2) is simplified. An integral over the interval (a, b) is presented as a sum of the integrals over the intervals $[t_{j-1}, t_j)$, $j = 1, \dots, J$ where both hazard and drop out rates λ_j and γ_j are constant. These integrals are calculated analytically

$$\int_a^b \kappa(t)dt = \sum_{j=1}^J I_{0j} \tag{D.3}$$

D Computing the Expected Number of Events

where

$$\begin{aligned} I_{0j} &= \int_{t_{j-1}}^{t_j} \kappa(t) dt = \lambda_j e^{-[\Lambda(t_{j-1})+H(t_{j-1})]} \int_{t_{j-1}}^{t_j} e^{-\lambda_{s,j}(t-t_{j-1})} dt \\ &= c_j \left[1 - e^{-\lambda_{s,j}(t_j-t_{j-1})} \right] \end{aligned}$$

with $\lambda_{s,j} = \lambda_j + \gamma_j$ and

$$c_j = \frac{\lambda_j}{\lambda_{s,j}} e^{-[\Lambda(t_{j-1})+H(t_{j-1})]} \tag{D.4}$$

Similarly

$$\int_a^b \kappa(t) t dt = \sum_{j=1}^J I_{1j} \tag{D.5}$$

where

$$\begin{aligned} I_{1j} &= \int_{t_{j-1}}^{t_j} \kappa(t) t dt = \lambda_j e^{-[\Lambda(t_{j-1})+H(t_{j-1})]} \int_{t_{j-1}}^{t_j} e^{\lambda_{s,j}(t-t_{j-1})} t dt \\ &= c_j \left[\left(t_{j-1} + \frac{1}{\lambda_{s,j}} \right) - e^{-\lambda_{s,j}(t_j-t_{j-1})} \left(t_j + \frac{1}{\lambda_{s,j}} \right) \right] \end{aligned} \tag{D.6}$$

In the following sections we present simplified versions of these general expressions that correspond to the more restrictive settings.

D.2 Fixed hazard rate, uniform accrual

- D.2.1 General setting
- D.2.2 No drop out and no fixed follow-up
- D.2.3 Drop out and no fixed follow-up.
- D.2.4 No drop out and fixed follow-up.

D.2.1 General setting

Consider a situation where the event hazard rate is constant ($\lambda(t) = \lambda$), the accrual rate is uniform ($a(t) = a$), and there are the drop outs hazard rate is constant ($\gamma(t) = \gamma$). A subject is followed up to a maximum of $m < \infty$ units of time if an event of interest or drop out does not occur first. The following derivation gives the formula for the expected number of events at calendar time l for all of the cases listed in Table 1.

$0 \leq l \leq S_a, l^* < 0$:

$$d(l) = a \varphi(0, l, l) = a \left[\int_0^l (l-t) \kappa(t) dt \right] = a \left[l \int_0^l \kappa(t) dt - \int_0^l \kappa(t) t dt \right]$$

An application of expressions (D.4) and (D.6) leads to the following results

$$\int_0^l \kappa(t)dt = \frac{\lambda}{\lambda + \gamma} \left[1 - e^{-(\lambda + \gamma)l} \right]$$

$$\int_0^l \kappa(t)t dt = \frac{\lambda}{\lambda + \gamma} \left[\frac{1}{\lambda + \gamma} - e^{-(\lambda + \gamma)l} \left(l + \frac{1}{\lambda + \gamma} \right) \right]$$

Therefore

$$d(l) = \frac{a\lambda}{\lambda + \gamma} \left[\left(l - \frac{1}{\lambda + \gamma} \right) + \frac{e^{-(\lambda + \gamma)l}}{\lambda + \gamma} \right] \tag{D.7}$$

$0 \leq l \leq S_a, 0 \leq l^* \leq S_a - m:$

$$d_A(l) = a(l - m) \int_0^m \kappa(t)dt = \frac{a\lambda(l - m)}{\lambda + \gamma} \left(1 - e^{-(\lambda + \gamma)m} \right)$$

$$\begin{aligned} d_B(l) &= a \varphi(l^*, l, l) = a \int_0^{l-l^*} (l - l^* - t) \kappa(t) dt = \\ &= a \left[m \int_0^m \kappa(t) dt - \int_0^m \kappa(t)t dt \right] \end{aligned}$$

An application of equations (D.4) – (D.6) leads to the following expressions

$$\int_0^m \kappa(t)dt = \frac{\lambda}{\lambda + \gamma} \left(1 - e^{-(\lambda + \gamma)m} \right)$$

$$\int_0^m \kappa(t)t dt = \frac{\lambda}{\lambda + \gamma} \left[\frac{1}{\lambda + \gamma} - e^{-(\lambda + \gamma)m} \left(m + \frac{1}{\lambda + \gamma} \right) \right]$$

Therefore

$$d_B(l) = \frac{a\lambda}{\lambda + \gamma} \left[\left(m - \frac{1}{\lambda + \gamma} \right) + \frac{e^{-(\lambda + \gamma)m}}{\lambda + \gamma} \right]$$

The resulting expression is

$$d(l) = \frac{a\lambda}{\lambda + \gamma} \left[\left(l - \frac{1}{\lambda + \gamma} \right) - e^{-(\lambda + \gamma)m} \left(l - m - \frac{1}{\lambda + \gamma} \right) \right] \tag{D.8}$$

D Computing the Expected Number of Events

$$S_a < l \leq S_a + m, l^* < 0 :$$

$$\begin{aligned} d(l) &= a \varphi(0, S_a, l) = a \left[S_a \int_0^{l-S_a} \kappa(t) dt + \int_{l-S_a}^l (l-t) \kappa(t) dt \right] \\ &= a \left[S_a \int_0^{l-S_a} \kappa(t) dt + l \int_{l-S_a}^l \kappa(t) dt - \int_{l-S_a}^l \kappa(t) t dt \right] \end{aligned}$$

An application of equations (D.4) – (D.6) leads to the following expressions

$$\int_0^{l-S_a} \kappa(t) dt = \frac{\lambda}{\lambda + \gamma} \left(1 - e^{-(\lambda+\gamma)(l-S_a)} \right)$$

$$\int_{l-S_a}^l \kappa(t) dt = \frac{\lambda}{\lambda + \gamma} e^{-(\lambda+\gamma)(l-S_a)} \left(1 - e^{-(\lambda+\gamma)S_a} \right)$$

$$\begin{aligned} \int_{l-S_a}^l \kappa(t) t dt &= \frac{\lambda}{\lambda + \gamma} e^{-(\lambda+\gamma)(l-S_a)} \\ &\quad \left[\left(l - S_a + \frac{1}{\lambda + \gamma} \right) - e^{-(\lambda+\gamma)S_a} \left(l + \frac{1}{\lambda + \gamma} \right) \right] \end{aligned}$$

and the resulting expression for $d(l)$ has the following form

$$d(l) = \frac{a\lambda}{\lambda + \gamma} \left[S_a - \frac{e^{-(\lambda+\gamma)l}}{\lambda + \gamma} \left(e^{(\lambda+\gamma)S_a} - 1 \right) \right] \quad (D.9)$$

$$S_a < l \leq S_a + m, 0 \leq l^* \leq S_a :$$

$$d_A(l) = \Psi(m) \int_0^{l^*} a(u) du = \left[\frac{\lambda}{\lambda + \gamma} \left(1 - e^{-(\lambda+\gamma)m} \right) \right] a(l - m)$$

$$\begin{aligned} d_B(l) &= a \varphi(l^*, S_a, l) = a \left[(S_a - l^*) \int_0^{l-S_a} \kappa(t) dt + \int_{l-S_a}^{l-l^*} (l-l^*-t) \kappa(t) dt \right] \\ &= a \left[(S_a - l^*) \int_0^{l-S_a} \kappa(t) dt + m \int_{l-S_a}^m \kappa(t) dt - \int_{l-S_a}^m \kappa(t) t dt \right] \end{aligned}$$

An application of equations (D.4) – (D.6) leads to the following expressions

$$\int_0^{l-S_a} \kappa(t)dt = \frac{\lambda}{\lambda + \gamma} \left(1 - e^{-(\lambda+\gamma)(l-S_a)}\right)$$

$$\int_{l-S_a}^m \kappa(t)dt = \frac{\lambda}{\lambda + \gamma} e^{-(\lambda+\gamma)(l-S_a)} \left(1 - e^{-(\lambda+\gamma)(m+S_a-l)}\right)$$

$$\int_{l-S_a}^m \kappa(t)tdt = \frac{\lambda}{\lambda + \gamma} e^{-(\lambda+\gamma)(l-S_a)} \left[\left(l - S_a + \frac{1}{\lambda + \gamma}\right) - e^{-(\lambda+\gamma)(m+S_a-l)} \left(m + \frac{1}{\lambda + \gamma}\right) \right]$$

Therefore

$$d_B(l) = \frac{a\lambda}{\lambda + \gamma} \left[(S_a + m - l) - \frac{e^{-(\lambda+\gamma)l}}{\lambda + \gamma} \left(e^{(\lambda+\gamma)S_a} - e^{(\lambda+\gamma)(l-m)} \right) \right]$$

and

$$d(l) = \frac{a\lambda}{\lambda + \gamma} \left[S_a - (l - m)e^{-(\lambda+\gamma)m} - \frac{e^{-(\lambda+\gamma)l}}{\lambda + \gamma} \left(e^{(\lambda+\gamma)S_a} - e^{(\lambda+\gamma)(l-m)} \right) \right] \tag{D.10}$$

$S_a + m < l, S_a < l^*$:

Events that occur at the calendar time l exceeding $S_a + m$ are not observed because the maximum follow-up time m is limited. The expression for $d(l)$ has the following form

$$d(l) = a S_a \Psi(m) = \frac{a S_a \lambda}{\lambda + \gamma} \left(1 - e^{-(\lambda+\gamma)m}\right) \tag{D.11}$$

D.2.2 No drop out and no fixed follow-up

In this situation the event hazard rate is constant ($\lambda(t) = \lambda$), the accrual rate is uniform ($a(u) = a$), there are no drop outs ($\gamma(t) = 0$), and subjects are followed up until the end of study ($m = \infty$). In the unlimited follow-up time setting $l^* = l - m$ is always negative and only the cases 1 and 3 from the Table 1 are to be considered. The

D Computing the Expected Number of Events

following expressions for $d(l)$ are obtained from equations (D.7, D.9) by a substitution of $\gamma = 0$.

$$0 < l \leq S_a, l^* < 0:$$

$$d(l) = a \left[\left(l - \frac{1}{\lambda} \right) + \frac{e^{-\lambda l}}{\lambda} \right]$$

$$S_a < l, l^* < 0:$$

$$d(l) = a \left[S_a - \frac{e^{-\lambda l}}{\lambda} (e^{\lambda S_a} - 1) \right]$$

D.2.3 Drop out and no fixed follow-up.

In this situation we consider the event hazard rate is constant ($\lambda(t) = \lambda$), the accrual rate is uniform ($a = a$), the drop out hazard rate $\gamma(t) = \gamma$ is non-zero, and subjects are followed up until the end of study ($m = \infty$). Once again, the cases 1 and 3 from Table 1 are to be considered and the expressions (D.7, D.9) are directly applicable.

$$0 < l \leq S_a, l^* < 0:$$

$$d(l) = \frac{a\lambda}{\lambda + \gamma} \left[\left(l - \frac{1}{\lambda + \gamma} \right) + \frac{e^{-(\lambda + \gamma)l}}{\lambda + \gamma} \right]$$

$$S_a < l, l^* < 0:$$

$$d(l) = \frac{a\lambda}{\lambda + \gamma} \left[S_a - \frac{e^{-(\lambda + \gamma)l}}{\lambda + \gamma} (e^{(\lambda + \gamma)S_a} - 1) \right]$$

D.2.4 No drop out and fixed follow-up.

Now consider a situation where the event hazard rate is constant ($\lambda(t) = \lambda$), the accrual rate is uniform ($a(t) = a$), and there are no drop outs ($\gamma(t) = 0$). However, each subject is now followed up to a maximum of $m < \infty$ units of time if an event of

interest or drop out does not occur first. The following expressions are obtained from equations (D.7–D.11) by a substitution of $\gamma = 0$.

$0 \leq l \leq S_a, l^* < 0$:

$$d(l) = a \left[\left(l - \frac{1}{\lambda} \right) + \frac{e^{-\lambda l}}{\lambda} \right]$$

$0 \leq l \leq S_a, 0 \leq l^* \leq S_a - m$:

$$d(l) = a \left[\left(l - \frac{1}{\lambda} \right) - e^{-\lambda m} \left(l - m - \frac{1}{\lambda} \right) \right]$$

$S_a < l \leq S_a + m, l^* < 0$:

$$d(l) = a \left[S_a - \frac{e^{-\lambda l}}{\lambda} (e^{\lambda S_a} - 1) \right]$$

$S_a < l \leq S_a + m, 0 \leq l^* \leq S_a$:

$$d(l) = a \left[S_a - (l - m)e^{-\lambda m} - \frac{e^{-\lambda l}}{\lambda} (e^{\lambda S_a} - e^{\lambda(l-m)}) \right]$$

$S_a + m < l, S_a < l^*$:

$$d(l) = a S_a (1 - e^{-\lambda m})$$

D.3 Piecewise constant hazard and drop out rates, no follow-up limit

Consider a setting where an accrual is uniform ($a(u) = a$) and hazard and drop-out rates are piece-wise constant so that $\lambda(t) = \lambda_k$ and $\gamma(t) = \gamma_k$ for $[\tau_{k-1} \leq t < \tau_k)$. We also assume that there is no follow-up limit ($m = \infty$). For the unlimited follow-up time m the value $l^* = l - m$ is always negative and therefore only the cases 1 and 3 from the Table 1 are to be considered.

$0 < l \leq S_a, l^* < 0$:

$$d(l) = a \varphi(0, l, l) = \int_0^l (l - t)\kappa(t)dt = l \int_0^l \kappa(t)tdt - \int_0^l \kappa(t)tdt$$

We denote by k^* the number of the interval $[\tau_{k^*-1}, \tau_{k^*})$ which contains l . The integrals $\int_0^l \kappa(t)tdt$ and $\int_0^l \kappa(t)tdt$ are calculated using the expressions (D.3) – (D.6) with $a = 0, b = l, J = k^*, t_j = \tau_j$ for $j = 0, \dots, J - 1$ and $t_J = l$.

D Computing the Expected Number of Events

$S_a < l, l^* < 0$:

$$\begin{aligned} d(l) &= a \varphi(0, S_a, l) = S_a \int_0^{l-S_a} \kappa(t) dt + \int_{l-S_a}^l (l-t) \kappa(t) dt \\ &= S_a \int_0^{l-S_a} \kappa(t) dt + l \int_{l-S_a}^l \kappa(t) dt - \int_{l-S_a}^l \kappa(t) t dt \end{aligned}$$

We denote by k^* the number of the interval $[\tau_{k^*-1}, \tau_{k^*})$ which contains S_a and by k' the number of the interval $\tau_{k'-1}, \tau_{k'}$ which contains $l - S_a$. The calculation of the integrals $\int_0^{S_a} \kappa(t) dt$, $\int_{l-S_a}^l \kappa(t) dt$ and $\int_{l-S_a}^l \kappa(t) t dt$ is based on the expressions (D.3) – (D.6). In the calculation of an integral over the interval $(0, S_a)$ we use $a = 0, b = S_a, J = k^*, t_j = \tau_j$ for $j = 0, \dots, J-1$ and $t_J = S_a$. The corresponding values used in the calculation of the integrals over the interval $(l - S_a, l)$ are $a = l - S_a, b = l, J = k^* - k' + 1, t_0 = l - S_a, t_j = \tau_{k'-1+j}, j = 1, \dots, J-1$ and $t_J = l$.

D.4 Non-uniform accrual, constant hazard and drop out rates

If the setting where an accrual $a(u)$ is not uniform but hazard and drop-out rates are constant the following simplified expressions for the integrals in the expression D.2 are available

$$\int_0^{l-u_{max}} \kappa(t) dt = \frac{\lambda}{\lambda_s} \left[1 - e^{-\lambda_s(l-u_{max})} \right]$$

$$\int_{l-u_{max}}^{l-u_{min}} \kappa(t) dt = \frac{\lambda}{\lambda_s} e^{-\lambda_s(l-u_{max})} \left[1 - e^{-\lambda_s(u_{max}-u_{min})} \right]$$

$$\begin{aligned} \int_{l-u_{max}}^{l-u_{min}} \kappa(t) t dt &= \frac{\lambda}{\lambda_s} e^{-\lambda_s(l-u_{max})} \\ &\left[l - u_{max} + \frac{1}{\lambda_s} - e^{-\lambda_s(u_{max}-u_{min})} \left(l - u_{min} + \frac{1}{\lambda_s} \right) \right] \end{aligned}$$

and $\lambda_s = \lambda + \gamma$.

E Generating Survival Simulations in EastSurv

East provides the user with powerful simulation tools for trials with time-to-event endpoints. In addition to easily verifying the operating characteristics of the many different design scenarios mentioned in Appendix B, the simulations may be used to actually design for non-standard problems where power and sample size calculations are analytically intractable. For instance, East allows the user to simulate trials in which the hazard rates for each treatment arm are non-proportional. By trial and error, running simulations under various parameter choices, the user may find an appropriate design for this kind of trial. East actually provides two simulation methods: a) Basic simulation and b) Enhanced simulation.

The Basic simulation method uses asymptotic theory when generating the data and is discussed in the main East manual. In East 3.1, the enhanced simulation also used asymptotic theory to generate the data, but allowed the user to change some of the design parameter values in order to simulate under various scenarios. EastSurv's enhanced simulation tool no longer generates the data using asymptotic theory. The purpose of this appendix in fact is to outline how the data are generated in the new enhanced survival simulations.

When initiating an enhanced survival simulation session, East uses as input all the parameters selected during the design stage. By clicking on the "Show Survival Parameters" button, a survival sheet is opened that allows the user to change these parameter values. In fact, the flexibility offered to the user in this screen is such that the piecewise exponential hazard curves in each treatment arm can be individually specified. This permits the user to specify late separating hazard curves or even crossing hazard curves. In addition, the user must also decide how each simulated trial will terminate by choosing whether to a) fix the number of events in the trial or b) fix the study duration. Once this is done, clicking either the "Run" button or "Single Step" button starts the simulations. East then proceeds as follows.

In each simulation:

1. For each accrual period
 - (a) East computes the number of subjects to be accrued in the control group and the treatment group.
 - (b) For each subject i
 - i. A random accrual time $t_{acc,i}$ of subject i is generated as a random value from the uniform distribution bounded by the starting and ending times of the current accrual period

E Generating Survival Simulations in EastSurv

- ii. A random survival time $t_{surv,i}$ is generated as a random value from the survival time distribution characterized by a piecewise hazard rate
- iii. A random dropout time $t_{drop,i}$ is generated as a random value from the exponential distribution characterized by the dropout rate.
- iv. An indicator of censoring C_i is computed as

$$\begin{cases} C_i = 0 & \text{if } t_{surv,i} \geq t_{drop,i} \text{ and } t_{surv,i} \geq t_{fix} \\ C_i = 1 & \text{otherwise} \end{cases}$$

where t_{fix} is the user-specified fixed maximum follow-up time of a subject.

2. Now for each look j
 - (a) If the timing of the look j is characterized by the time S_j since the initiation of the study then the value of S_j is predefined.
 - (b) If the timing of the look j is driven by the number of events so that the look j occurs immediately after observing N_j events then S_j is calculated based on the study times $t_{study,i} = t_{acc,i} + t_{surv,i}$ of the uncensored observations (with $C_i = 0$).
 - (c) At the time S_j the subset of observations of interest is limited to the observations from accrued subjects ($t_{acc,i} \leq S_j$). For the look-based analysis the observations with $t_{study,i} > S_j$ are treated as censored.
 - (d) The calculated values of S_j or N_j are stored for the subsequent calculation of average values across the simulations.
 - (e) East computes the test statistic and checks if a stopping boundary has been crossed.
 - i. If yes, or if the last look has been reached without crossing a stopping boundary, it proceeds to the next simulation.
 - ii. Otherwise, it proceeds to the next look.

F Spending Functions Derived from Power Boundaries

East provides several families of published spending functions, each with a well defined functional form. These spending functions are all documented in Section B.2.4 of Appendix B. The general approach is to select one of these published spending functions for generating the stopping boundaries at the design stage and to select the same spending function to re-compute the stopping boundaries at the interim monitoring stage. This gives us the flexibility to change the number and spacing of the interim looks during the interim monitoring stage.

However, the Wang-Tsiatis (1987) and Pampallona-Tsiatis (1994) power boundaries are not derived from spending functions. If these boundaries is used for the study design they should also be used for interim monitoring. This could be problematic if the number and spacing of the interim looks changes from what was specified at the design stage. For this reason we construct special “ten-look” spending functions that correspond to the members of the Wang-Tsiatis or Pampallona-Tsiatis family. The next section shows how this is accomplished.

F.1 Inverting Ten-Look Power Boundaries

For each Wang-Tsiatis power boundary of the form

$$C(\Delta, \alpha, K)t_j^\Delta, \quad j = 1, 2, \dots, K,$$

we compute the type-1 errors, as they accumulate at each of the equally spaced looks, t_1, t_2, \dots, t_K , according to the selected values of Δ and α , but with a preset value for the maximum number of looks, $K = 10$. For example, suppose we wish to generate a spending function that corresponds to a one-sided Wang-Tsiatis power boundary for a specific value of α and Δ . The first step is to compute the actual boundary values at the ten equally spaced looks t_1, t_2, \dots, t_{10} , where $t_j = j/10$, using the procedure described in Section B.2.2 of Appendix B. Denote these ten boundary values by c_1, c_2, \dots, c_{10} . Next, compute the cumulative errors $\alpha(t_j)$, $j = 1, 2, \dots, 10$, where

$$\alpha(t_1) = P_0[W(t_1) \geq c_1],$$

and for $j = 2, 3, \dots, 10$,

$$\alpha(t_j) = \alpha(t_{j-1}) + P_0[W(t_1) < c_1, \dots, W(t_{j-1}) < c_{j-1}, W(t_j) \geq c_j].$$

These computations are clearly unaffected by the type of end point since the test statistic can be expressed in the general framework of Section B.1. Linear interpolation between these cumulative errors is then applied for setting up approximate spending functions for the type-1 and type-2 error probabilities to be used at the interim monitoring stage. This approach will make the resulting re-computed boundaries at the

F Spending Functions Derived from Power Boundaries

interim monitoring stage enjoy approximately the same properties as the corresponding original boundaries obtained at the design stage while still providing flexibility to deviate from the pre-specified number and timing of the interim looks. However, as a consequence of fixing $K = 10$ for deriving the spending function in the interim monitoring module, even though we might have used a different value of K in the design module, there can be slight differences in the boundary values computed at the design stage and the boundary values computed at the interim monitoring stage. In practice this difference is negligible, as we show below in Section F.2..

F.2 Comparison of Design Boundaries and Interim Monitoring Boundaries

At the design stage East computes the Wang and Tsatis (1987) or Pampallona and Tsatis (1990) power boundaries directly, as documented in Appendix B, Sections B.2.2 and B.2.3. These boundaries depend on K , the number of equally spaced interim looks. At the interim monitoring stage, however, East re-computes the stopping boundaries by inverting a ten-look error spending function, as documented above Section F.1. This implies that, even if the interim monitoring actually takes place as designed at K equally spaced looks, the design boundaries won't match the interim monitoring boundaries, unless $K = 10$. This is not of much practical importance since, as a consequence of the flexible spending function methodology, interim monitoring will rarely occur at precisely the same time-points as was specified in the design. Table F.1 and Table F.2, display the O'Brien-Fleming power boundaries obtained at the design stage, for $K = 5$ and $K = 3$, respectively, and the corresponding boundaries obtained by inverting a ten-look error spending function, for a two-sided test at $\alpha = 0.05$. We observe that the difference between the design and interim monitoring boundaries is very small.

Table F.1: Design and Interim Monitoring Boundaries for Five Equally Spaced Looks

| Look No. | Information Fraction | Design Boundary | Monitoring Boundary |
|----------|----------------------|-----------------|---------------------|
| 1 | 0.2 | ± 4.562 | ± 4.692 |
| 2 | 0.4 | ± 3.226 | ± 3.285 |
| 3 | 0.6 | ± 2.634 | ± 2.656 |
| 4 | 0.8 | ± 2.281 | ± 2.285 |
| 5 | 1.0 | ± 2.040 | ± 2.035 |

Table F.2: Design and Interim Monitoring Boundaries for Three Equally Spaced Looks

| Look No. | Information Fraction | Design Boundary | Monitoring Boundary |
|----------|----------------------|-----------------|---------------------|
| 1 | 0.333 | ±3.471 | ±3.518 |
| 2 | 0.667 | ±2.454 | ±2.487 |
| 3 | 1.000 | ±2.004 | ±1.998 |

F.3 Comparison of Ten-Look and Lan and DeMets Spending Functions

We stated in Section B.2.2 of Appendix B that the power boundaries proposed by Wang and Tsiatis (1987) generate, as a special case, the boundaries of O'Brien and Fleming (1979) if the shape parameter takes on the value $\Delta = 0$. We also stated in Section B.2.4 of Appendix B that the **LD (OF)** spending function (Lan-DeMets spending function with O'Brien-Fleming flavor) of the form

$$\alpha(t) = 4 - 4\Phi\left(\frac{z_{\alpha/4}}{\sqrt{t}}\right) \quad (\text{F.1})$$

generates two-sided boundaries similar to those proposed by O'Brien and Fleming. It is therefore of interest to see how the spending function derived from the ten-look design compares with $\alpha(t)$. The figure below shows that the two spending functions

F Spending Functions Derived from Power Boundaries

have very similar behaviors.

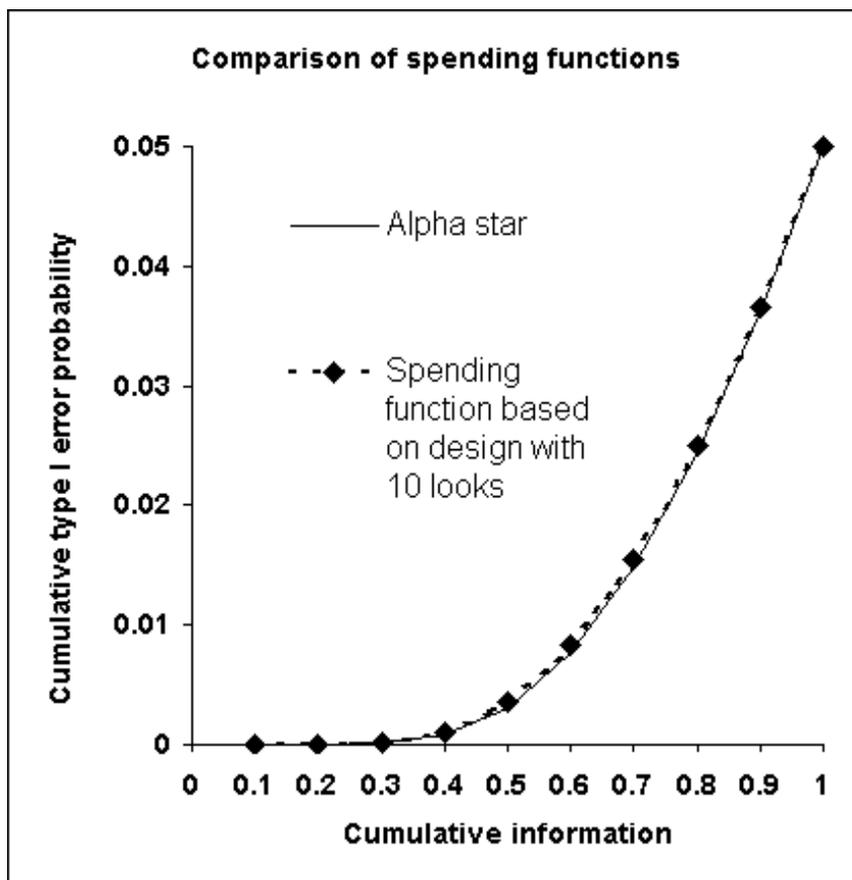


Table F.3 displays the amount of type-I error actually spent, at each of five equally spaced looks by the two error spending functions given an overall type-I error of $\alpha = 0.05$. Corresponding stopping boundaries are also displayed. We note that the differences are very minor. The last column of Table F.3 displays the actual O'Brien and Fleming power boundaries based on shape parameter $\Delta = 0$, and number of looks $K = 5$, using the computations discussed in Appendix B, Section B.2.2. These boundaries too are very similar to the boundaries derived from the two error spending functions.

Table F.3: Comparing 10-Look and Lan-DeMets Spending Functions

| Look No. | Fraction (t) | Error Spent | | Stopping Boundaries | | |
|----------|----------------|-------------|-------------|---------------------|-------------|-------------|
| | | 10-Look | $\alpha(t)$ | 10-Look | $\alpha(t)$ | 5-look |
| 1 | 0.2 | 0.000003 | 0.000001 | ± 4.692 | ± 4.877 | ± 4.562 |
| 2 | 0.4 | 0.001020 | 0.000789 | ± 3.285 | ± 3.357 | ± 3.226 |
| 3 | 0.6 | 0.008262 | 0.007617 | ± 2.656 | ± 2.680 | ± 2.634 |
| 4 | 0.8 | 0.025008 | 0.024424 | ± 2.285 | ± 2.290 | ± 2.281 |
| 5 | 1 | 0.05 0000 | 0.050000 | ± 2.035 | ± 2.031 | ± 2.040 |

G *The Recursive Integration Algorithm*

Substantial savings in computational effort can be achieved in the computations of the group sequential boundaries. We will give details of this savings using one-sided tests of hypothesis with boundary only for the rejection of H_0 . But the same applies to other situations.

At the time of the j^{th} interim monitoring, the group sequential boundary is determined by

$$\Pr_0(W(t_1) < b_1, \dots, W(t_{j-1}) < b_{j-1}, W(t_j) \geq b_j) = \alpha^*(t_j) - \alpha^*(t_{j-1}). \quad (G.1)$$

The probability above is evaluated by the recursive integration formula by Armitage, McPherson and Rowe (1969); the density function for $W(t)$ in the discrete sequential procedure is given by

$$f_1(w; \eta) = t_1^{-1/2} \phi[t_1^{-1/2}(w - \eta t_1)],$$

and, by recursion

$$f_j(w; \eta) = \int_{-\infty}^{b_{j-1}} f_{j-1}(v; \eta) \Delta t_j^{-1/2} \phi[\Delta t_j^{-1/2}(w - v - \eta \Delta t_j)] dv \quad (G.2)$$

where $\Delta t_j = t_j - t_{j-1}$, for $j = 1, \dots, K$, with $t_0 = 0$, and ϕ is the standard normal density function. Equation (G.2) follows from the fact that, as discussed in Section B.1 of Appendix B, the distribution of $W(t_j)$ is $N(\eta t_j, t_j)$ with an independent increments structure..

To find the boundary for the j^{th} interim monitoring, we simply need to find the value of b_j such that

$$\int_{b_j}^{\infty} f_j(w; \eta) dw = \alpha^*(t_j) - \alpha^*(t_{j-1}).$$

Therefore, at each time of interim monitoring, instead of repeating the recursive numerical integration, we need to evaluate the numerical integration only once by storing internally previous boundary values b_1, \dots, b_{j-1} and the coordinates of the density function $f_{j-1}(w; \eta)$ for $-\infty < w < b_{j-1}$.

H Theory - Multiple Comparison Procedures

H.1 Parametric Procedures

H.1.1 Introduction

H.1.2 Single Step Dunnett Test

H.1.3 Step Down Dunnett Test

H.1.1 Introduction

Assume that there are k arms including the placebo arm. Let n_0 be the number of subjects for placebo arm and n_i the number of subjects for i^{th} treatment arm ($i = 1, 2, \dots, k - 1$). Let $N = \sum_{i=0}^{k-1} n_i$ be the total sample size. Let Y_{ij} be the response from subject j in treatment arm i and y_{ij} be the observed value of Y_{ij} ($i = 0, 1, \dots, k - 1, j = 1, 2, \dots, n_i$). Suppose that

$$Y_{ij} = \mu_i + e_{ij} \tag{H.1}$$

where $e_{ij} \sim N(0, \sigma^2)$. Let \bar{y}_i ($i = 0, 1, \dots, k - 1$) be the sample mean for treatment arm i and s^2 be the pooled sample variance for all arms. Let $T_i = \frac{\bar{y}_i - \bar{y}_0}{s \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}}$ be the

test statistic for comparing treatment effect of arm i with placebo. Let $T_{(1)} \geq T_{(2)} \geq \dots \geq T_{(k-1)}$ be the ordered statistics of T_i . Let t_i ($i = 1, \dots, k - 1$) be the observed values of T_i and $t_{(1)} \geq t_{(2)} \geq \dots \geq t_{(k-1)}$ be the observed values of $T_{(1)} \geq T_{(2)} \geq \dots \geq T_{(k-1)}$. We are interested in the following hypotheses

- For the right tailed test: $H_i : \mu_i - \mu_0 \leq 0$ vs $K_i : \mu_i - \mu_0 > 0$
- For the left tailed test: $H_i : \mu_i - \mu_0 \geq 0$ vs $K_i : \mu_i - \mu_0 < 0$
- For the global null hypothesis: $H_0 : \mu_0 = \mu_1 = \mu_2 = \dots = \mu_{k-1}$ vs $H_{01} : \text{At least one } \mu_i > \mu_0$ for right tailed test ($\mu_i < \mu_0$ for left tailed test)

H.1.2 Single Step Dunnett Test in One-Way ANOVA Design

Let $F(x)$ denote the distribution function of $T_{(1)}$ under the global null hypothesis H_0 , i.e.

$$F(x) = \Pr(T_{(1)} \leq x) = \int_0^\infty \int_{-\infty}^\infty J d\Phi(z) d\psi_\nu(u) \tag{H.2}$$

where $J = \prod_{i=1}^{k-1} \Phi\left[\frac{\gamma_i z + x u}{\sqrt{1 - \gamma_i^2}}\right]$ and $\Phi(\cdot)$ be the cumulative distribution function of standard normal variable such that

$$\frac{d\Phi(z)}{dz} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{H.3}$$

is the standard normal density function and

$$\frac{d\psi_\nu(u)}{du} = \frac{\nu^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right) 2^{\frac{\nu}{2}-1}} u^{\nu-1} \exp\left(-\frac{\nu u^2}{2}\right) \tag{H.4}$$

H Theory - Multiple Comparison Procedures

is the density of $\sqrt{\frac{V}{\nu}}$, where V is a Chi-squared random variable with ν degrees of freedom and $\nu = N - k$. The parameter γ_i is

$$\gamma_i = \sqrt{\frac{n_i}{n_0 + n_i}} \quad (\text{H.5})$$

- Test statistics:

$$T_i = \frac{\bar{y}_i - \bar{y}_0}{s\sqrt{\frac{1}{n_i} + \frac{1}{n_0}}} \quad (i = 1, 2, \dots, k - 1) \quad (\text{H.6})$$

where

$$s^2 = \frac{1}{N - k} \sum_{i=0}^{k-1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (\text{H.7})$$

is the pooled sample variance.

- The critical values for single step Dunnett, denoted by c_α , satisfied the following equation
 - For the right tailed test

$$\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^{k-1} \Phi \left[\frac{\gamma_i z + c_\alpha u}{\sqrt{1 - \gamma_i^2}} \right] d\Phi(z) d\psi_\nu(u) = 1 - \alpha \quad (\text{H.8})$$

- For the left tailed test

$$\int_0^\infty \int_{-\infty}^\infty J d\Phi(z) d\psi_\nu(u) = 1 - \alpha \quad (\text{H.9})$$

where $J = \prod_{i=1}^{k-1} \left(1 - \Phi \left[\frac{\gamma_i z + c_\alpha u}{\sqrt{1 - \gamma_i^2}} \right] \right)$.

- Decisions:
 - For the right tailed test, reject H_i if $t_i > c_\alpha$
 - For the left tailed test, reject H_i if $t_i < c_\alpha$
- Adjusted p -values for individual hypothesis H_i : $\tilde{p}_i = 1 - F(t_i)$ where
 - For the right tailed tests:

$$F(t_i) = \int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^{k-1} \Phi \left[\frac{\gamma_i z + t_i u}{\sqrt{1 - \gamma_i^2}} \right] d\Phi(z) d\psi_\nu(u) \quad (\text{H.10})$$

– For the left tailed tests:

$$F(t_i) = \int_0^\infty \int_{-\infty}^\infty Jd\Phi(z) d\psi_\nu(u) \tag{H.11}$$

where $J = \prod_{i=1}^{k-1} \left(1 - \Phi \left[\frac{\gamma_i z + t_i u}{\sqrt{1 - \gamma_i^2}} \right] \right)$.

■ Adjusted p - value for testing the global null hypothesis H_0 :

– For the right tailed tests $\tilde{p} = 1 - F(t_{(1)})$ where $t_{(1)} = \max \{t_i : i = 1, \dots, k - 1\}$ and

$$F(t_{(1)}) = \int_0^\infty \int_{-\infty}^\infty Jd\Phi(z) d\psi_\nu(u) \tag{H.12}$$

where $J = \prod_{i=1}^{k-1} \Phi \left[\frac{\gamma_i z + t_{(1)} u}{\sqrt{1 - \gamma_i^2}} \right]$.

– For the left tailed tests $\tilde{p} = 1 - F(t_{(k-1)})$ where $t_{(k-1)} = \min \{t_i : i = 1, \dots, k - 1\}$ and

$$F(t_{(k-1)}) = \int_0^\infty \int_{-\infty}^\infty Jd\Phi(z) d\psi_\nu(u) \tag{H.13}$$

where $J = \prod_{i=1}^{k-1} \left(1 - \Phi \left[\frac{\gamma_i z + t_{(k-1)} u}{\sqrt{1 - \gamma_i^2}} \right] \right)$.

H.1.3 Step Down Dunnett Test in One-Way ANOVA

Let $H_{(i)}$ be the associated null hypothesis with $t_{(i)}$ ($i = 1, \dots, k - 1$). Let $n_{(i)}$ be the number of subjects for the treatment arm associated with $H_{(i)}$. Let R_{k-1} be the correlation matrix of the unordered statistics associated with $H_{(1)}, H_{(2)}, \dots, H_{(k-1)}$ which has the element at i th row and j th column $\rho_{ij} = \gamma_i \gamma_j$ where $\gamma_i = \sqrt{\frac{n_{(i)}}{n_{(i)} + n_0}}$ and $\gamma_j = \sqrt{\frac{n_{(j)}}{n_{(j)} + n_0}}$. Let $\nu = N - k$. Let c_i ($i = 1, 2, \dots, k - 1$) be the critical values for step-down Dunnett procedure. Let $\Phi(\cdot)$ be the cumulative distribution function of standard normal variable such that

$$\frac{d\Phi(z)}{dz} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{H.14}$$

is the standard normal density function and

$$\frac{d\psi_\nu(u)}{du} = \frac{\nu^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right) 2^{\frac{\nu}{2}-1}} u^{\nu-1} \exp\left(-\frac{\nu u^2}{2}\right) \tag{H.15}$$

H Theory - Multiple Comparison Procedures

is the density of $U = \sqrt{\frac{V}{\nu}}$, where V is a Chi-squared random variable with ν degrees of freedom and $\nu = N - k$.

- Test statistics:

$$T_i = \frac{\bar{y}_i - \bar{y}_0}{s \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}} \quad (\text{H.16})$$

where

$$s^2 = \frac{1}{N - k} \sum_{i=0}^{k-1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (\text{H.17})$$

is the pooled sample variance for all arms.

- Critical values c_i satisfy the following equations

- For the right tailed tests

$$G_i(c_i) = \int_0^\infty \int_{-\infty}^\infty J d\Phi(z) d\psi_\nu(u) = 1 - \alpha \quad (\text{H.18})$$

where $J = \prod_{j=i}^{k-1} \Phi \left[\frac{\gamma_j z + c_i u}{\sqrt{1 - \gamma_j^2}} \right]$.

- For the left tailed tests

$$G_i(c_i) = \int_0^\infty \int_{-\infty}^\infty J d\Phi(z) d\psi_\nu(u) = 1 - \alpha \quad (\text{H.19})$$

where $J = \prod_{j=1}^i \left(1 - \Phi \left[\frac{\gamma_j z + c_i u}{\sqrt{1 - \gamma_j^2}} \right] \right)$.

- Decisions: The step down Dunnett procedure can be carried out as follows:

- For the right tailed tests

- * Step 1: If $t_{(1)} > c_1$, reject $H_{(1)}$ and go to the next step; otherwise retain all hypotheses and stop.
- * Step $i = 2, \dots, k - 2$: If $t_{(i)} > c_i$, reject H_i and go to the next step; otherwise retain $H_{(i)}, H_{(i+1)}, \dots, H_{(k-1)}$ and stop.
- * Step $k - 1$: If $t_{(k-1)} > c_{k-1}$, reject $H_{(k-1)}$ and stop; otherwise retain $H_{(k-1)}$ and stop.

- For the left tailed tests

- * Step 1: If $t_{(k-1)} < c_1$, reject $H_{(k-1)}$ and go to the next step; otherwise retain all hypotheses and stop.

- * Step $i = 2, \dots, k - 2$: If $t_{(k-i)} < c_i$, reject H_{k-i} and go to the next step; otherwise retain $H_{(1)}, H_{(2)}, \dots, H_{(k-i)}$ and stop.
- * Step $k - 1$: If $t_{(1)} < c_{k-1}$, reject $H_{(1)}$ and stop; otherwise retain $H_{(1)}$ and stop.

■ Adjusted p-values for individual hypothesis:

- For the right tailed test

$$\tilde{p}_{(i)} = \begin{cases} p_i & \text{if } i = 1 \\ \max \{ \tilde{p}_{(i-1)}, p_i \} & \text{if } i = 2, \dots, k - 1 \end{cases} \quad (\text{H.20})$$

where

$$p_i = 1 - F(t_{(i)}) \quad (\text{H.21})$$

$$G_i(t_{(i)}) = \int_0^\infty \int_{-\infty}^\infty \prod_{j=i}^{k-1} \Phi \left[\frac{\gamma_j z + t_{(i)} u}{\sqrt{1 - \gamma_j^2}} \right] d\Phi(z) d\psi_\nu(u) \quad (\text{H.22})$$

- For the left tailed test

$$\tilde{p}_{(i)} = \begin{cases} p_i & \text{if } i = k - 1 \\ \max \{ \tilde{p}_{(i+1)}, p_i \} & \text{if } i = k - 2, \dots, 1 \end{cases} \quad (\text{H.23})$$

where

$$p_i = 1 - F(t_{(i)}) \quad (\text{H.24})$$

$$G_i(t_{(i)}) = \int_0^\infty \int_{-\infty}^\infty \prod_{j=1}^i \left(1 - \Phi \left[\frac{\gamma_j z + t_{(i)} u}{\sqrt{1 - \gamma_j^2}} \right] \right) d\Phi(z) d\psi_\nu(u) \quad (\text{H.25})$$

■ Adjusted p-value for the global null hypothesis

- For the right tailed test $\tilde{p} = \tilde{p}_{(1)} = p_1$
- For the left tailed test $\tilde{p} = \tilde{p}_{(k-1)} = p_{k-1}$

H Theory - Multiple Comparison Procedures

H.2 P-value based procedures

H.2.1 Hypotheses etc.-continuous response

H.2.2 Hypotheses etc. binary response

H.2.3 Bonferroni Procedure

H.2.4 Sidak Procedure

H.2.5 Weighted Bonferroni Procedure

H.2.6 Holm Step-Down Procedure

H.2.7 Hochberg Step-Up Procedure

H.2.8 Fixed Sequence Testing Procedure

H.2.9 Hommel Step-Up Procedure

H.2.10 Fallback Procedures

H.2.1 Hypotheses, test statistics and marginal p-values for continuous response

- Individual hypotheses:

- For the right tailed tests

$$H_i : \mu_i \leq \mu_0 \text{ vs } K_i : \mu_i > \mu_0 \quad (i = 1, \dots, k - 1) \quad (\text{H.26})$$

- For the left tailed tests

$$H_i : \mu_i \leq \mu_0 \text{ vs } K_i : \mu_i > \mu_0 \quad (i = 1, \dots, k - 1) \quad (\text{H.27})$$

where k is the total number of arms.

- Global null hypothesis:

$$H_0 : \mu_0 = \mu_1 = \dots = \mu_{k-1} \quad (\text{H.28})$$

against the alternative hypothesis H_{01} : at least one $\mu_i > \mu_0$ for right tailed test or $\mu_i < \mu_0$ for left tailed test.

- Test statistics: The calculation for test statistics is slightly different depending on whether the checkbox for **Common Standard Deviation** is checked or not.
 - If **Common Standard Deviation** for design is checked (or **Equal Variance** for analysis is selected),

$$T_i = \frac{\bar{y}_i - \bar{y}_0}{s \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}} \quad (i = 1, 2, \dots, k - 1) \quad (\text{H.29})$$

where

$$s^2 = \frac{1}{N - k} \sum_{i=0}^{k-1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (\text{H.30})$$

is the variance estimate pooled for all arms, y_{ij} is the response for j^{th} subject in i^{th} arm, \bar{y}_i is the sample mean for the i^{th} arm, N is the total sample size and n_i ($i = 0, 1, \dots, k - 1$) is the number of subjects in arm i

- If **Common Standard Deviation** for design is not checked (**Unequal Variance** for analysis is selected),

$$T_i = \frac{\bar{y}_i - \bar{y}_0}{\sqrt{\frac{1}{n_i} s_i^2 + \frac{1}{n_0} s_0^2}} \quad (i = 1, 2, \dots, k - 1) \quad (\text{H.31})$$

where

$$s_0^2 = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} (y_{0j} - \bar{y}_0)^2 \quad (\text{H.32})$$

is the variance estimate for the control arm and

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (\text{H.33})$$

is the variance estimate for the i^{th} arm.

- Marginal p-values:

- For the right tailed test

$$p_i = P(T_i > t_i) = 1 - P(T_i < t_i) = \Phi(-t_i) \quad (\text{H.34})$$

- For the left tailed test

$$p_i = P(T_i < t_i) = \Phi(t_i) \quad (\text{H.35})$$

where T_i follows t distribution with degree of freedom ν and $\Phi(\cdot)$ is the cumulative distribution function of the t distribution with degree of freedom ν and the value of ν depends on whether the checkbox for **Common Standard Deviation** for design is checked (or the radio button for **Equal Variance** or **Unequal Variance** for analysis is selected)

- If **Common Standard Deviation** for design is checked (or **Equal Variance** for analysis is selected)

$$\nu = N - k \quad (\text{H.36})$$

H Theory - Multiple Comparison Procedures

- If **Common Standard Deviation** for design is not checked (or **Unequal Variance** for analysis is selected)

$$\nu = \left(\frac{s_i^2}{n_i} + \frac{s_0^2}{n_0} \right)^2 / \left[\frac{(s_i^2/n_i)^2}{n_i - 1} + \frac{(s_0^2/n_0)^2}{n_0 - 1} \right] \quad (\text{H.37})$$

H.2.2 Hypotheses, test statistics and marginal p-values for binary response

- Individual hypotheses:

- For the right tailed test

$$H_i : \pi_i - \pi_0 = 0 \text{ vs } K_i : \pi_i - \pi_0 > 0 \quad (i = 1, 2, \dots, k - 1) \quad (\text{H.38})$$

- For the left tailed test

$$H_i : \pi_i - \pi_0 = 0 \text{ vs } K_i : \pi_i - \pi_0 < 0 \quad (i = 1, 2, \dots, k - 1) \quad (\text{H.39})$$

where k is the total number of arms.

- Global null hypothesis

$$H_0 : \pi_0 = \pi_1 = \dots = \pi_{k-1} \quad (\text{H.40})$$

against the alternative H_{01} : at least one $\pi_i > \pi_0$ for right tailed test ($\pi_i < \pi_0$ for left tailed test).

- Test statistics: The calculation for test statistics is slightly different depending on whether **Pooled Variance** or **Unpooled Variance** is selected.

- If **Pooled Variance** is selected,

$$T_i = \frac{\hat{\pi}_i - \hat{\pi}_0}{\sqrt{\tilde{\pi}_i (1 - \tilde{\pi}_i) \left(\frac{1}{n_0} + \frac{1}{n_i} \right)}} \quad (i = 1, 2, \dots, k - 1) \quad (\text{H.41})$$

where $\hat{\pi}_i$ is the sample proportion for the i^{th} arm, $\hat{\pi}_0$ is the sample proportion for the control arm, $\tilde{\pi}_i = \frac{n_i \hat{\pi}_i + n_0 \hat{\pi}_0}{n_i + n_0}$ is the pooled sample proportion, N is the total sample size and n_i ($i = 0, 1, \dots, k - 1$) is the number of subjects in arm i

- If **Unpooled Variance** is selected,

$$T_i = \frac{\hat{\pi}_i - \hat{\pi}_0}{\sqrt{\frac{1}{n_i} \hat{\pi}_i (1 - \hat{\pi}_i) + \frac{1}{n_0} \hat{\pi}_0 (1 - \hat{\pi}_0)}} \quad (i = 1, 2, \dots, k - 1) \quad (\text{H.42})$$

where $\hat{\pi}_i$ is the sample proportion for the i^{th} arm $\hat{\pi}_0$ is the sample proportion for the control arm.

- Marginal p-values:

- For the right tailed test

$$p_i = P(T_i > t_i) = 1 - P(T_i < t_i) = \Phi(-t_i) \quad (\text{H.43})$$

- For the left tailed test

$$p_i = P(T_i < t_i) = \Phi(t_i) \quad (\text{H.44})$$

where T_i follows standard normal distribution and $\Phi(\cdot)$ is the cumulative distribution function

H.2.3 Bonferroni Procedure

Suppose p_1, p_2, \dots, p_{k-1} are the marginal p -values associated with H_i ($i = 1, 2, \dots, k - 1$). Let $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k-1)}$ be the ordered p -values. Suppose α is the significance level.

- The Bonferroni procedure will reject H_i , if $p_i < \frac{\alpha}{k-1}$, $i = 1, 2, \dots, k - 1$.
- The adjusted p -value for the individual hypothesis H_i is given by

$$\tilde{p}_i = \min(1, (k - 1) p_i), \quad i = 1, 2, \dots, k - 1 \quad (\text{H.45})$$

- The adjusted p -value for the global null hypothesis is given by

$$\begin{aligned} \tilde{p} &= \min\{\tilde{p}_i : i = 1, 2, \dots, k - 1\} \\ &= \min(1, mp_{(1)}) \end{aligned} \quad (\text{H.46})$$

H.2.4 Sidak Procedure

Let p_1, p_2, \dots, p_{k-1} be the marginal p -values associated with H_i ($i = 1, 2, \dots, k - 1$). Let $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k-1)}$ be the ordered p -values. Let α be the significance level.

H Theory - Multiple Comparison Procedures

- The Sidak procedure will reject H_i if $p_i < 1 - (1 - \alpha)^{\frac{1}{k-1}}$, $i = 1, 2, \dots, k - 1$.
- The adjusted p -value for the individual hypothesis H_i is given by

$$\tilde{p}_i = 1 - (1 - p_i)^{k-1}, \quad i = 1, 2, \dots, k - 1 \quad (\text{H.47})$$

- The adjusted p -value for the global null hypothesis is given by

$$\begin{aligned} \tilde{p} &= \min \{ \tilde{p}_i : i = 1, 2, \dots, k - 1 \} \\ &= 1 - (1 - p_{(1)})^{k-1} \end{aligned} \quad (\text{H.48})$$

H.2.5 Weighted Bonferroni Procedure

Let p_1, p_2, \dots, p_{k-1} be the marginal p -values associated with H_i ($i = 1, 2, \dots, k - 1$). Let α be the significance level. Let α be the overall type I error rate. Let w_1, w_2, \dots, w_{k-1} be the proportions indicating the allocations of α to each hypothesis such that $\sum_{i=1}^{k-1} w_i = 1$. Let $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k-1)}$ be the order p -values.

- The weighted Bonferroni procedure will reject H_i if $p_i < w_i \alpha$, $i = 1, 2, \dots, k - 1$.
- The adjusted p -value for the individual hypothesis H_i is given by

$$\tilde{p}_i = \min \left(1, \frac{p_i}{w_i} \right), \quad i = 1, 2, \dots, k - 1 \quad (\text{H.49})$$

- The adjusted p -value for the global null hypothesis is given by

$$\tilde{p} = \min \{ \tilde{p}_i : i = 1, 2, \dots, k - 1 \} \quad (\text{H.50})$$

Note that, if $w_1 = w_2 = \dots = w_{k-1} = \frac{1}{k-1}$, the weighted Bonferroni procedure is reduced to the regular Bonferroni procedure.

H.2.6 Holm Step-Down Procedure

Let p_1, p_2, \dots, p_{k-1} be the marginal p -values. Let $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k-1)}$ be the order p -values and $H_{(i)}$ ($i = 1, 2, \dots, k - 1$) be the associated hypotheses. Let α be the significance level. Holm (1979) step-down procedure is carried out as follows:

- Step 1: If $p_{(1)} \leq \frac{\alpha}{k-1}$, reject $H_{(1)}$ and go to the next step. Otherwise retain all hypotheses and stop

- Step $i = 2, \dots, k - 2$: If $p_{(i)} \leq \frac{\alpha}{k-i}$, reject $H_{(i)}$ and go to the next step. Otherwise retain $H_{(i)}, \dots, H_{(k-1)}$ and stop
- Step $k - 1$. If $p_{(k-1)} \leq \alpha$, reject $H_{(k-1)}$ and stop. Otherwise retain $H_{(k-1)}$ and stop.

The adjusted p -value for the individual hypothesis $H_{(i)}$ ($i = 1, 2, \dots, k - 1$) is given by

$$\tilde{p}_{(i)} = \begin{cases} \min(1, (k - 1) p_{(i)}) & \text{if } i = 1, \\ \max(\tilde{p}_{(i-1)}, (k - i) p_{(i)}, 1) & \text{if } i = 2, \dots, k - 1. \end{cases} \quad (\text{H.51})$$

The adjusted p -value for the global hypothesis H_0 is

$$\tilde{p}_{(1)} = \min(1, (k - 1) p_{(1)}) \quad (\text{H.52})$$

H.2.7 Hochberg Step-Up Procedure

Let p_1, p_2, \dots, p_{k-1} be the marginal p -values. Let $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k-1)}$ be the order p -values and $H_{(i)}$ ($i = 1, 2, \dots, k - 1$) be the associated hypotheses. Let α be the significance level. Hochberg (1988) step-up procedure is carried out as follows:

- Step 1: If $p_{(k-1)} > \alpha$, retain $H_{(k-1)}$ and go to the next step. Otherwise reject all hypotheses and stop
- Step $i = 2, \dots, k - 2$: if $p_{(k-i)} > \frac{\alpha}{i}$, retain $H_{(k-i)}$ and go to the next step. Otherwise reject all remaining hypotheses and stop.
- Step $k - 1$: If $p_{(1)} > \frac{\alpha}{k-1}$, retain $H_{(1)}$ and stop. Otherwise reject $H_{(1)}$ and stop.

The adjusted p -values for individual hypothesis is given by

$$\tilde{p}_{(i)} = \begin{cases} p_{(i)} & \text{if } i = k - 1 \\ \min(\tilde{p}_{(i+1)}, (k - i) p_{(i)}) & \text{if } i = k - 2, k - 3, \dots, 1 \end{cases} \quad (\text{H.53})$$

The adjusted p -value for the global null hypothesis is

$$\tilde{p} = \min \{ \tilde{p}_{(i)} : i = 1, 2, \dots, k - 1 \} \quad (\text{H.54})$$

$$= \min \{ p_{(k-1)}, 2p_{(k-2)}, \dots, ip_{(k-i)}, \dots, (k - 1) p_{(1)} \}$$

Compared with Simes adjusted p -value, Hochberg adjusted p -value tends to be larger for testing the global hypothesis.

H Theory - Multiple Comparison Procedures

H.2.8 Fixed Sequence Testing Procedure

Assume that H_1, H_2, \dots, H_{k-1} are ordered hypotheses and the order is prespecified so that H_1 is tested first followed by H_2 and so on. Let p_1, p_2, \dots, p_{k-1} be the associated raw marginal p values. Let α be the significance level. The fixed sequence testing procedure can be carried out as follows:

- Step 1: If $p_1 < \alpha$, reject H_1 and go the next step. Otherwise retain all hypotheses and stop.
- Step $i = 2, 3, \dots, k - 2$: If $p_i < \alpha$, reject H_i and go the the next step. Otherwise retain $H_i, H_{i+1}, \dots, H_{k-1}$.
- Step $k - 1$: If $p_{k-1} < \alpha$, reject H_{k-1} and stop. Otherwise retain H_{k-1} and stop.

The adjusted p - values for individual hypothesis H_i ($i = 1, \dots, k$) is given by

$$\tilde{p}_i = \max \{p_1, p_2, \dots, p_i\} \tag{H.55}$$

The adjusted p - value for the global null hypothesis is given by

$$\tilde{p} = p_1 \tag{H.56}$$

H.2.9 Hommel Step-Up Procedure

Let p_1, p_2, \dots, p_{k-1} be the marginal p -values. Let $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k-1)}$ be the ordered p -values and $H_{(i)}$ ($i = 1, 2, \dots, k - 1$) be the associated hypotheses. Let α be the significance level. The Hommel procedure is carried out as follows:

- Step 1: If $p_{(k-1)} > \alpha$, retain $H_{(k-1)}$ and go to the next step. Otherwise reject all hypotheses and stop.
- Step $i = 2, \dots, k - 2$: If $p_{(k-j)} > \frac{i-j+1}{i} \alpha$ for $j = 1, \dots, i$, retain $H_{(k-i)}$ and go to the next step. Otherwise reject all remaining hypotheses with $p_{(k-1)} < \frac{\alpha}{i-1}$ and stop.
- Step $k - 1$: If $p_{(k-j)} > \frac{k-j}{k-1} \alpha$ for $j = 1, \dots, k - 1$, retain $H_{(1)}$; otherwise reject $H_{(1)}$ if $p_{(1)} < \frac{\alpha}{k-2}$.

Another way of describing Hommel procedure is as follows:

Let $J \subseteq \{1, 2, \dots, k - 1\}$ be defined as $J = \{i \mid i \text{ belongs to } \{1, 2, \dots, k - 1\} \text{ such that } p_{(k-j)} > \frac{i-j+1}{i} \alpha \text{ for all } j = 1, 2, \dots, i\}$. If J is nonempty, reject H_{k-1} whenever $p_{k-1} \leq \frac{\alpha}{v'}$ with $v' = \max_{i \in J} \{i\}$. If J is empty, reject all H_i ($i = 1, \dots, k - 1$).

The adjusted p -values for Hommel procedure can be calculated as

$$\tilde{p}_i = \max \{p_I : i \in I\} \tag{H.57}$$

where p_I denotes the p -value for testing the intersection hypothesis H_I using Simes (1986) test.

H.2.10 Fallback Procedures

Assume that H_1, H_2, \dots, H_{k-1} are ordered hypotheses and the order is prespecified so that H_1 is tested first followed by H_2 and so on. Let p_1, p_2, \dots, p_{k-1} be the marginal p values. Let α be the overall type I error rate. Let w_1, w_2, \dots, w_{k-1} be the proportions indicating the allocations of α to each hypothesis such that $\sum_{i=1}^{k-1} w_i = 1$. The amount of type I error assigned to hypothesis H_i ($i = 1, 2, \dots, k - 1$) is $w_i\alpha$. The fallback procedures can be carried out as follows:

- Step 1: Test H_1 at $\alpha_1 = w_1\alpha$. If $p_1 \leq \alpha_1$, reject H_1 and go to the next step; otherwise retain it and go to the next step
- Step $i = 2, \dots, k - 2$: Test H_i at $\alpha_i = \alpha_{i-1} + w_i\alpha$ if H_{i-1} is rejected and at $\alpha_i = w_i\alpha$ if H_{i-1} is retained. If $p_i \leq \alpha_i$, reject H_i ; otherwise retain it and go to the next step.
- Step $k - 1$: Test H_{k-1} at $\alpha_{k-1} = \alpha_{k-2} + w_{k-1}\alpha$ if H_{k-2} is rejected and at $\alpha_{k-1} = w_{k-1}\alpha$ if H_{k-2} is retained. If $p_{k-1} \leq \alpha_{k-1}$, reject H_{k-1} ; otherwise retain it.

The adjusted p -values for the fallback procedure can be computed as

$$\tilde{p}_i = \max_{J:i \in J} \{p_J\} \tag{H.58}$$

where p_J denotes the p -value for testing the intersection hypothesis H_J using weighted Bonferroni test. The algorithm is described in Appendix A in the paper by Wiens and Dmitrienko (2005). The fallback procedure is equivalent to the closed test using Weighted Bonferroni for the intersection hypotheses. The following algorithm described in Appendix A in the paper by Wiens and Dmitrienko (2005) is used to assign weights to each elementary hypothesis of a particular intersection hypothesis. Let $I = \{1, 2, \dots, k - 1\}$ be the index set. Assume that H_1, H_2, \dots, H_{k-1} is already ordered so that H_1 is tested first followed by H_2 and so on as described in the fall back procedure above. Let w_1, w_2, \dots, w_{k-1} be the associated weights initially assigned to H_1, H_2, \dots, H_{k-1} respectively such that $\sum_{i=1}^{k-1} w_i \leq 1$. For any intersection hypothesis H_J , let $\mathbf{v} = (v_1(H_J), v_2(H_J), \dots, v_{k-1}(H_J))$ be the decision vector to test H_J . This decision vector represents a weighted Bonferroni test for H_J in the following sense. We will compare p_1 with $v_1(H_J)\alpha$, p_2 with $v_2(H_J)\alpha, \dots, p_{k-1}$ with $v_{k-1}(H_J)\alpha$. The following algorithm shows how to determine the decision vector for a particular intersection hypothesis H_J .

H Theory - Multiple Comparison Procedures

- Step 1: $v_1(H_J) = w_1$ if H_J contains H_1 and 0 otherwise
- Step 2: $v_2(H_J) = w_1 + w_2 - v_1(H_J)$ if H_J contains H_2 and 0 otherwise.
-
- Step i : $v_i(H_J) = w_1 + w_2 + \dots + w_i - v_1(H_J) - v_2(H_J) - \dots - v_{i-1}(H_J)$ if H_J contains H_i and 0 otherwise.
-
- Step $k - 1$:
 $v_{k-1}(H_J) = w_1 + \dots + w_{k-1} - v_1(H_J) - v_2(H_J) - \dots - v_{k-2}(H_J)$ if H_J contains H_{k-1} and 0 otherwise.

Once we obtain the decision vector \mathbf{v} according to the above algorithm, we can compute the weighted Bonferroni adjusted p -values as follows for a particular intersection hypothesis H_J as follows

$$p_J = \min_{i=1, \dots, k-1} \{p_i/v_i(H_J)\} \tag{H.59}$$

Consequently, the adjusted p -value for fallback procedure is

$$\tilde{p}_i = \max_{J:i \in J} \{p_J\} \tag{H.60}$$

For example, suppose we have three hypotheses of interest H_1, H_2, H_3 and w_1, w_2, w_3 are the associated weights. The fallback procedure is carried out as follows:

- Step 1: Test H_1 at $\alpha_1 = w_1\alpha$. If $p_1 \leq \alpha_1$, reject H_1 and go to the next step; otherwise retain it and go to the next step
- Step 2: Test H_2 at $\alpha_2 = \alpha_1 + w_2\alpha$ if H_1 is rejected and at $\alpha_2 = w_2\alpha$ if H_1 is retained. If $p_2 \leq \alpha_2$, reject H_2 ; otherwise retain it and go to the next step.
- Step 3: Test H_3 at $\alpha_3 = \alpha_2 + w_3\alpha$ if H_2 is rejected and at $\alpha_3 = w_3\alpha$ if H_2 is retained. If $p_3 \leq \alpha_3$, reject H_3 ; otherwise retain it.

To calculate the adjusted p -values, we first need to obtain the decision vectors for all the intersection hypotheses. In this example, we have 7 intersection hypotheses including the three single hypotheses. The decision vectors are given in the following table

Hence the adjusted p -value for H_1 is $\max \{p_{\{123\}}, p_{\{12\}}, p_{\{13\}}, p_{\{1\}}\}$. Similarly the adjusted p -value for H_2 is $\max \{p_{\{123\}}, p_{\{12\}}, p_{\{23\}}, p_{\{2\}}\}$ and that for H_3 is

$$\max \{p_{\{123\}}, p_{\{13\}}, p_{\{23\}}, p_{\{3\}}\}.$$

| Intersection | Decision Vectors | Bonferrni p -values |
|---------------|--------------------|---|
| $H_{\{123\}}$ | (w_1, w_2, w_3) | $p_{\{123\}} = \min \left\{ \frac{p_1}{w_1}, \frac{p_2}{w_2}, \frac{p_3}{w_3} \right\}$ |
| $H_{\{12\}}$ | (w_1, w_2) | $p_{\{12\}} = \min \left\{ \frac{p_1}{w_1}, \frac{p_2}{w_2} \right\}$ |
| $H_{\{13\}}$ | $(w_1, w_2 + w_3)$ | $p_{\{13\}} = \min \left\{ \frac{p_1}{w_1}, \frac{p_3}{w_2 + w_3} \right\}$ |
| $H_{\{23\}}$ | $(w_1 + w_2, w_3)$ | $p_{\{23\}} = \min \left\{ \frac{p_2}{w_2 + w_3}, \frac{p_3}{w_3} \right\}$ |
| $H_{\{1\}}$ | w_1 | $p_{\{1\}} = \frac{p_1}{w_1}$ |
| $H_{\{2\}}$ | $w_1 + w_2$ | $p_{\{2\}} = \frac{p_2}{w_1 + w_2}$ |
| $H_{\{3\}}$ | $w_1 + w_2 + w_3$ | $p_{\{3\}} = \frac{p_3}{w_1 + w_2 + w_3}$ |

H.3 Generate Means/Proportions through DR Curves

- Four Parameter Logistic

$$E(Y | D) = \beta + \frac{\delta}{1 + \exp\left(\frac{\theta - D}{\tau}\right)} \tag{H.61}$$

where $\tau > 0, -\infty < \beta, \delta, \theta < \infty$

- Linear

$$E(Y | D) = a + bD \tag{H.62}$$

where E_0 is the intercept and b represents the slope.

- Quadratic

$$E(Y | D) = E_0 + B_1 * D + B_2 * D^2 \tag{H.63}$$

where E_0 represents the mean response for placebo, B_1 represents the linear coefficient and B_2 represents the quadratic coefficient.

- E_{max}

$$E(Y | D) = E_0 + \frac{E_{max}}{1 + \exp\{S [\ln(ED_{50}) - \ln(D)]\}} \tag{H.64}$$

where E_0 represents the y -intercept, E_{max} is the difference between the mean response at a very large dose and placebo, $ED_{50} > 0$ is the value of the dose that gives a response of $E_0 + \frac{1}{2}E_{max}$ and $S > 0$ is a slope factor (Hill parameter) that controls the rate at which response increases as a function of dose at ED_{50} .

I Theory - Multiple Endpoint Procedures

I.1 Serial Gatekeeping

Assume that we are interested in testing K endpoints which are grouped into m families F_1, F_2, \dots, F_m . A family is called a serial gatekeeper if all hypotheses must be rejected within that family in order to proceed to test the hypotheses in the next family. In other words, if $F_i (i = 1, 2, \dots, m - 1)$ is a serial gatekeeper, then hypotheses in the next family F_{i+1} are tested only if all the hypotheses in F_i are rejected. Serial gatekeeping over m families is implemented in the following m steps. Note that in the following serial gatekeeping testing procedure any α -level FWER-controlling multiple testing procedure can be used for testing the preceding $m-1$ families. But since we need to reject all hypotheses in one family in order to proceed to test the next family, the most powerful test is the intersection-union (IU) test. The IU test is a min test which is tailored to test a composite null hypothesis. For example, the IU test would reject $\cup_{j=1}^{n_i} H_{in_i}$ if all the hypotheses H_{in_i} in F_i are rejected at their α -level tests, i.e. $\max_{j=1, \dots, n_i} p_{ij} \leq \alpha$.

Serial gatekeeping procedure based on intersection-union test

- Step 1: Test all the hypotheses in F_1 at their nominal α levels using the intersection-union test; i.e., reject all H_{1j} if $\max_{j=1, \dots, n_1} p_{1j} \leq \alpha$, $j = 1, 2, \dots, n_1$. If all the n_1 hypotheses are rejected, go to Step 2, otherwise stop. The term intersection-union test arises because, as shown by Berger (1982), this procedure offers level- α protection against rejecting the null hypothesis $H_1 = \cup_{j=1}^{n_1} H_{1j}$ in favor of the alternative hypothesis $\bar{H}_1 = \cap_{j=1}^{n_1} \bar{H}_{1j}$.
 - Step 2: Test all the hypotheses in F_2 at their nominal α levels using the intersection-union test. If all the hypotheses are rejected, go to step 3, otherwise stop.
- ⋮
- Step $m-1$: Test all the hypotheses in F_{m-1} at their nominal α levels using the intersection-union test. If all the hypotheses are rejected, go to step m , otherwise stop.
 - Step m : Test all the hypotheses in F_m using any multiple testing procedure that guarantees strong control of type-1 error within the family F_m .

To obtain adjusted p-values, let p_i^* denote the largest p-value in F_i , for

$i = 1, 2, \dots, m - 1$. Then

$$\tilde{p}_{ij} = \begin{cases} \max(p_1^*, p_2^*, \dots, p_i^*) & \text{if } i = 1, 2, \dots, m - 1 \\ \max(p'_{ij}, p_1^*, p_2^*, \dots, p_{i-1}^*) & \text{if } i = m \end{cases}$$

where p'_{mj} is the adjusted p-value for H_{mj} based on the multiple testing procedure that has been adopted for family F_m . In terms of adjusted p-values, the serial gatekeeping condition implies that hypotheses in family F_{i+1} will only be tested if

$$\max_{j=1, \dots, n_i} \tilde{p}_{ij} \leq \alpha$$

1.2 Parallel Gatekeeping

Assume that we are interested in testing K endpoints which are grouped into m families F_1, F_2, \dots, F_m . F_i is termed a parallel gatekeeper if at least one hypothesis within it must be rejected in order to proceed to family F_{i+1} ($i = 1, 2, \dots, m - 1$). We consider the general multistage parallel gatekeeping procedure proposed by Dmitrienko, Tamhane and Wiens 2008. Control of the FWER relies on using a so-called “separable” multiple testing procedure. In order to define separable tests we require the concept of an error rate function. Consider the problem of testing a single family of n null hypotheses H_1, H_2, \dots, H_n . Let $I \subseteq N$ be the index set of true null hypotheses. The error rate function $e(I)$ of a multiple testing procedure is the maximum probability of making at least one type-1 error.

$$e(I) = \sup P \left(\bigcup_{i \in I} \{\text{Reject } H_i\} \mid H_I \right)$$

where $H_I = \cap_{i \in I} H_i$ and the supremum is computed over the the entire parameter space of the hypotheses in $N \setminus I$. In other words, $e(I)$ is error that the multiple testing procedure will produce under the worst configuration of alternative hypotheses for a specific set of $I \subseteq N$ null hypotheses. An explicit expression for $e(I)$ is not generally available, but an upper bound can be used instead. A multiple testing procedure is separable if its error rate is strictly less than α unless all hypotheses are true. That is, $e(I) < \alpha$ for all. Suppose $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$ are the ordered p-values for corresponding null hypotheses $H_{(1)}, H_{(2)}, \dots, H_{(n)}$. Then the following three multiple testing procedures are separable.

I Theory - Multiple Endpoint Procedures

Bonferroni Test: Bonferroni Test: The upper bound of the error rate function for Bonferroni test is given by

$$e(I) = \frac{|I|}{n} \alpha$$

where $|I|$ is the cardinality of set I .

Note that Bonferroni procedure is separable. But regular Holm or Hochberg is not separable. To see this, consider a family of two hypotheses where one hypothesis is true and the other hypothesis is infinitely false. The type I error of regular Holm applied to such a family of hypotheses would be α . Similar argument applies to regular Hochberg procedure. Hence regular Holm or Hochberg can't be directly used in parallel gatekeeping procedure. However we can modify them by taking the convex combination of their own critical values and the critical values of Bonferroni test. The modified procedures are separable which we call truncated Holm and truncated Hochberg described as follows.

Truncated Holm: For any prespecified truncation fraction γ , the truncated Holm test performs as follows

- Step 1: If $p_{(1)} \leq \frac{\alpha}{n}$, then reject $H_{(1)}$ and go to the next step, otherwise retain all hypotheses.
- Step 2: If $p_{(2)} \leq \left(\frac{\gamma}{n-1} + \frac{1-\gamma}{n} \right) \alpha$, then reject $H_{(2)}$ and go to the next step, otherwise retain $H_{(2)}, H_{(3)}, \dots, H_{(n)}$ and stop.
- ⋮
- Step i : If $p_{(i)} \leq \left(\frac{\gamma}{n-i+1} + \frac{1-\gamma}{n} \right) \alpha$, then reject $H_{(i)}$ and go to the next step, otherwise retain $H_{(i)}, H_{(i+1)}, \dots, H_{(n)}$ and stop.
- ⋮
- Step n : If $p_{(n)} \leq \left(\gamma + \frac{1-\gamma}{n} \right) \alpha$, then reject $H_{(n)}$, otherwise retain $H_{(n)}$.

The upper bound of the error rate function for truncated Holm is given by

$$e(I) = \begin{cases} \left(\gamma + \frac{(1-\gamma)|I|}{n}\right) \alpha & \text{if } |I| > 0 \\ 0 & \text{if } |I| = 0 \end{cases}$$

Truncated Hochberg: For any prespecified truncation fraction γ , the truncated Hochberg test performs as follows:

- Step 1: If $p_{(n)} \leq \left(\gamma + \frac{1-\gamma}{n}\right) \alpha$, then rejects all hypotheses and stop, otherwise retain $H_{(n)}$ and go to the next step to test $H_{(n-1)}$.
- Step 2: If $p_{(n-1)} \leq \left(\frac{\gamma}{2} + \frac{1-\gamma}{n}\right) \alpha$, then rejects $H_{(1)}, H_{(2)}, \dots, H_{(n-1)}$ and stop, otherwise retain $H_{(n-1)}$ and go to the next step to test $H_{(n-2)}$.
- ⋮
- Step i : If $p_{(i)} \leq \left(\frac{\gamma}{n-i+1} + \frac{1-\gamma}{n}\right) \alpha$, then reject $H_{(1)}, H_{(2)}, \dots, H_{(i)}$ and stop, otherwise retain $H_{(i)}$ and go to the next step to test $H_{(i-1)}$.
- ⋮
- Step n : If $p_{(1)} \leq \frac{\alpha}{n}$, then reject $H_{(1)}$ and stop, otherwise retain $H_{(1)}$ and stop.

The upper bound of truncated Hochberg test is given by

$$e(I) = \begin{cases} 1 - P \left\{ p_{(i)}(I) > \left[\frac{\gamma}{|I|-i+1} + \frac{1-\gamma}{n} \right] \alpha \text{ for all } i \in I \right\} & \text{if } |I| > 0 \\ 0 & \text{if } |I| = 0 \end{cases}$$

In general, the upper bound on $e(I)$ for truncated Holm can also be used for truncated Hochberg. Using the above expression, we can obtain more stringent upper bound than the one for truncated Holm. Consequently, more type I error will be carried over to the next family. Observe that the above expression for the error rate function requires knowledge of the joint distribution of the p-values. If the p-values are for comparisons of multiple treatments versus a common control, then the correlations among them are known and the error rate function can be evaluated. If, however, the p-values are for comparisons of a single treatment versus a control with respect to multiple endpoints, we typically will not know the correlations amongst these endpoints. In that case we can obtain a conservative upper bound for the error rate function by assuming

I Theory - Multiple Endpoint Procedures

independence of p-values and using the following result due to Sen (1999): Let $U_{(1)} < \dots < U_{(k)}$ denote the order statistics of $k > 1$ i.i.d. observations from a uniform (0,1) distribution. For any $0 < a_1 < \dots < a_k < 1$,

$$P(a_1, a_2, \dots, a_k) = P(U_{(i)} > a_i \text{ for all } i = 1, \dots, k) = k!H_k(1)$$

where $H_i(u) = \int_{a_i}^u H_{i-1}(v)dv$, $i = 1, \dots, k$ and $H_0(u) = I(u \geq a_1)$ and $I(\cdot)$ is an indicator function.

Consider $m \geq 2$ families, $F_i = \{H_{i1}, \dots, H_{in_i}\}$ ($1 \leq i \leq m$) of null hypotheses. Let $N_i = \{1, 2, \dots, n_i\}$ and $A_i \subseteq N_i$ be the index set corresponding to the accepted hypotheses in F_i . Parallel gatekeeping is implemented in the following m steps.

Step 1 : Let $\alpha_1 = \alpha$ and test all hypotheses in F_1 at level α_1 using any separable multiple testing procedure (Bonferroni, Truncated Holm, Truncated Hochberg) with a suitable upper bound on the error rate function $e_1(I)$. If $A_1 = N_1$, i.e., no hypotheses in F_1 are rejected, then stop testing and retain all hypotheses in F_2, \dots, F_m ; otherwise go to the next step.

Step 2: Let $\alpha_2 = \alpha_1 - e_1(A_1)$ and test all hypotheses in F_2 at level α_2 using any of the separable multiple test procedures with a suitable upper bound on the error rate function $e_2(I)$. If $A_2 = N_2$, i.e. no hypotheses in F_2 are rejected, then stop testing and retain all hypotheses in F_3, \dots, F_m ; otherwise go to the next step.

⋮

Step i: Let $\alpha_i = \alpha_{i-1} - e_{i-1}(A_{i-1})$ and test all hypotheses in F_i at level α_i using any of the separable multiple test procedures with a suitable upper bound on the error rate function $e_i(I)$. If $A_i = N_i$, i.e. no hypotheses in F_i are rejected, then stop testing and retain all hypotheses in F_{i+1}, \dots, F_m ; otherwise go to the next step.

⋮

Step m: Let $\alpha_m = \alpha_{m-1} - e_{m-1}(A_{m-1})$ and test all hypotheses in F_m at level α_m using any of multiple test procedures which don't have to be separable.

Adjusted P Values:

The adjusted p-values associated with the gatekeeping procedure can be computed by looping through a discrete grid of significance levels. Let $\alpha = \frac{k}{K}$ ($0 < k < K$) for some sufficiently large value of K . The adjusted p-value, \tilde{p}_{ij} for hypotheses H_{ij} is the smallest α (corresponding to the smallest k) for which H_{ij} is rejected.

J Theory-Multi-arm Multi-stage Group Sequential Design

J.1 Notations

Let $\delta_i = \mu_i - \mu_0$ be the mean difference for group i versus control group. Suppose that there are K_1 analysis times including final one. Assume unequal sample size allocation. We use the first subscript index to denote doses and the second subscript to denote interim analysis time. Let $n_{i1} < \dots < n_{iK_1}$ ($i = 0, 1, 2, \dots, D$) be the cumulative sample size for group i at each interim where 0 denotes control arm. Let $n_{i(j)}$ be the incremental sample size from look $j - 1$ to look j . Let σ_i^2 ($i = 0, 1, 2, \dots, D$) be the variance for responses in the i th group. Let $\bar{X}_{i(j)}$ ($i = 0, 1, 2, \dots, D; j = 1, 2, \dots, K_1$) be the sample mean based on the incremental data from look $j - 1$ to j for the i th group. Let $\hat{\delta}_{i(j)} = \bar{X}_{i(j)} - \bar{X}_{0(j)}$ ($i = 1, 2, \dots, D$) be the observed mean difference from control group for group i . Let $\xi_{i(j)} = [\text{var}(\bar{X}_{i(j)})]^{-1} = \frac{n_{i(j)}}{\sigma_i^2}$ be the incremental information from look $j - 1$ to look j for the i th group. Let $\xi_{ij} = \sum_{h=1}^j \xi_{i(h)}$. Let $I_{i(j)} = [\text{var}(\bar{X}_{i(j)} - \bar{X}_{0(j)})]^{-1} = [\xi_{i(j)}^{-1} + \xi_{0(j)}^{-1}]^{-1}$. Let $Z_{i(j)} = \hat{\delta}_{i(j)} \sqrt{I_{i(j)}}$ be the Z statistic for the comparison of group i versus control based on incremental data. Let $W_{i(j)} = \hat{\delta}_{i(j)} I_{i(j)} = Z_{i(j)} \sqrt{I_{i(j)}}$ be the score statistic based on incremental data. Let $W_{ij} = \sum_{h=1}^j W_{i(h)} = \sum_{h=1}^j Z_{i(h)} \sqrt{I_{i(h)}} = \sum_{h=1}^j \hat{\delta}_{i(h)} I_{i(h)}$. Assume that we will monitor the trial based on the processes $W_{1j}, W_{2j}, \dots, W_{Dj}$. Let $I_{ij} = \sum_{h=1}^j I_{i(h)}$ be the cumulative information up to look j for W_{ij} .

Now let N be the total sample size for the whole study. Let $\lambda_i = \frac{n_{i(1)}}{n_{0(1)}} = \frac{n_{i(2)}}{n_{0(2)}} = \dots = \frac{n_{i(K_1)}}{n_{0(K_1)}}$ ($i = 0, 1, 2, \dots, D$) be the sample size allocation ratio of dose i to control group. Note that as long as the allocation ratio for a particular dose to control remains the same across all interim looks, the W_{ij} is the same as $Z_{ij} \sqrt{I_{ij}}$. Let $\frac{n_{0K_1}}{N} = \lambda_0$ be the fraction of total sample size for control arm to total sample size of the whole study. Let $t_{(j)} = \frac{n_{0(j)}}{n_{0K_1}} = \frac{n_{1(j)}}{n_{1K_1}} = \dots = \frac{n_{D(j)}}{n_{DK_1}}$ and let $t_j = \frac{n_{0j}}{n_{0K_1}} = \frac{n_{1j}}{n_{1K_1}} = \dots = \frac{n_{Dj}}{n_{DK_1}}$ be the cumulative sample size fraction up to look j for control arm. Note that $t_{(j)} = \frac{I_{i(j)}}{I_{iK_1}}$, ($i = 0, 1, \dots, D$) and $t_j = \frac{I_{ij}}{I_{iK_1}}$, ($i = 0, 1, \dots, D$). Then we have

$$\xi_{i(j)} = \frac{n_{i(j)}}{\sigma_i^2}$$

$$I_{i(j)} = \left[\frac{\sigma_i^2}{n_{i(h)}} + \frac{\sigma_0^2}{n_{0(h)}} \right]^{-1} = \left[\frac{\sigma_i^2}{\lambda_i} + \sigma_0^2 \right]^{-1} n_{0(j)}$$

$$I_{ij} = \sum_{h=1}^j I_{i(j)} = \left[\frac{\sigma_i^2}{\lambda_i} + \sigma_0^2 \right]^{-1} \sum_{h=1}^j n_{0(h)} = \left[\frac{\sigma_i^2}{\lambda_i} + \sigma_0^2 \right]^{-1} n_{0j}$$

$$E(W_{i(j)}) = \delta_i I_{i(j)}$$

$$Var(W_{i(j)}) = I_{i(j)}$$

$$Cov(W_{k(j)}, W_{l(j)}) = \xi_{0(j)}^{-1} I_{k(j)} I_{l(j)} = \left[\left(\frac{\sigma_k^2}{\lambda_k} + \sigma_0^2 \right) \left(\frac{\sigma_l^2}{\lambda_l} + \sigma_0^2 \right) \right]^{-1} \sigma_0^2 * n_{0(j)}$$

For the cumulative process W_{ij} , we have

$$E(W_{ij}) = E\left(\sum_{h=1}^j W_{i(h)}\right) = \delta_i I_{ij}$$

$$Var(W_{ij}) = Var\left(\sum_{h=1}^j W_{i(h)}\right) = I_{ij}$$

$$Cov(W_{kj}, W_{lj}) = \sum_{h=1}^j Cov(W_{k(h)}, W_{l(h)}) = \left[\left(\frac{\sigma_k^2}{\lambda_k} + \sigma_0^2 \right) \left(\frac{\sigma_l^2}{\lambda_l} + \sigma_0^2 \right) \right]^{-1} \sigma_0^2 * n_{0j}$$

Next we derive the conditional distribution of $\vec{W}_{j_2} = (W_{1j_2}, W_{2j_2}, \dots, W_{Dj_2})$ given $\vec{W}_{j_1} = \vec{x}_{j_1} = (x_{1j_1}, x_{2j_1}, \dots, x_{Dj_1})$ for $j_1 < j_2$. For each process W_{ij_2} , $W_{ij_2} = W_{ij_1} + \sum_{h=j_1+1}^{j_2} W_{i(h)}$. Hence conditional on $W_{ij_1} = x_{ij_1}$, W_{ij_2} has a normal distribution with mean $x_{ij_1} + \delta_i (I_{ij_2} - I_{ij_1})$ and variance $I_{ij_2} - I_{ij_1}$. And conditional on \vec{W}_{j_1} , the covariance between W_{kj_2} and W_{lj_2} is given by

$$Cov(W_{kj_2}, W_{lj_2} | \vec{W}_{j_1}) = Cov\left(\sum_{h=j_1+1}^{j_2} W_{k(h)}, \sum_{h=j_1+1}^{j_2} W_{l(h)}\right)$$

J Theory-Multi-arm Multi-stage Group Sequential Design

$$\begin{aligned}
 &= \sum_{h=j_1+1}^{j_2} Cov(W_{k(h)}, W_{l(h)}) \\
 &= \left[\left(\frac{\sigma_k^2}{\lambda_k} + \sigma_0^2 \right) \left(\frac{\sigma_l^2}{\lambda_l} + \sigma_0^2 \right) \right]^{-1} \sigma_0^2 * \sum_{h=j_1+1}^{j_2} n_{0(h)} \\
 &= \left[\left(\frac{\sigma_k^2}{\lambda_k} + \sigma_0^2 \right) \left(\frac{\sigma_l^2}{\lambda_l} + \sigma_0^2 \right) \right]^{-1} \sigma_0^2 * (n_{0j_2} - n_{0j_1})
 \end{aligned}$$

Then the transition density of the process \vec{W} from look j_1 to look j_2 is given by

$$\begin{aligned}
 p_{\vec{\delta}, \Sigma_{j_2|j_1}} \left(\left(\vec{I}_{j_1}, \vec{x}_{j_1} \right), \left(\vec{I}_{j_2}, \vec{x}_{j_2} \right) \right) &= (2\pi)^{-D/2} | \Sigma_{j_2|j_1} |^{-\frac{1}{2}} \\
 \exp \left\{ - \frac{ \left(\vec{x}_{j_2} - \left[\vec{x}_{j_1} + (A_{j_2} - A_{j_1}) \vec{\delta} \right] \right)^T \left[\Sigma_{j_2|j_1} \right]^{-1} \left(\vec{x}_{j_2} - \left[\vec{x}_{j_1} + (A_{j_2} - A_{j_1}) \vec{\delta} \right] \right) }{2} \right\} & \tag{J.1}
 \end{aligned}$$

where $x_{j_2} = (x_{1j_2}, \dots, x_{Dj_2})^T$, $x_{j_1} = (x_{1j_1}, \dots, x_{Dj_1})^T$, $\vec{\delta} = (\delta_1, \dots, \delta_D)^T$, $\vec{I}_{j_1} = (I_{1j_1}, I_{2j_1}, \dots, I_{Dj_1})^T$, and the matrix $\Sigma_{j_2|j_1} = (\zeta_{kl})_{D \times D}$ and A_j has the following form

$$\zeta_{kl} = \begin{cases} I_{kj_2} - I_{kj_1} = \left[\frac{\sigma_k^2}{\lambda_k} + \sigma_0^2 \right]^{-1} (n_{0j_2} - n_{0j_1}) & \text{if } k = l \\ \left[\left(\frac{\sigma_k^2}{\lambda_k} + \sigma_0^2 \right) \left(\frac{\sigma_l^2}{\lambda_l} + \sigma_0^2 \right) \right]^{-1} \sigma_0^2 * (n_{0j_2} - n_{0j_1}) & \text{if } k \neq l \end{cases}$$

$$A_j = \begin{pmatrix} I_{1j}^{(1)} & 0 & 0 & 0 \\ 0 & I_{2j}^{(1)} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & I_{Dj}^{(1)} \end{pmatrix}$$

J.2 Design

J.2.1 Look 1

J.2.2 Look 2

J.2.3 Look 3

Assume that a group sequential design with K_1 looks including the final look is planned initially. Let $e_j^{(1)}$ ($j = 1, 2, \dots, K_1$) be the level α exit boundaries for the initial design using $W_{ij}^{(1)}$ to test H_0 . The boundaries $e_j^{(1)}$ satisfy

$$P\left(\bigcup_{j=1}^{K_1} \left\{ \max_i \left\{ W_{ij}^{(1)} \right\} > e_j^{(1)} \right\} \mid \vec{\delta} = 0\right) = \alpha$$

Let α_j ($j = 1, \dots, K_1$) be the cumulative type I error by look j such that

$$P\left(\bigcup_{h=1}^j \left\{ \max_i \left\{ W_{ih}^{(1)} \right\} > e_h^{(1)} \right\} \mid \vec{\delta} = 0\right) = \alpha_j$$

Let $T = \max_{i=1, \dots, D} \{I_{iK_1}\} = n_{0K_1} * \max_h \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2 \right]^{-1} \right\}$. Now let $U_{ij} = \frac{W_{ij}}{\sqrt{T}}$. Then the process U_{ij} is a Brownian process with mean $\eta_i \tilde{t}_{ij}$ and variance \tilde{t}_{ij} where

$$\eta_i = \delta_i \sqrt{T}$$

$$\tilde{t}_{ij} = \frac{\left[\frac{\sigma_i^2}{\lambda_i} + \sigma_0^2 \right]^{-1}}{\max_h \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2 \right]^{-1} \right\}} t_j$$

Next we derive the conditional distribution for the process U . Note that

$U_{ij_2} = U_{ij_1} + \frac{1}{\sqrt{T}} \sum_{h=j_1+1}^{j_2} W_{i(h)}$. Hence, conditional on $U_{ij_1} = y_{ij_1}$, U_{ij_2} is normal with mean $y_{ij_1} + \frac{1}{\sqrt{T}} \sum_{h=j_1+1}^{j_2} E(W_{i(h)}) = y_{ij_1} + \eta_i (\tilde{t}_{ij_2} - \tilde{t}_{ij_1})$ and variance $\tilde{t}_{ij_2} - \tilde{t}_{ij_1}$ where and $\vec{\eta} = (\eta_1, \eta_2, \dots, \eta_D)^T$ and $\eta_i = \delta_i \sqrt{T}$. Conditional on \vec{U}_{j_1} , the covariance between U_{kj_2} and U_{lj_2} is given by

$$\begin{aligned} Cov\left(U_{kj_2}, U_{lj_2} \mid \vec{U}_{j_1}\right) &= Cov\left(U_{kj_1} + \frac{1}{\sqrt{T}} \sum_{h=j_1+1}^{j_2} W_{k(h)}, U_{lj_1} + \frac{1}{\sqrt{T}} \sum_{h=j_1+1}^{j_2} W_{l(h)}\right) \\ &= \frac{\left[\left(\frac{\sigma_k^2}{\lambda_k} + \sigma_0^2 \right) \left(\frac{\sigma_l^2}{\lambda_l} + \sigma_0^2 \right) \right]^{-1}}{\max_h \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2 \right]^{-1} \right\}} \sigma_0^2 * (t_{j_2} - t_{j_1}) \end{aligned}$$

J Theory-Multi-arm Multi-stage Group Sequential Design

Hence the transition density of the process U is given by

$$p_{\vec{\eta}, \tilde{\Sigma}_{j_2|j_1}}((t_{j_1}, \vec{y}_{j_1}), (t_{j_2}, \vec{y}_{j_2})) = (2\pi)^{-D/2} |\tilde{\Sigma}_{j_2|j_1}|^{-\frac{1}{2}} \exp \left\{ -\frac{(\vec{y}_{j_2} - [\vec{y}_{j_1} + (A_{j_2} - A_{j_1})\vec{\eta}])^T [\tilde{\Sigma}_{j_2|j_1}]^{-1} (\vec{y}_{j_2} - [\vec{y}_{j_1} + (A_{j_2} - A_{j_1})\vec{\eta}])}{2} \right\}$$

where $\vec{\eta} = (\eta_1, \eta_2, \dots, \eta_D)^T$ and $\eta_i = \delta_i \sqrt{T}$, $\vec{y}_{j_2} = (y_{1j_2}, y_{2j_2}, \dots, y_{Dj_2})^T$ and the covariance matrix $\tilde{\Sigma}_{j_2|j_1} = (\zeta_{kl})_{D \times D}$ has the form

$$\zeta_{kl} = \begin{cases} \tilde{t}_{kj_2} - \tilde{t}_{kj_1} = \frac{\left[\frac{\sigma_k^2}{\lambda_k} + \sigma_0^2\right]^{-1}}{\max_h \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2\right]^{-1} \right\}} (t_{j_2} - t_{j_1}) & \text{if } k = l \\ = \frac{\left[\left(\frac{\sigma_k^2}{\lambda_k} + \sigma_0^2\right) \left(\frac{\sigma_l^2}{\lambda_l} + \sigma_0^2\right)\right]^{-1}}{\max_h \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2\right]^{-1} \right\}} \sigma_0^2 * (t_{j_2} - t_{j_1}) & \text{if } k \neq l \end{cases}$$

and the matrix A_{j_2} has the form

$$A_{j_2} = \begin{pmatrix} \tilde{t}_{1j_2} & 0 & 0 & 0 \\ 0 & \tilde{t}_{2j_2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \tilde{t}_{Dj_2} \end{pmatrix}$$

Note that

$$P \left(\bigcup_{h=1}^j \left\{ \max_i \{W_{ih}^{(1)}\} > e_h^{(1)} \right\} \mid \vec{\delta} = 0 \right) = \alpha_j$$

is equivalent to

$$P \left(\bigcup_{h=1}^j \left\{ \max_i \{U_{ih}\} > \frac{e_h^{(1)}}{\sqrt{T}} \right\} \mid \vec{\delta} = 0 \right) = \alpha_j$$

For Boundary computation, we will work on the process U . Let $b_j = \frac{e_j^{(1)}}{\sqrt{T}}$ be the boundary based on the process U . We can find b_j recursively and the computation for boundary b_j is independent of sample size.

J.2.1 Look 1

The boundaries b_j ($j = 1, \dots, K_1$) satisfy the following equation

$$P\left(\bigcap_{h=1}^{j-1} \left\{ \max_i \{U_{ih}\} \leq b_j \right\} \cap \left\{ \max_i \{U_{ij}\} > b_j \right\} \mid \vec{\delta} = 0\right) = \alpha_j - \alpha_{j-1}$$

More specifically, b_1 satisfies the following equation

$$P\left(\max_i \{U_{i1}\} > b_1 \mid \vec{\delta} = 0\right) = \alpha_1$$

The left hand side of the above equation under any values of $\vec{\delta}$ is

$$1 - \int_{-\infty}^{b_1} \dots \int_{-\infty}^{b_1} p_{\vec{\eta}, \tilde{\Sigma}_{1|0}}\left((0, 0), \left(t_1^{(1)}, \vec{y}_1^{(1)}\right)\right) d\vec{y}_1^{(1)} \quad (J.2)$$

i.e.

$$\int_{-\infty}^{b_1} \dots \int_{-\infty}^{b_1} p_{\vec{\eta}, \tilde{\Sigma}_{1|0}}\left((0, 0), \left(t_1^{(1)}, \vec{y}_1^{(1)}\right)\right) d\vec{y}_1^{(1)} = 1 - \alpha_1$$

where $\vec{y}_1^{(1)} = \left(y_{11}^{(1)}, \dots, y_{D1}^{(1)}\right)^T$, $\vec{\eta} = (\eta_1, \dots, \eta_D)^T$ and $p_{\vec{\eta}, \tilde{\Sigma}_{1|0}}\left((0, 0), \left(t_1^{(1)}, \vec{y}_1^{(1)}\right)\right)$ is the joint density function of U_{11}, \dots, U_{D1} given by

$$p_{\vec{\eta}, \tilde{\Sigma}_{1|0}}\left((0, 0), \left(t_1^{(1)}, \vec{y}_1^{(1)}\right)\right) = (2\pi)^{-\frac{D}{2}} \left| \tilde{\Sigma}_{1|0} \right|^{-\frac{1}{2}}$$

$$\exp\left\{-\frac{\left(\vec{y}_1^{(1)} - A_1 \vec{\eta}\right)^T \left(\tilde{\Sigma}_{1|0}\right)^{-1} \left(\vec{y}_1^{(1)} - A_1 \vec{\eta}\right)}{2}\right\}$$

and

$$A_1 = \begin{pmatrix} \tilde{t}_{11}^{(1)} & 0 & 0 & 0 \\ 0 & \tilde{t}_{21}^{(1)} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \tilde{t}_{D1}^{(1)} \end{pmatrix}$$

J Theory-Multi-arm Multi-stage Group Sequential Design

J.2.2 Look 2

The boundary b_2 satisfies the following equation

$$P \left(\left\{ \max_i \{U_{i1}\} \leq b_1 \right\} \cap \left\{ \max_i \{U_{i2}\} > b_2 \right\} \mid \vec{\delta} = 0 \right) = \alpha_2 - \alpha_1$$

The left hand side of the above equation under any $\vec{\delta}$ is

$$\int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_1} p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left((0, 0), \left(t_1^{(1)}, \vec{y}_1^{(1)} \right) \right)$$

$$\left[1 - \int_{-\infty}^{b_2} \cdots \int_{-\infty}^{b_2} p_{\vec{\eta}, \tilde{\Sigma}_{2|1}} \left(\left(t_1^{(1)}, \vec{y}_1^{(1)} \right), \left(t_2^{(1)}, \vec{y}_2^{(1)} \right) \right) d\vec{y}_2^{(1)} \right] d\vec{y}_1^{(1)} \quad (J.3)$$

i.e.

$$\int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_1} p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left((0, 0), \left(t_1^{(1)}, \vec{y}_1^{(1)} \right) \right)$$

$$\left[\int_{-\infty}^{b_2} \cdots \int_{-\infty}^{b_2} p_{\vec{\eta}, \tilde{\Sigma}_{2|1}} \left(\left(t_1^{(1)}, \vec{y}_1^{(1)} \right), \left(t_2^{(1)}, \vec{y}_2^{(1)} \right) \right) d\vec{y}_2^{(1)} \right] d\vec{y}_1^{(1)}$$

$$= 1 - \alpha_2$$

where $\vec{y}_2^{(1)} = \left(y_{12}^{(1)}, \dots, y_{D2}^{(1)} \right)^T$ and

$$p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left((0, 0), \left(t_1^{(1)}, \vec{y}_1^{(1)} \right) \right) = (2\pi)^{-\frac{D}{2}} \left| \tilde{\Sigma}_{1|0} \right|^{-\frac{1}{2}}$$

$$\exp \left\{ - \frac{\left(\vec{y}_1^{(1)} - A_1 \vec{\eta} \right)^T \tilde{\Sigma}_{1|0}^{-1} \left(\vec{y}_1^{(1)} - A_1 \vec{\eta} \right)}{2} \right\}$$

$$p_{\vec{\eta}, \tilde{\Sigma}_{2|1}} \left(\left(t_1^{(1)}, \vec{y}_1^{(1)} \right), \left(t_2^{(1)}, \vec{y}_2^{(1)} \right) \right) = (2\pi)^{-\frac{D}{2}} \left| \tilde{\Sigma}_{2|1} \right|^{-\frac{1}{2}}$$

$$\exp \left\{ - \frac{\left(\vec{y}_2^{(1)} - \left[\vec{y}_1^{(1)} + (A_2 - A_1) \vec{\eta} \right] \right)^T \tilde{\Sigma}_{2|1}^{-1} \left(\vec{y}_2^{(1)} - \left[\vec{y}_1^{(1)} + (A_2 - A_1) \vec{\eta} \right] \right)}{2} \right\}$$

J.2.3 Look 3

Now let's consider the calculation of the boundary b_3 which satisfies the following

$$P \left(\left\{ \max_i \{U_{i1}\} \leq b_1 \right\} \cap \left\{ \max_i \{U_{i2}\} \leq b_2 \right\} \cap \left\{ \max_i \{U_{i3}\} > b_3 \right\} \mid \vec{\delta} = 0 \right) = \alpha_3 - \alpha_2$$

The left hand side of the above equation under any $\vec{\delta}$ is the following integration

$$\int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_1} p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left((0, 0), \left(t_1^{(1)}, \vec{y}_1^{(1)} \right) \right) \int_{-\infty}^{b_2} \cdots \int_{-\infty}^{b_2} p_{\vec{\eta}, \tilde{\Sigma}_{2|1}} \left(\left(t_1^{(1)}, \vec{y}_1^{(1)} \right), \left(t_2^{(1)}, \vec{y}_2^{(1)} \right) \right) \left\{ 1 - \int_{-\infty}^{b_3} \cdots \int_{-\infty}^{b_3} p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left(\left(t_2^{(1)}, \vec{y}_2^{(1)} \right), \left(t_3^{(1)}, \vec{y}_3^{(1)} \right) \right) d\vec{y}_3^{(1)} \right\} d\vec{y}_2^{(1)} d\vec{y}_1^{(1)} \quad (J.4)$$

i.e.

$$\int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_1} \int_{-\infty}^{b_2} \cdots \int_{-\infty}^{b_2} \int_{-\infty}^{b_3} \cdots \int_{-\infty}^{b_3} p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left((0, 0), \left(t_1^{(1)}, \vec{y}_1^{(1)} \right) \right) p_{\vec{\eta}, \tilde{\Sigma}_{2|1}} \left(\left(t_1^{(1)}, \vec{y}_1^{(1)} \right), \left(t_2^{(1)}, \vec{y}_2^{(1)} \right) \right) p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left(\left(t_2^{(1)}, \vec{y}_2^{(1)} \right), \left(t_3^{(1)}, \vec{y}_3^{(1)} \right) \right) d\vec{y}_3^{(1)} d\vec{y}_2^{(1)} d\vec{y}_1^{(1)} = 1 - \alpha_3$$

J Theory-Multi-arm Multi-stage Group Sequential Design

where $p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left((0, 0), (t_1^{(1)}, \vec{y}_1^{(1)}) \right)$ and $p_{\vec{\eta}, \tilde{\Sigma}_{2|1}} \left((t_1^{(1)}, \vec{y}_1^{(1)}), (t_2^{(1)}, \vec{y}_2^{(1)}) \right)$ are defined as before. $p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left((t_2^{(1)}, \vec{y}_2^{(1)}), (t_3^{(1)}, \vec{y}_3^{(1)}) \right) d\vec{y}_3^{(1)}$ is given as

$$p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left((t_2^{(1)}, \vec{y}_2^{(1)}), (t_3^{(1)}, \vec{y}_3^{(1)}) \right) d\vec{y}_3^{(1)} = (2\pi)^{-\frac{D}{2}} |\tilde{\Sigma}_{3|2}|^{-\frac{1}{2}} \exp \left\{ -\frac{\left(\vec{y}_3^{(1)} - \left[\vec{y}_2^{(1)} + (A_3 - A_2)\vec{\eta} \right] \right)^T \tilde{\Sigma}_{3|2}^{-1} \left(\vec{y}_3^{(1)} - \left[\vec{y}_2^{(1)} + (A_3 - A_2)\vec{\eta} \right] \right)}{2} \right\}$$

J.3 Conditional Power and Conditional Type I Error

J.3.1 look $l + 1$

J.3.2 Look $l + 2$

Assume that the trial didn't cross the boundaries at look $1, \dots, l$. At look l , we observed $\vec{x}_l^{(1)}$. The conditional rejection probability under any $\vec{\delta}$ is given by

$$P \left(\bigcup_{j=l+1}^{K_1} \left\{ \max_i \{W_{ij}\} > e_j^{(1)} \right\} \mid \vec{x}_l^{(1)} \right)$$

where $\vec{x}_l^{(1)} = (x_{1l}^{(1)}, x_{2l}^{(1)}, \dots, x_{Dl}^{(1)})$. Then the above probability is reduced to

$$P \left(\left\{ \max_i \{W_{i,l+1}\} > e_{l+1}^{(1)} \right\} \mid \vec{x}_l^{(1)} \right)$$

$$+ P \left(\left\{ \max_i \{W_{i,l+1}\} \leq e_{l+1}^{(1)} \right\} \cap \left\{ \max_i \{W_{i,l+2}\} > e_{l+2}^{(1)} \right\} \mid \vec{x}_l^{(1)} \right) + \dots \quad (J.5)$$

For computational purpose, we will work on the U process which is defined as follows. Let $T = \max_{i=1, \dots, D} \{I_{iK_1}\} = n_{0K_1} * \max_h \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2 \right]^{-1} \right\}$. Now let $U_{ij} = \frac{W_{ij}}{\sqrt{T}}$. Then the process U_{ij} is a Brownian process with mean $\eta_i \tilde{t}_{ij}$ and variance \tilde{t}_{ij} where

$$\eta_i = \delta_i \sqrt{T}$$

$$\tilde{t}_{ij} = \frac{\left[\frac{\sigma_i^2}{\lambda_i} + \sigma_0^2 \right]^{-1}}{\max_h \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2 \right]^{-1} \right\}} t_j$$

Let b_j be the initial boundaries based on the process U . Then the conditional power can be calculated as follows.

$$P \left(\left\{ \max_i \{U_{i,l+1}\} > b_{l+1} \right\} \mid \vec{y}_l^{(1)} \right)$$

$$+ P \left(\left\{ \max_i \{U_{i,l+1}\} \leq b_{l+1} \right\} \cap \left\{ \max_i \{U_{i,l+2}\} > b_{l+2} \right\} \mid \vec{y}_l^{(1)} \right) + \dots$$

where $\vec{y}_l^{(1)} = \frac{1}{\sqrt{T}} (x_{1l}, x_{2l}, \dots, x_{Dl})^T$.

J.3.1 look $l + 1$

Note that the first probability is as follows

$$P \left(\left\{ \max_i \{U_{i,l+1}\} > b_{l+1} \right\} \mid \vec{y}_l^{(1)} \right)$$

$$= 1 - \int_{-\infty}^{b_{l+1}} \dots \int_{-\infty}^{b_{l+1}} p_{\vec{\eta}, \tilde{\Sigma}_{l+1|l}} \left((t_l^{(1)}, \vec{y}_l^{(1)}), (t_{l+1}^{(1)}, \vec{y}_{l+1}^{(1)}) \right) d\vec{y}_{l+1}^{(1)}$$

where the transition density from J.1 is given by

$$p_{\vec{\eta}, \tilde{\Sigma}_{l+1|l}} \left((t_l^{(1)}, \vec{y}_{j_1}^{(1)}), (t_{l+1}^{(1)}, \vec{y}_{j_2}^{(1)}) \right) = (2\pi)^{-D/2} |\tilde{\Sigma}_{l+1|l}|^{-\frac{1}{2}} \exp \left\{ - \frac{\left(\vec{y}_{l+1}^{(1)} - \left[\vec{y}_l^{(1)} + (A_{j_2} - A_{j_1}) \vec{\eta} \right] \right)^T \left[\tilde{\Sigma}_{l+1|l} \right]^{-1} \left(\vec{y}_{l+1}^{(1)} - \left[\vec{y}_l^{(1)} + (A_{j_2} - A_{j_1}) \vec{\eta} \right] \right)}{2} \right\}$$

where the matrix $\tilde{\Sigma}$ and A are defined as in section 2.

Hence under $\vec{\delta} = 0$, the probability

$$P \left(\left\{ \max_i \{U_{i,l+1}\} > b_{l+1} \right\} \mid \vec{y}_l^{(1)} \right)$$

J Theory-Multi-arm Multi-stage Group Sequential Design

$$= 1 - \int_{-\infty}^{b_{l+1}} \cdots \int_{-\infty}^{b_{l+1}} p_{\tilde{\Sigma}_{l+1|l}} \left((t_l^{(1)}, \vec{y}_l^{(1)}), (t_{l+1}^{(1)}, \vec{y}_{l+1}^{(1)}) \right) d\vec{y}_{l+1}^{(1)}$$

where the transition density is given by

$$p_{\tilde{\Sigma}_{l+1|l}} \left((t_l^{(1)}, \vec{y}_{j_1}^{(1)}), (t_{l+1}^{(1)}, \vec{y}_{j_2}^{(1)}) \right) = (2\pi)^{-D/2} |\tilde{\Sigma}_{l+1|l}|^{-\frac{1}{2}} \exp \left\{ -\frac{(\vec{y}_{l+1}^{(1)} - \vec{y}_l^{(1)})^T [\tilde{\Sigma}_{l+1|l}]^{-1} (\vec{y}_{l+1}^{(1)} - \vec{y}_l^{(1)})}{2} \right\}$$

J.3.2 Look $l + 2$

The second term in J.5 under any $\vec{\delta}$ is

$$P \left(\left\{ \max_i \{W_{i,l+1}\} \leq e_{l+1}^{(1)} \right\} \cap \left\{ \max_i \{W_{i,l+2}\} > e_{l+2}^{(1)} \right\} \mid \vec{x}_l^{(1)} \right)$$

which can be expressed in term of the process U

$$\begin{aligned} & P \left(\left\{ \max_i \{U_{i,l+1}\} \leq b_{l+1} \right\} \cap \left\{ \max_i \{U_{i,l+2}\} > b_{l+2} \right\} \mid \vec{y}_l^{(1)} \right) \\ &= \int_{-\infty}^{b_{l+1}} \cdots \int_{-\infty}^{b_{l+1}} p_{\vec{\eta}, \tilde{\Sigma}_{l+1|l}} \left((t_l^{(1)}, \vec{y}_l^{(1)}), (t_{l+1}^{(1)}, \vec{y}_{l+1}^{(1)}) \right) \\ & \left[1 - \int_{-\infty}^{b_{l+2}} \cdots \int_{-\infty}^{b_{l+2}} p_{\vec{\eta}, \tilde{\Sigma}_{l+2|l+1}} \left((t_{l+1}^{(1)}, \vec{y}_{l+1}^{(1)}), (t_{l+2}^{(1)}, \vec{y}_{l+2}^{(1)}) \right) d\vec{y}_{l+2}^{(1)} \right] d\vec{y}_{l+1}^{(1)} \end{aligned}$$

where

$$p_{\vec{\eta}, \tilde{\Sigma}_{l+1|l}} \left((t_l^{(1)}, \vec{y}_l^{(1)}), (t_{l+1}^{(1)}, \vec{y}_{l+1}^{(1)}) \right) = (2\pi)^{-\frac{D}{2}} |\tilde{\Sigma}_{l+1|l}|^{-\frac{1}{2}} \exp \left\{ -\frac{(\vec{y}_{l+1}^{(1)} - [\vec{y}_l^{(1)} + (A_{l+1} - A_l) \vec{\eta}])^T \tilde{\Sigma}_{l+1|l}^{-1} (\vec{y}_{l+1}^{(1)} - [\vec{y}_l^{(1)} + (A_{l+1} - A_l) \vec{\eta}])}{2} \right\}$$

$$p_{\vec{\eta}, \tilde{\Sigma}_{l+2|l+1}} \left((t_{l+1}^{(1)}, \vec{y}_{l+1}^{(1)}), (t_{l+2}^{(1)}, \vec{y}_{l+2}^{(1)}) \right) = (2\pi)^{-\frac{D}{2}} |\tilde{\Sigma}_{l+2|l+1}|^{-\frac{1}{2}}$$

$$\exp \left\{ - \frac{\left(\vec{y}_{l+2}^{(1)} - \left[\vec{y}_{l+1}^{(1)} + (A_{l+2} - A_{l+1}) \vec{\eta} \right] \right)^T \tilde{\Sigma}_{l+2|l+1}^{-1} \left(\vec{y}_{l+2}^{(1)} - \left[\vec{y}_{l+1}^{(1)} + (A_{l+2} - A_{l+1}) \vec{\eta} \right] \right)}{2} \right\}$$

Under $\vec{\delta} = 0$, we will obtain the conditional type I error by replacing $\vec{\eta}$ by 0.

J.4 Compute power and sample size

J.4.1 Compute power for user-specified sample size

J.4.1 Compute power for user-specified sample size

To compute power for user-specified sample size, we first need to compute boundaries b_j ($j = 1, \dots, K_1$) using the method in Section 2. Once the boundaries have been computed, we can compute power for user-specified sample size. The power is given by

$$\begin{aligned} & P \left(\bigcup_{j=1}^{K_1} \left\{ \max_i \left\{ W_{ij}^{(1)} \right\} > e_j^{(1)} \right\} \mid \vec{\delta} \right) \\ &= P \left(\max_i \left\{ W_{i1}^{(1)} \right\} > e_1^{(1)} \mid \vec{\delta} \right) \\ &+ P \left(\left\{ \max_i \left\{ W_{i1}^{(1)} \right\} \leq b_1^{(1)} \right\} \cap \left\{ \max_i \left\{ W_{i2}^{(1)} \right\} > b_2^{(1)} \right\} \mid \vec{\delta} \right) + \dots \end{aligned}$$

Let N be the total sample size for the study. Assume we want to power the study at some $\vec{\delta} = (\delta_1, \delta_2, \dots, \delta_D)$. To compute the power for a sample size of N , we work with the process U which is defined as $U_{ij} = \frac{W_{ij}^{(1)}}{\sqrt{T}}$ and

$$T = \max_{h=1, \dots, D} \{ I_{hK_1} \} = n_{0K_1} * \max_h \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2 \right]^{-1} \right\}$$

From Section 2, we have

$$P \left(\max_i \left\{ W_{i1}^{(1)} \right\} > e_1^{(1)} \mid \vec{\delta} \right) = P \left(\max_i \left\{ U_{i1} \right\} > b_1 \mid \vec{\delta} \right)$$

$$1 - \int_{-\infty}^{b_1} \dots \int_{-\infty}^{b_1} p_{\vec{\eta}, \tilde{\Sigma}_1|0} \left((0, 0), \left(t_1^{(1)}, \vec{y}_1^{(1)} \right) \right) d\vec{y}_1^{(1)}$$

J Theory-Multi-arm Multi-stage Group Sequential Design

where

$$\begin{aligned}
 & p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left((0, 0), \left(t_1^{(1)}, \vec{y}_1^{(1)} \right) \right) = (2\pi)^{-\frac{D}{2}} |\tilde{\Sigma}_{1|0}|^{-\frac{1}{2}} \\
 & \exp \left\{ -\frac{\left(\vec{y}_1^{(1)} - A_1 \vec{\eta} \right)^T \left(\tilde{\Sigma}_{1|0} \right)^{-1} \left(\vec{y}_1^{(1)} - A_1 \vec{\eta} \right)}{2} \right\} \\
 & P \left(\left\{ \max_i \{ W_{i1}^{(1)} \} \leq e_1^{(1)} \right\} \cap \left\{ \max_i \{ W_{i2}^{(1)} \} > e_2^{(1)} \right\} \mid \vec{\delta} \right) \\
 & = P \left(\left\{ \max_i \{ U_{i1} \} \leq b_1 \right\} \cap \left\{ \max_i \{ U_{i2} \} > b_2 \right\} \mid \vec{\delta} \right) \\
 & = \int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_1} p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left((0, 0), \left(t_1^{(1)}, \vec{y}_1^{(1)} \right) \right) \\
 & \left[1 - \int_{-\infty}^{b_2} \cdots \int_{-\infty}^{b_2} p_{\vec{\eta}, \tilde{\Sigma}_{2|1}} \left(\left(t_1^{(1)}, \vec{y}_1^{(1)} \right), \left(t_2^{(1)}, \vec{y}_2^{(1)} \right) \right) d\vec{y}_2^{(1)} \right] d\vec{y}_1^{(1)}
 \end{aligned}$$

where $\vec{y}_2^{(1)} = \left(y_{12}^{(1)}, \dots, y_{D2}^{(1)} \right)^T$ and

$$\begin{aligned}
 & p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left((0, 0), \left(t_1^{(1)}, \vec{y}_1^{(1)} \right) \right) = (2\pi)^{-\frac{D}{2}} |\tilde{\Sigma}_{1|0}|^{-\frac{1}{2}} \\
 & \exp \left\{ -\frac{\left(\vec{y}_1^{(1)} - A_1 \vec{\eta} \right)^T \tilde{\Sigma}_{1|0}^{-1} \left(\vec{y}_1^{(1)} - A_1 \vec{\eta} \right)}{2} \right\} \\
 & p_{\vec{\eta}, \tilde{\Sigma}_{2|1}} \left(\left(t_1^{(1)}, \vec{y}_1^{(1)} \right), \left(t_2^{(1)}, \vec{y}_2^{(1)} \right) \right) = (2\pi)^{-\frac{D}{2}} |\tilde{\Sigma}_{2|1}|^{-\frac{1}{2}} \\
 & \exp \left\{ -\frac{\left(\vec{y}_2^{(1)} - \left[\vec{y}_1^{(1)} + (A_2 - A_1) \vec{\eta} \right] \right)^T \tilde{\Sigma}_{2|1}^{-1} \left(\vec{y}_2^{(1)} - \left[\vec{y}_1^{(1)} + (A_2 - A_1) \vec{\eta} \right] \right)}{2} \right\}
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 & P \left(\bigcap_{j=1}^2 \left\{ \max_i \{W_{i,j}^{(1)}\} \leq e_j^{(1)} \right\} \cap \left\{ \max_i \{W_{i,3}^{(1)}\} > e_3^{(1)} \right\} \mid \vec{\delta} \right) \\
 &= P \left(\bigcap_{j=1}^2 \left\{ \max_i \{U_{i,j}\} \leq b_j \right\} \cap \left\{ \max_i \{U_{i,3}\} > b_3 \right\} \mid \vec{\delta} \right) \\
 &= \int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_1} p_{\vec{\eta}, \tilde{\Sigma}_{1|0}} \left((0, 0), (t_1^{(1)}, \vec{y}_1^{(1)}) \right) \\
 &\quad \int_{-\infty}^{b_2} \cdots \int_{-\infty}^{b_2} p_{\vec{\eta}, \tilde{\Sigma}_{2|1}} \left((t_1^{(1)}, \vec{y}_1^{(1)}), (t_2^{(1)}, \vec{y}_2^{(1)}) \right) \\
 &\quad \left\{ 1 - \int_{-\infty}^{b_3} \cdots \int_{-\infty}^{b_3} p_{\vec{\eta}, \tilde{\Sigma}_{3|2}} \left((t_2^{(1)}, \vec{y}_2^{(1)}), (t_3^{(1)}, \vec{y}_3^{(1)}) \right) d\vec{y}_3^{(1)} \right\} d\vec{y}_2^{(1)} d\vec{y}_1^{(1)}
 \end{aligned}$$

To compute sample size needed to achieve a target power specified by users, we will need to use bisection search algorithm to find the required sample size for a target power.

J.5 Simulation

Let $n_{i,j}^{(1)}$ be the cumulative sample size up to look j for each group for the primary trial. Let L_1 be the look number at which dose selection occurs. For $\alpha_1 < \alpha_2 < \dots < \alpha_{K_1} = \alpha$ be the cumulative α spent by each interim look, let $e_j^{(1)}$ ($j = 1, 2, \dots, K_1$) be the exiting boundaries for the process $W_{i,j}^{(1)}$ ($i = 1, \dots, D; j = 1, \dots, K_1$) such that

$$P \left(\bigcup_{h=1}^j \left\{ \max_i \{W_{i,h}^{(1)}\} > e_h^{(1)} \right\} \mid \vec{\delta} = 0 \right) = \alpha_j$$

We first generate the incremental data for each dose group and control group at each look. Then we calculate $W_{i,j}^{(1)}$ ($i = 1, \dots, D; j = 1, \dots, L_1$). If there exists such a $j \leq L_1$ such that $\max_i \{W_{i,j}^{(1)}\} > e_j^{(1)}$, then stop the trial. If the trial doesn't cross the boundaries until look L_1 and suppose we observed $W_{i,L_1}^{(1)} = x_{i,L_1}^{(1)}$ ($i = 1, \dots, D$), we

J Theory-Multi-arm Multi-stage Group Sequential Design

will drop those doses with $x_{iL_1}^{(1)} < \gamma$. Let $S \subseteq F \equiv \{1, 2, \dots, D\}$ be the index set of the doses selected and denote the cardinality of S by D^* . Let α' be the conditional type I error at look L_1 based on the initial design.

Suppose K_2 additional looks will be planned after dose selection. Let $n_{ij}^{(2)}$ ($1 \leq j \leq K_2$) be the cumulative sample size by look j after dose selection. Suppose that the remaining sample size for the dropped doses are reallocated to other dose groups according to the same allocation ratios. Then n_{ij} Suppose user specifies how to spend this conditional type I error α' with $\alpha'_1 < \alpha'_2 \dots < \alpha'_{K_2} = \alpha'$. Next we need to compute the new boundaries after dose selection. Let $e_1^{(2)}, e_2^{(2)}, \dots, e_{K_2}^{(2)}$ be the exiting boundaries after dose selection. Then $e_1^{(2)}, e_2^{(2)}, \dots, e_{K_2}^{(2)}$ satisfy the following equation

$$P \left(\bigcup_{j=1}^{K_2} \left\{ \max_{i \in S} \{W_{ij}^{(2)}\} > e_j^{(2)} \right\} \mid x_{iL_1}^{(1)} (i \in S); \vec{\delta} = 0 \right) = \alpha' \quad (J.6)$$

Suppose sample size adaptation is planned at look L_2 with $0 \leq L_2 < K_2$. If $L_2 = 0$, then dose selection and sample size adaptation will be performed at the same look. We will generate the incremental statistics at each look j ($0 < j \leq L_2$) (Note that if $L_2 = 0$, skip this step). If there exists a j such that the boundary $e_j^{(2)}$ is crossed, then the trial stops. If the trial didn't cross any of the boundaries up to look L_2 , we will perform sample size adaptation. We will increase the total sample size by a flat rate, say 50%, for each of the selected doses and control group if

$\nu_{low} < \max_{i \in S} \{W_{iL_2}^{(2)}\} < \nu_{up}$, otherwise keep the total sample size for each selected dose and control group as planned.

Suppose K_3 additional looks will be performed after sample size adaptation. Let $n_{ij}^{(3)}$ ($1 \leq j \leq K_3$) be the cumulative sample size by look j after sample size adaptation. If the sample size is adapted, we will need to recalculate the boundaries such that the conditional type I error at look L_2 , denoted by α'' is preserved. Suppose we observed $W_{iL_2}^{(2)} = x_{iL_2}^{(2)}$ at look L_2 . Suppose user specifies how to spend this conditional type I error α'' with $\alpha''_1 < \alpha''_2 \dots < \alpha''_{K_3} = \alpha''$. Next we need to compute the new boundaries after sample size adaptation. Let $e_1^{(3)}, e_2^{(3)}, \dots, e_{K_3}^{(3)}$ be the exiting

boundaries after adaptations. We need to find $e_1^{(3)}, \dots, e_{K_3}^{(3)}$ such that

$$P \left(\bigcup_{j=1}^{K_3} \left\{ \max_{i \in S} \{W_{ij}^{(3)}\} > e_j^{(3)} \right\} \mid x_{iL_2}^{(2)}; \vec{\delta} = 0 \right) = \alpha'' \quad (J.7)$$

The new boundary $e_1^{(2)}, e_2^{(2)}, \dots, e_{K_2}^{(2)}$ after dose selection and the new boundary $e_1^{(3)}, e_2^{(3)}, \dots, e_{K_2}^{(3)}$ after sample size adaptation can be calculated by recursively solving the equation J.6 and J.7. The technical details for new boundary calculation is described in Section 5.1 and 5.2.

J.5.1 Compute boundaries after dose selection at look L_1

To compute the boundaries $e_1^{(2)}, e_2^{(2)}, \dots, e_{K_2}^{(2)}$, we first need to compute the conditional type I error α' . The conditional type I error for the primary trial at look L_1 is the following probability.

$$P \left(\bigcup_{j=L_1+1}^{K_1} \max_{i \in F} \{W_{ij}^{(1)}\} > e_j^{(1)} \mid x_{iL_1}^{(1)}; \vec{\delta} = 0 \right) \quad (J.8)$$

To compute J.8, we need to work with the process $U^{(1)}$ which is defined as $U_{ij}^{(1)} = \frac{W_{ij}^{(1)}}{\sqrt{T^{(1)}}}$ and $T^{(1)} = \max_{i \in F} \{I_{iK_1}^{(1)}\} = n_{0K_1}^{(1)} * \max_h \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2 \right]^{-1} \right\}$. Let $b_j^{(1)}$ be the boundaries based on the process $U^{(1)}$. Then the equation J.8 is equivalent to

$$P \left(\bigcup_{j=L_1+1}^{K_1} \max_{i \in F} \{U_{ij}^{(1)}\} > b_j^{(1)} \mid U_{iL_1}^{(1)} = y_{iL_1}^{(1)}; \vec{\delta} = 0 \right) = \alpha'$$

where $y_{iL_1}^{(1)} = \frac{x_{iL_1}^{(1)}}{\sqrt{T^{(1)}}}$. The above probability can be obtained by recursively computing the following probability for all $K_1 \geq j > L_1$ and then adding up these probabilities for all j with $K_1 \geq j > L_1$

$$P \left(\bigcap_{h=L_1+1}^{j-1} \left\{ \max_{i \in F} \{U_{ih}^{(1)}\} \leq b_h^{(1)} \right\} \cap \left\{ \max_{i \in F} \{U_{ij}^{(1)}\} > b_j^{(1)} \right\} \mid y_{iL_1}^{(1)}; \vec{\delta} = 0 \right)$$

J Theory-Multi-arm Multi-stage Group Sequential Design

Once we obtain α' , we can calculate the new boundary $e_1^{(2)}, e_2^{(2)}, \dots, e_{K_2}^{(2)}$ which satisfy the following equation.

$$P \left(\bigcup_{j=1}^{K_2} \max_{i \in S} \{W_{ij}^{(2)}\} > e_j^{(2)} \mid x_{iL_1}^{(1)} (i \in S) \right) = \alpha' \tag{J.9}$$

To compute the probability sitting to the left hand side of the equation J.9, we define the process $U^{(2)}$ as $U_{ij}^{(2)} = \frac{W_{ij}^{(2)}}{\sqrt{T^{(2)}}}$ and

$T^{(2)} = \max_{i \in F} \{I_{iK_2}^{(2)}\} = n_{0K_2}^{(2)} * \max_{h \in F} \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2 \right]^{-1} \right\}$. If there is no sample size reallocation after dropping doses, then $n_{0K_2}^{(2)} = n_{0K_1}^{(1)}$. If there is sample size reallocation after dropping doses and assume that we keep the same allocation ratios for the selected doses to control arm, then the sample size allocated to control arm after look L_1 is $\frac{N - n_{0L_1}^{(1)}(1 + \sum_{i \in F} \lambda_i)}{1 + \sum_{i \in S} \lambda_i}$ and $n_{0K_2}^{(2)} = \frac{N - n_{0L_1}^{(1)}(1 + \sum_{i \in F} \lambda_i)}{1 + \sum_{i \in S} \lambda_i} + n_{0L_1}^{(1)}$ and $n_{iK_2}^{(2)} = \lambda_i n_{0K_2}^{(2)}$. Let $U_{i0}^{(2)} = y_{i0}^{(2)} = \frac{x_{iL_1}^{(1)}}{\sqrt{T^{(2)}}}$ where we use subscript 0 to indicate the starting state for the secondary trial. We first compute the boundaries $b_1^{(2)}, b_2^{(2)}, \dots, b_{K_2}^{(2)}$ where $b_j^{(2)} = \frac{e_j^{(2)}}{\sqrt{T^{(2)}}}$ such that

$$P \left(\bigcup_{j=1}^{K_2} \max_{i \in S} \{U_{ij}^{(2)}\} > b_j^{(2)} \mid U_{i0}^{(2)} = y_{i0}^{(2)} (i \in S) \right) = \alpha' \tag{J.10}$$

Note that $U_{ij}^{(2)} = y_{i0}^{(2)} + \frac{1}{\sqrt{T^{(2)}}} \sum_{h=1}^j W_{i(h)}^{(2)}$. Hence $U_{ij}^{(2)}$ is normal distributed with mean $y_{i0}^{(2)} + \frac{\delta_i \sum_{h=1}^j I_{i(h)}^{(2)}}{\sqrt{T^{(2)}}} = y_{i0}^{(2)} + \eta_i^{(2)} \frac{\sum_{h=1}^j I_{i(h)}^{(2)}}{T^{(2)}}$ and variance $\frac{\sum_{h=1}^j I_{i(h)}^{(2)}}{T^{(2)}}$ where

$$\eta_i^{(2)} = \delta_i \sqrt{T^{(2)}}$$

$$I_{i(j)}^{(2)} = \left[\frac{\sigma_i^2}{\lambda_i} + \sigma_0^2 \right]^{-1} n_{0(j)}^{(2)} \text{ where } j = 1, \dots, K_2$$

The covariance between $U_{kj}^{(2)}$ and $U_{lj}^{(2)}$ is

$$\begin{aligned} Cov\left(U_{kj}^{(2)}, U_{lj}^{(2)} \mid y_{i0}^{(2)}\right) &= \frac{1}{T^{(2)}} Cov\left(\sum_{h=1}^j W_{k(h)}^{(2)}, \sum_{h=1}^j W_{l(h)}^{(2)}\right) \\ &= \frac{\left[\left(\frac{\sigma_k^2}{\lambda_k} + \sigma_0^2\right)\left(\frac{\sigma_l^2}{\lambda_l} + \sigma_0^2\right)\right]^{-1} \sigma_0^2 * (n_{0j}^{(2)} - n_{00}^{(2)})}{\max_{h \in F} \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2\right]^{-1} \right\} n_{0K_2}^{(2)}} \end{aligned}$$

where $n_{00}^{(2)} = n_{0L_1}^{(1)}$. We can find $b_j^{(2)}$ recursively by solving the following equation

$$P\left(\bigcap_{h=1}^{j-1} \left\{ \max_{i \in S} \{U_{ih}^{(2)}\} \leq b_h^{(2)} \right\} \cap \left\{ \max_{i \in S} \{U_{ij}^{(2)}\} > b_h^{(2)} \right\} \mid y_{i0}^{(2)} (i \in S)\right) = \alpha'_j$$

Once we obtain $b_j^{(2)}$, we can compute $e_j^{(2)}$ as $e_j^{(2)} = b_j^{(2)} \sqrt{T^{(2)}}$.

J.5.2 Compute boundaries after sample size adaptation at look L_2

To compute the boundaries $e_1^{(3)}, e_2^{(3)}, \dots, e_{K_2}^{(3)}$, we first need to compute the conditional type I error α'' at look L_2 of the secondary trial. The conditional type I error for the secondary trial at look L_2 is the following probability.

$$P\left(\bigcup_{j=L_2+1}^{K_2} \max_{i \in S} \{W_{ij}^{(2)}\} > e_j^{(2)} \mid x_{iL_2}^{(2)} (i \in S); \vec{\delta} = 0\right) \tag{J.11}$$

To compute J.11, we need to work with the process $U_{ij}^{(2)} = \frac{W_{ij}^{(2)}}{\sqrt{T^{(2)}}}$ and

$T^{(2)} = \max_{i \in F} \{I_{iK_2}^{(2)}\} = n_{0K_2}^{(2)} * \max_{h \in F} \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2\right]^{-1} \right\}$. By this step, the

boundaries $b_1^{(2)}, b_2^{(2)}, \dots, b_{K_2}^{(2)}$ have been computed. Then J.11 is equivalent to

$$P\left(\bigcup_{j=L_2+1}^{K_2} \max_{i \in F} \{U_{ij}^{(2)}\} > b_j^{(2)} \mid y_{iL_2}^{(2)}; \vec{\delta} = 0\right)$$

where $y_{iL_2}^{(2)} = \frac{x_{iL_1}^{(2)}}{\sqrt{T^{(2)}}}$. The above probability can be obtained by recursively computing the following probability for all $K_1 \geq j > L_1$ and then adding up these probabilities

J Theory-Multi-arm Multi-stage Group Sequential Design

for all j with $K_2 \geq j > L_2$

$$P \left(\bigcap_{h=L_2+1}^{j-1} \left\{ \max_{i \in F} \{U_{ih}^{(2)}\} \leq b_h^{(2)} \right\} \cap \left\{ \max_{i \in F} \{U_{ij}^{(2)}\} > b_j^{(2)} \right\} \mid y_{iL_2}^{(2)}; \vec{\delta} = 0 \right)$$

Once we obtain α'' , we can calculate the new boundary $e_1^{(3)}, e_2^{(3)}, \dots, e_{K_3}^{(3)}$ which satisfy the following equation.

$$P \left(\bigcup_{j=1}^{K_3} \max_{i \in S} \{W_{ij}^{(3)}\} > e_j^{(3)} \mid x_{i,L_2}^{(2)} (i \in S) \right) = \alpha'' \quad (J.12)$$

To compute the probability sitting to the left hand side of the equation J.12, we define the process $U^{(3)}$ as $U_{ij}^{(3)} = \frac{W_{ij}^{(3)}}{\sqrt{T^{(3)}}}$ and

$$T^{(3)} = \max_{i \in F} \{I_{iK_2}^{(3)}\} = n_{0K_2}^{(3)} * \max_{h \in F} \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2 \right]^{-1} \right\}. \text{ Let } U_{i0}^{(3)} = y_{i0}^{(3)} = \frac{x_{iL_2}^{(2)}}{\sqrt{T^{(3)}}}.$$

We first compute the boundaries $b_1^{(3)}, b_2^{(3)}, \dots, b_{K_3}^{(3)}$ where $b_j^{(3)} = \frac{e_j^{(3)}}{\sqrt{T^{(3)}}}$ such that

$$P \left(\bigcup_{j=1}^{K_3} \max_{i \in S} \{U_{ij}^{(3)}\} > b_j^{(3)} \mid y_{i0}^{(3)} (i \in S) \right) = \alpha''$$

Note that $U_{ij}^{(3)} = y_{i0}^{(3)} + \frac{1}{\sqrt{T^{(3)}}} \sum_{h=1}^j W_{i(h)}^{(3)}$. Hence $U_{ij}^{(3)}$ is normally distributed with mean $y_{i0}^{(3)} + \frac{\delta_i \sum_{h=1}^j I_{i(h)}^{(3)}}{\sqrt{T^{(3)}}} = y_{i0}^{(3)} + \eta_i^{(3)} \frac{\sum_{h=1}^j I_{i(h)}^{(3)}}{T^{(3)}}$ and variance $\frac{\sum_{h=1}^j I_{i(h)}^{(3)}}{T^{(3)}}$ where

$$\eta_i^{(3)} = \delta_i \sqrt{T^{(3)}}$$

$$I_{i(j)}^{(3)} = \left[\frac{\sigma_i^2}{\lambda_i} + \sigma_0^2 \right]^{-1} n_{0(j)}^{(3)}$$

The covariance between $U_{kj}^{(3)}$ and $U_{lj}^{(3)}$ is

$$\begin{aligned} Cov \left(U_{kj}^{(3)}, U_{lj}^{(3)} \mid y_{i0}^{(3)} \right) &= \frac{1}{T^{(3)}} Cov \left(\sum_{h=1}^j W_{k(h)}^{(3)}, \sum_{h=1}^j W_{l(h)}^{(3)} \right) \\ &= \frac{\left[\left(\frac{\sigma_k^2}{\lambda_k} + \sigma_0^2 \right) \left(\frac{\sigma_l^2}{\lambda_l} + \sigma_0^2 \right) \right]^{-1} \sigma_0^2 * \left(n_{0j}^{(3)} - n_{00}^{(3)} \right)}{\max_{h \in F} \left\{ \left[\frac{\sigma_h^2}{\lambda_h} + \sigma_0^2 \right]^{-1} \right\} n_{0K_3}^{(3)}} \end{aligned}$$

where $n_{00}^{(3)} = n_{0L_2}^{(2)}$. We can find $b_j^{(3)}$ recursively by solving the following equation

$$P \left(\bigcap_{h=1}^{j-1} \left\{ \max_{i \in S} \{U_{ih}^{(3)}\} \leq b_h^{(3)} \right\} \cap \left\{ \max_{i \in S} \{U_{ij}^{(3)}\} > b_h^{(3)} \right\} \mid y_{i0}^{(3)} (i \in S) \right) = \alpha_j''$$

Once we obtain $b_j^{(3)}$, we can compute $e_j^{(3)}$ as $e_j^{(3)} = b_j^{(3)} \sqrt{T^{(3)}}$.

K ***Theory - MultiArm Two Stage Designs*** ***Combining p-values***

K.1 Introduction

This appendix describes theory behind two stage multi arm designs which combine p-values. These designs are available for difference of means test and difference of proportions test. We also provide details of various computations used in East for these designs.

K.2 Treatment Effect Scales

K.2.1 Estimated Mean

K.2.2 Estimated Delta

K.2.3 Estimated

Standardized Effect Size

K.2.4 Test Statistic

K.2.5 Conditional Power

K.2.6 Isotonic Mean

K.2.7 Isotonic Delta

K.2.8 Isotonic Standardized Effect Size

K.2.9 Estimated Proportion

K.2.10 Isotonic Proportion

East provides various treatment effect scales for selecting treatments for stage 2 for difference of means as well as difference of proportions tests. This section describes these treatment effect scales. Treatment effect scale is used along with treatment selection rule for selecting treatments for stage 2.

For any treatment effect scale, if a tie or ties are observed then they are broken using following conventions.

1. If responses are generated using dose response curve then the treatment with the lowest dose among tied treatments is selected.
2. If responses are not generated using dose response curve then treatment is selected at random among tied treatments.

For isotonic computations 'Pooled Adjacent Violators Algorithm' (PAVA) proposed by Ayers et. al. (1955) is used.

K.2.1 Estimated Mean

This treatment effect scale is available only for difference of mean test. The estimated mean response for each treatment is used in this treatment effect scale.

K.2.2 Estimated Delta

This treatment effect scale is available for difference of mean as well as difference of proportions test.

For difference of mean test, the estimated δ which is the difference between estimated mean for a particular treatment and estimated mean for control is used in this treatment effect scale.

For difference of proportions test, the estimated δ which is the difference between estimated proportion for a particular treatment and estimated proportion for control is used in this treatment effect scale.

K.2.3 Estimated Standardized Effect Size

This treatment effect scale is available only for difference of means test. It is available only if variance option is equal for t statistic and common standard deviation option is selected for Z statistic.

The estimated δ for each treatment is difference between estimated mean for a particular treatment and estimated mean for control. If test statistic option is Z then common standard deviation provided by user is used in the computations. If test statistic is t then estimated pooled standard deviation across all data is used in the computations.

K.2.4 Test Statistic

This treatment effect scale is available for difference of mean as well as difference of proportions test.

For difference of means test, test statistic (Z or t) considering variance option (equal or unequal) is used for this treatment effect scale.

For difference of proportions test, test statistic Z considering pooled or un-pooled option is used for this treatment effect scale.

K.2.5 Conditional Power

This treatment effect scale is available for difference of mean as well as difference of proportions test.

The computation of conditional power is done under the assumption that only control and specified treatment are carried forward to stage 2. The details of computation of conditional power for each specific treatment as given below.

$w^{(1)}$: Weight for stage 1

$Z^{(1)}$: Incremental statistic for stage 1

RB : Cumulative efficacy boundary on Z scale for stage 2 for right tailed test

LB : Cumulative efficacy boundary on Z scale for stage 2 for left tailed test

p : Raw p value for stage 1

q : Raw p value for stage 2

SN : Standard Normal random variable

$n_t^{(2)}$: Sample size corresponding to the specified treatment in stage 2.

$n_c^{(2)}$: Sample size corresponding to the control in Stage 2.

λ : Allocation ratio for specified treatment as specified in initial allocation.

K Theory - MultiArm Two Stage Designs Combining p-values

$n_t^{(2)}$ and $n_c^{(2)}$ are computed using stage 2 sample size as planned initially and allocation ratio under the assumption that only specified treatment and control are carried to stage 2.

Right Tailed Test

$$CP = P(SN > RC - B) \tag{K.1}$$

Where

$$RC = \frac{RB - w^{(1)}\Phi^{(-1)}(1 - p)}{w^{(2)}} \tag{K.2}$$

Left Tailed Test

$$CP = P(SN < LC - B) \tag{K.3}$$

Where

$$LC = \frac{LB - w^{(1)}\Phi^{(-1)}(p)}{w^{(2)}} \tag{K.4}$$

For Difference of Means Test

$$B = \frac{\hat{\delta}}{D} \tag{K.5}$$

Where,

$$\hat{\delta} = \bar{d}_t^{(1)} - \bar{d}_c^{(1)}$$

If Variance option is equal then

$$D = \sigma \sqrt{\frac{1}{n_t^{(2)}} + \frac{1}{n_c^{(2)}}} \tag{K.6}$$

If Test statistic option is t then

σ : Estimate of Pooled standard deviation for stage 1

If Test statistic option is Z then

σ : Design common standard deviation

If Variance option is un-equal then

$$D = \sqrt{\frac{\sigma_t^2}{n_t^{(2)}} + \frac{\sigma_c^2}{n_c^{(2)}}} \tag{K.7}$$

If test statistic option is t then

σ_t^2 : Estimate of variance for specified treatment based on stage 1

σ_c^2 : Estimate of variance for control based on stage 1

If test statistic option is Z then

σ_t^2 : Design variance for specified treatment

σ_c^2 : Design variance for control

For Difference of Proportions Test

$$B = \frac{\hat{\delta}}{D} \quad (\text{K.8})$$

Where

$$\hat{\delta} = \hat{\pi}_t - \hat{\pi}_c \quad (\text{K.9})$$

Where

$\hat{\pi}_t$: Estimate of Proportion for treatment based on stage 1

$\hat{\pi}_c$: Estimate of Proportion for control based on stage 1

When variance is Un-Pooled

$$D = \sqrt{\frac{\hat{\pi}_t(1 - \hat{\pi}_t)}{n_t^{(2)}} + \frac{\hat{\pi}_c(1 - \hat{\pi}_c)}{n_c^{(2)}}} \quad (\text{K.10})$$

When variance is Pooled

$$D = \sqrt{\bar{\pi}(1 - \bar{\pi}) \left(\frac{1}{n_t^{(2)}} + \frac{1}{n_c^{(2)}} \right)} \quad (\text{K.11})$$

Where $\bar{\pi}$: Estimate of pooled proportion based on stage 1

K.2.6 Isotonic Mean

This treatment effect scale is available only for difference of mean test. Isotonic means are computed after applying PAVA algorithm to estimated means of all treatments and control.

K.2.7 Isotonic Delta

This treatment effect scale is available for difference of mean as well as for difference of proportions test.

First Isotonic means are computed by applying PAVA algorithm to estimated means of all treatments and control. Using these computed isotonic means, the value of isotonic δ for each treatment is computed.

K Theory - MultiArm Two Stage Designs Combining p-values

K.2.8 Isotonic Standardized Effect Size

This treatment effect scale is available for difference of mean test only. It is available only if variance option is equal for t statistic and common standard deviation option is selected for Z statistic.

Isotonic $\frac{\delta}{\sigma}$ values are computed by first computing isotonic δ values for all treatments. If test statistic option is Z then value of σ used is the value of common standard deviation and if test statistic is t then estimated pooled standard deviation across all data is used in the computations.

K.2.9 Estimated Proportion

This treatment effect scale is available only for difference of proportions test. The estimated proportion for each treatment is used in this treatment effect scale.

K.2.10 Isotonic Proportion

This treatment effect scale is available only for difference of proportions test. Isotonic proportions are computed after applying PAVA algorithm to estimated proportions of all treatments and control.

K.3 Combination Method

East uses "Inverse Normal" combination function to combine p values (or adjusted p values) from two stages.

Default values of weights for two stages are computed as follows.

$$w^{(1)} = \sqrt{\frac{n^{(1)}}{n}} \quad (\text{K.12})$$

$$w^{(2)} = \sqrt{\frac{n^{(2)}}{n}} \quad (\text{K.13})$$

Where $n^{(1)}$ and $n^{(2)}$ are the total sample sizes corresponding to stage 1 and stage 2 respectively and n is the total sample size.

East allows the user to change the weights as long as the weights satisfy the following condition

$$w^{(1)} * w^{(1)} + w^{(2)} * w^{(2)} = 1 \quad (\text{K.14})$$

East uses "Inverse Normal" combination function to combine p values (or adjusted p values) from two stages.

Let $p^{(1)}$ and $p^{(2)}$ be p-values (or adjusted p-values) from stage 1 and stage 2 respectively. The combined p value is given by the formula

$$p = 1 - \Phi \left(w^{(1)} \Phi^{-1} \left(1 - p^{(1)} \right) + w^{(2)} \Phi^{-1} \left(1 - p^{(2)} \right) \right) \quad (\text{K.15})$$

K.4 Closed Testing

No elementary hypothesis can be rejected unless all intersection hypotheses which contain that elementary hypothesis are rejected is the closed testing principle of Marcus et. al. (1976). This principle is applied in analysis after both stages.

For multiplicity adjustment, East provides Bonferroni, Sidak, Simes and Dunnett methods. Dunnett method is available only for difference of means test. For details of these methods please see appendix of multiple comparison procedures.

After stage 1, multiplicity adjusted p-values are computed for each intersection hypothesis and then closed testing is used to perform hypothesis test of each individual hypothesis.

After stage 2, multiplicity adjusted p values from both stages are combined for each intersection hypothesis using combination method and then closed testing is used to perform hypothesis test of each individual hypothesis.

K.5 Stopping Boundaries

East allows stopping after stage 1 using efficacy or futility boundaries. Trial is terminated after stage 1 if any of the treatment arms crosses efficacy boundary. Trial is terminated for futility after stage 1 if all treatment arms cross futility boundary.

At the end of stage 2, for efficacy futility design, if no arm has crossed efficacy boundary then trial is declared futile.

For efficacy as well as futility stopping, adjusted p value obtained using combination and closed testing procedures is used.

For futility stopping, user can specify futility boundary in terms of δ for difference of proportions test and in terms of $\frac{\delta}{\sigma}$ for difference of means test.

K Theory - MultiArm Two Stage Designs Combining p-values

K.6 Sample Size Re-estimation

Sample size re-estimation allows the user to increase sample size after stage 1. User can specify a target conditional power which will be used to compute re-estimated sample size. User may also directly specify re-estimated sample size. Sample size reduction is not allowed.

Promising zone approach used in adaptive simulations in East is also used here. If a trial lands in the promising zone then only sample size is re-estimated and used. If a trial lands in unfavorable to favorable zones then sample size is not re-estimated.

The conditional power calculation will be based on the assumption that only the control treatment and the best treatment (according to the treatment effect scale) are carried to the second stage. $Z^{(1)}$: Incremental statistic for stage 1 corresponding to best treatment.

$n_c^{(2)}$: Sample size corresponding to control in stage 2.

$n_t^{(2)}$: Sample size corresponding to the best treatment in Stage 2.

p : Raw p-value for the best treatment at stage 1.

$\lambda_b^{(2)}$: Allocation ratio for the best treatment as specified in treatment selection tab

RBA: Adjusted Cumulative efficacy boundary on Z scale for stage 2 for right tailed test. Adjusted using $\frac{\alpha}{k}$ where α is the design type I error and k is the number of active treatments in stage 1.

LBA: Adjusted Cumulative efficacy boundary on Z scale for stage 2 for left tailed test. Adjusted using $\frac{\alpha}{k}$ where α is the design type I error and k is the number of active treatments in stage 1.

tCP: Target conditional power *SN*: Standard Normal random variable

For right tailed test, the formula for conditional power is given by

$$CP = P(SN > RC - B) = tCP \quad (K.16)$$

Where

$$RC = \frac{RBA - w^{(1)}\Phi^{(-1)}(1 - p)}{w^{(2)}} \quad (K.17)$$

For Left tailed test, the formula for conditional power is given by

$$CP = P(SN < LC - B) = tCP \quad (K.18)$$

Where

$$LC = \frac{LBA - w^{(1)}\Phi^{(-1)}(p)}{w^{(2)}} \quad (K.19)$$

For Difference of Means Test

$$B = \frac{\hat{\delta}}{D} \tag{K.20}$$

Where,

$$\hat{\delta} = \bar{d}_t^{(1)} - \bar{d}_c^{(1)}$$

If Variance option is equal then

$$D = \sigma \sqrt{\frac{1}{n_t^{(2)}} + \frac{1}{n_c^{(2)}}} \tag{K.21}$$

Let us define

$$D_1 = \sigma \sqrt{\frac{1}{\lambda_b^{(2)}} + 1} \tag{K.22}$$

If Test statistic option is t then

σ : Estimate of Pooled standard deviation for stage 1

If Test statistic option is Z then

σ : Design common standard deviation

If Variance option is un-equal then

$$D = \sqrt{\frac{\sigma_t^2}{n_t^{(2)}} + \frac{\sigma_c^2}{n_c^{(2)}}} \tag{K.23}$$

Let us define

$$D_1 = \sqrt{\frac{\sigma_t^2}{\lambda_b^{(2)}} + \sigma_c^2} \tag{K.24}$$

If test statistic option is t then

σ_t^2 : Estimate of variance for specified treatment based on stage 1

σ_c^2 : Estimate of variance for control based on stage 1

If test statistic option is Z then

σ_t^2 : Design variance for specified treatment

σ_c^2 : Design variance for control

K Theory - MultiArm Two Stage Designs Combining p-values

For Difference of Proportions Test

$$B = \frac{\hat{\delta}}{D} \tag{K.25}$$

Where

$$\hat{\delta} = \hat{\pi}_t - \hat{\pi}_c \tag{K.26}$$

Where

$\hat{\pi}_t$: Estimate of Proportion for treatment based on stage 1

$\hat{\pi}_c$: Estimate of Proportion for control based on stage 1

When variance is Un-Pooled

$$D = \sqrt{\frac{\hat{\pi}_t(1 - \hat{\pi}_t)}{n_t^{(2)}} + \frac{\hat{\pi}_c(1 - \hat{\pi}_c)}{n_c^{(2)}}} \tag{K.27}$$

Let us define

$$D_1 = \sqrt{\frac{\hat{\pi}_t(1 - \hat{\pi}_t)}{\lambda_b^{(2)}} + \hat{\pi}_c(1 - \hat{\pi}_c)} \tag{K.28}$$

When variance is Pooled

$$D = \sqrt{\bar{\pi}(1 - \bar{\pi}) \left(\frac{1}{n_t^{(2)}} + \frac{1}{n_c^{(2)}} \right)} \tag{K.29}$$

Let us define

$$D = \sqrt{\bar{\pi}(1 - \bar{\pi}) \left(\frac{1}{\lambda_b^{(2)}} + 1 \right)} \tag{K.30}$$

Where $\bar{\pi}$: Estimate of pooled proportion based on stage 1

Finally the formulae for sample size on control arm on stage 2 are as follows. For Right Tailed Test

$$n_c^{(2)} = (RC - \Phi^{-1}(1 - tCP))^2 * \frac{D_1^2}{\hat{\delta}^2} \tag{K.31}$$

For Left Tailed Test

$$n_c^{(2)} = (LC - \Phi^{-1}(tCP))^2 * \frac{D_1^2}{\hat{\delta}^2} \tag{K.32}$$

Once the re-estimated control sample size $n_c^{(2)}$ is computed then we will consider the allocation ratio specified in the treatment selection tab (for second stage) and compute the sample size for a specific treatment which is carried forward to stage two.

Sample size re-estimation is not performed if estimated delta value after stage 1 is in opposite direction of the rejection type of the test.

L Technical Details - Predicted Interval Plots

Predicted interval plots (PIP) are useful tool in accessing magnitude of future treatment effect and its associated uncertainty given current data. Predicted interval plots are available in regular interim monitoring as well as Muller and Schafer interim monitoring. In this appendix, we describe technical details related to PIP. We have divided the appendix in following four sections.

1. Inputs for PIP
2. Estimation of Parameters from Data
3. Simulating the future for PIP
4. Computing and Displaying PIP

L.1 Inputs for PIP

Below we describe details about inputs required for PIP.

1. **PIP for Look** - This corresponds to the various choices about the future that you want generated in PIP. There are three choices.
 - (a) **PIP at Final Look** - This option corresponds to the final look in Design. With this option, it is assumed that there is only one specified look in the future and future will be generated so that completers or events corresponding to the final look as per design are achieved.
 - (b) **PIP at Any Future Look** - All looks specified in design which have not yet happened in Interim Monitoring Sheet are considered in this option. In the future looks, early stopping is also considered with this option.
 - (c) **PIP at User specified Look** - With this option, it is assumed that there is only one specified look in the future but here user can alter completers or events which correspond to this future look. This option is not available in PIP for Muller and Schafer interim monitoring.
2. **Population ID** - The variable corresponding to the Population ID must be binary which contains only two distinct values. In user interface, you can specify which value corresponds to the control arm and which value corresponds to the treatment arm.
3. **Arrival Time** - The variable corresponding to Arrival Time is required only for Survival end point tests. This variable must be numeric (values strictly greater than zero) representing time of entry into the trial for each subject.
4. **Status Indicator** - The variable corresponding to Status Indicator is mandatory for Survival end point trials. It is required for normal or binomial end point trials if data contains delayed responses. In this variable, value of 1 represents that response has been observed for that subject. Value of -1 represents that the subject has dropped out before giving response. Value 0 represents that the subject is still in the trial but has not responded.

5. **Response Variable** - The variable which corresponds to the Response Variable must be numeric for Normal End Point test. It must be binary which contains only two distinct values for discrete end point test. In user interface, you can specify which value corresponds to the control arm and which value corresponds to the treatment Arm. For Survival End Point tests, this variable must be continuous representing time spent in the trial until event (for subject whose Status Indicator is 1) or until drop out (for subject whose Status Indicator is -1) or total follow up time (for pending subject whose Status Indicator is 0).

L.2 Estimation of Parameters from Data

Estimation of Parameter from current Data is optional. By default design values of parameters are copied whenever possible. Of course after estimating parameters from current data, you can edit their values as you desire. Before estimating parameters from data, invalid observations are ignored from the data. Sample size is calculated as total number of subjects accrued in the trial, based on the current data. Number of completers or events (if they are different than Sample size because of response lag or drop out probability) are computed as subjects whose response is observed. Sample Size and Number of Completers are not editable for regular PIP. Sample size is editable for MS PIP. You should verify that number of completers or events computed from the data match with the number used in Interim Monitoring Sheet. In case of Normal or Binomial end point, if data does not contain delayed responses then all subjects in the data are assumed to be responders and their responses are used for parameter estimation. If the data contains delayed responses then response for subjects who have Status Indicator value equal to 1 are used for parameter estimation. In case of Survival end point, Data Base Lock Time (DBLT) is computed from that data which is the maximum observed calendar time in the data. For subjects whose Status Indicator is 0 i.e. for pending subjects, the Response value must be the difference between DBLT and arrival time for that subject. If this condition is not met for any subject then Response value will be correctly updated for such a subject and used in parameter estimation. Here are the formulae for estimation of various parameters.

1. Difference of Means Test

Let

n_c : Number of responders on control arm.

n_t : Number of responders on treatment arm.

$x_{i,c}$: Response value of i^{th} subject on control arm.

$x_{i,t}$: Response value of i^{th} subject on treatment arm.

Estimate of Mean Control is given by

$$\mu_c = \frac{\sum_{i=1}^{n_c} x_{i,c}}{n_c} \quad (\text{L.1})$$

L Technical Details - Predicted Interval Plots

Estimate of Mean Treatment is given by

$$\mu_t = \frac{\sum_{i=1}^{n_t} x_{i,t}}{n_t} \quad (\text{L.2})$$

Estimate of Difference of Means is given by

$$\delta = \mu_t - \mu_c \quad (\text{L.3})$$

Estimate of Probability of Dropout is given by

$$PD = \frac{\text{No. of DropOuts}}{\text{Sample Size}} \quad (\text{L.4})$$

Estimate of Std. Deviation is given by pooled standard deviation computed from the data.

2. Difference of Proportions and Ratio of Proportions Test

Let

x_c : Number of responses on control arm

y_t : Number of responses on treatment arm

n_c : Number of subjects on control arm

n_t : Number of subjects on treatment arm

Estimate of Proportion under control is given by

$$\pi_c = \frac{(x_c + 0.5)}{(n_c + 1)} \quad (\text{L.5})$$

Estimate of Proportion under treatment is given by

$$\pi_t = \frac{(x_t + 0.5)}{(n_t + 1)} \quad (\text{L.6})$$

Estimate of Difference of Proportions is given by

$$\delta = \pi_t - \pi_c \quad (\text{L.7})$$

Estimate of Probability of Dropout is given by

$$PD = \frac{\text{No. of DropOuts}}{\text{Sample Size}} \quad (\text{L.8})$$

3. Survival (GADAR and GADSD) Tests Let E_c : Number of Events on Control Arm. E_t : Number of Events on Treatment Arm. D_c : Number of Drop outs on Control Arm. D_t : Number of Drop outs on Treatment Arm . FT_c : Total follow up time of all patients on Control Arm. FT_t : Total follow up time of all patients on Treatment Arm.

Estimate of Hazard Rate for Control is given by

$$\lambda_c = \frac{E_c}{FT_c} \quad (\text{L.9})$$

Estimate of Hazard Ratio (HR) is computed from the Cox proportional hazard model.

Estimate of Hazard Rate for Treatment is given by

$$\lambda_t = \lambda_c * HR \quad (\text{L.10})$$

Estimate of Drop out Hazard Rate for Control is given by

$$\gamma_c = \frac{D_c}{FT_c} \quad (\text{L.11})$$

Estimate of Drop out Hazard Rate for Treatment is given by

$$\gamma_t = \frac{D_t}{FT_t} \quad (\text{L.12})$$

L.3 Simulating the future for PIP

For simulating the future for PIP, parameters estimated from data (editable) are used. Other values of parameters required for simulation (like allocation ratio for example) are used from the corresponding design.

For Normal and Binary end point, response value is generated for pending subjects but treatment indicator for them is preserved from current data. Treatment indicator and response values both are generated for future subjects.

For Survival end point also for pending subjects Treatment Indicator is preserved as is in the current data. For generating survival and dropout times for pending subjects, memory less property of exponential distribution is used. Generation of arrival times for future subjects starts after Data Base Lock Time. Generation of survival and drop out times for future subjects is similar to enhanced simulation.

L Technical Details - Predicted Interval Plots

L.4 Computing and Displaying PIP

The current data is fixed in all simulations but generated future data varies across simulations. For each simulation, the cumulative estimate of Delta (HR for survival end point) and associated standard error is computed. For PIP in non-adaptive IM as well as PIP in MS IM, futility boundary if present in design is ignored in all simulation computations. For regular PIP, two sided confidence interval is always computed even for one sided trial design. Boundary computation at a particular future look or any future look is similar to the boundary computation performed in Interim Monitoring sheet. Repeated confidence interval (RCI) is computed for each simulation with computations similar to that of Interim Monitoring sheet.

For MS PIP, one sided RCI is always computed using shift method.

For Efficacy Two Sided and for Efficacy Futility Two Sided designs with type I error α , the confidence coefficient that is used in regular PIP is $100 \times (1 - \alpha) \%$

For Efficacy One Sided and for Efficacy Futility one sided designs with type I error α , the confidence coefficient that is used in regular PIP is $100 \times (1 - 2 \times \alpha) \%$

For Efficacy One Sided and for Efficacy Futility one sided designs with type I error α , the confidence coefficient that is used in MS PIP is $100 \times (1 - \alpha) \%$

For Any Future Look, RCI is computed at stopping look or last look. For Final Look, RCI is computed at final look as per design. RCI's computed for all simulations are sorted on estimated value of Delta (or HR for survival end point) and are displayed on X -axis. On Y-axis estimated values of Delta are plotted. Current (Interim Monitoring) value of confidence interval is displayed by a black horizontal line in the PIP. Color coding is applied which helps in deciding the density of the observed estimated values of Delta (or HR). Read-offs on PIP is a simple matter of computation of counting the number of repeated confidence intervals which satisfy particular condition.

M Enrollment/Events Prediction - Theory

The terms 'Enrollment' and 'Accruals' are used interchangeably in this Appendix chapter. The Predict module in East 6.4 simulates subject enrollment and events in a clinical trial. These simulations are also part of Enrollment/Events simulation at design stage (Chapter 66 and Enrollment/Events simulation at Interim Monitoring stage (Chapter 67). The underlying theory of generating accruals and/or events is same for both the situations. In this Appendix we present the theory and algorithms based on which the arrival times, time to event data (survival data) and drop out times are simulated. Generation of these quantities make the realizations of accruals, events and drop outs possible which are further used in deriving estimates of average accrual duration, average follow up time, average study duration etc which are of much use to the investigator.

13.1 Enrollments Generation

In East 6.4, the subjects are enrolled assuming a Poisson process for the arrivals. In case the arrivals are across a number of sites, the option of Uniform arrivals is also provided. The arrivals are assumed to occur independently of each other. **Exponential Distribution** : In the Poisson process, the inter-arrival times follow an exponential distribution which has a density function as follows:

$$f(x) = \lambda e^{-\lambda x}, x \in [0, \infty)$$

The 'Poisson' option in the Predict module of East generates subject enrollments by randomly sampling successive inter-arrival times from an exponential distribution with parameter λ . The inter-arrival times obtained describe the time difference (in terms of days, months or years depending on the chosen unit of analysis) between the arrivals of consecutive subjects. In East 6.4 the accruals are assumed to occur in a specified period with fixed accrual rate λ . **Input** The primary input for East Predict simulation is the enrollment plan. It specifies for a set of regions/sites the activation periods (the duration over which the site is to be initialized), the accrual rates per site and the maximum number of subjects that may be enrolled in that region/site. The tables below display examples of enrollment plans by region and by site.

| Region ID | Number of Sites | Site Initiation Period | | Accrual Rate/Site | Enrollment Cap |
|-----------|-----------------|------------------------|-----|-------------------|----------------|
| | | Start | End | | |
| Region 1 | 5 | 0 | 0 | 3 | 1000 |
| Region 2 | 5 | 0 | 2 | 4 | 1000 |
| Region 3 | 10 | 2 | 5 | 2 | 1000 |

13 Enrollment/Events Prediction - Theory

| Site ID | Site Initiation Period | | Accrual Rate/Site | Enrollment Cap |
|---------|------------------------|-----|-------------------|----------------|
| | Start | End | End | |
| | | | Start | |
| Site 1 | 0 | 1 | 5 | 1200 |
| Site 2 | 0 | 1 | 5 | 1200 |
| Site 3 | 1 | 2 | 8 | 1200 |

13.2 Enrollment Simulation Algorithm

Suppose the number of accruals to be simulated in every simulation run is N . Let g : # distinct regions in the study

s : Total # Sites in the study s_i : # Sites in Region $i, i = 1, 2, \dots, g$. The algorithm will involve following steps: For every simulation,

13.2.1 Generation of Site Initiation Times

For a multi-center trial, the arrivals could be from different sites which may be grouped into a number of regions. At the beginning of the trial, some of the sites may be unopened which would get opened later. In order to simulate this scenario, East provides the option of specifying an Enrollment Plan (Chapter 67, Enrollments/Events prediction at Design Stage) which stores the information about Site Start time and Site End time for every site. The input can be either region wise or site wise. A region is comprised of many sites. If the input is region wise, then the variables **Site Initiation Period Start, Site Initiation Period End, Accrual Rate per Site** and **Enrollment Cap** are applicable region wise. For all the sites belonging to a region, the same values of the above mentioned variables apply. The site initiation can be anytime between Site Initiation Period Start and Site Initiation Period End. For the unopened sites, the Site Initiation Times are generated as Uniform random numbers between (Start Time, End Time) Generate a Site Initiation Time from Uniform (SIPStart, SIPEnd) as follows: - Generate a random value from Uniform (0,1), say u - Then, $X = \text{SIPStart} + (\text{SIPEnd} - \text{SIPStart}) * u$ X is the generated Site Initiation time random value from Uniform(SIPStart, SIPEnd) At the end of this step we will have the Site Initiation times (SI Times) for all the sites.

13.2.2 Generation of Enrollments

Sort the Sites data in order of the **Region IDs** and then in order of their **SI times**. The

enrolments will start at each site as per the individual site accrual rate. Suppose ‘a’ is the site accrual rate. (i) Poisson Process: Inter-arrival Times Exponential R : random number between (0,1) $R = F(x) = Exp(-ax) \ x = -\ln(R)/a \ c_{ij} = SITime + x \ c_{ij} =$ arrival time for the next subject at the j -th site in the i -th region. (ii) Uniform Process R : random number between (0,1)

$$R = F(x) = (x - Min)/(Max - Min)$$

$$x = Minimum + (Max-Min)R$$

Minimum = SITime Maximum = SITime+ $\frac{1}{a} \ c_{ij} = SITime + x \ c_{ij} =$ arrival time for the next subject at the j -th site in the i -th region.

13.2.3 Generation of Time on Study

There is no generation of response for normal and binomial end point studies in Predict module. Only for survival studies, the ‘Time to event’ data are generated. The generation of Time on Study follows the procedure described below: **Input: Hazard rates specified. Notation** : c_{ij} : survival time to be generated for i -th subject $(\tau_i, \tau_{i+1}]$: i -th interval in which survival information is specified, k : number of hazard pieces τ_i : starting time of i -th hazard piece with $\tau_0 = 0$. λ_i : hazard rate in i -th hazard piece For the l -th subject, generate its survival time as follows. Compute the survival time for this subject using the formula given below.

$$S_l = \tau_{i-1} - \frac{1}{\lambda_{i-1}} \ln \left(1 - v_l \left(1 - e^{-\lambda_{i-1}(\tau_i - \tau_{i-1})} \right) \right)$$

Where u_i and v_l are random numbers between (0,1) .

13.2.4 Generation of Dropout Times

The drop out time generation is on similar lines as that of time on study.

N Dose Escalation - Theory

N.1 The 3 + 3 Design

The 3 + 3 design method for finding the Maximum Tolerated Dose (MTD) in Phase I clinical trials is described in detail in this section. The 3 + 3 is a rule based design method which starts by allocating the lowest dose level to the first cohort and adaptively escalates/de-escalates to the next dose level based on observed number of dose limiting toxicities (DLTs), until either the MTD is obtained or the trial is stopped for excessive toxicity. It requires no modeling of the dose-toxicity curve beyond the classical assumption for cytotoxic drugs that toxicity increases with dose.

There are three different versions of the 3 + 3: $3 + 3^L$, $3 + 3^L$ (modified), and $3 + 3^H$. The $3 + 3^L$ algorithm proceeds as follows:

1. At each dose level, treat 3 patients beginning with dose level 1. Escalate to the next dose level or de-escalate to the previous dose according to the following rules:
 - (a) If 0 of 3 patients have a dose limiting toxicity (DLT), increase dose to next level.
 - (b) If 2 or more patients has a DLT, decrease dose to previous level¹
 - (c) If 1 of 3 patients has a DLT, treat 3 more patients at current dose level.
 - i. If 1 of 6 has DLT, increase to next dose level.
 - ii. If 2 or more of 6 have DLT, decrease to previous level.
 - (d) If a dose has de-escalated to previous level:
 - i. If only 3 had been treated at the previous level, enroll 3 more patients.
 - ii. If 6 have already been treated at the previous level, stop study and declare it the MTD.
2. The maximum tolerated dose (MTD) is defined as the largest dose for which 1 or fewer DLTs occurred.
3. Escalation never occurs to a dose at which 2 or more DLTs have already occurred.

If we have observed 1 DLT out of 6 patients at the current dose: $3 + 3^H$ and $3 + 3^L$ will recommend escalation, $3 + 3^L$ (modified) will declare the current dose as MTD.

If we have observed 2 DLTs out of 6 patients at the current dose: $3 + 3^H$ will declare the current dose as MTD, $3 + 3^L$ and $3 + 3^L$ (modified) will recommend de-escalation

¹ if de-escalation occurs at the first dose level, then the study is discontinued

N.2 The Continual Reassessment Method

N.2.1 Model

N.2.2 Dose Escalation rules

The Continual Reassessment Method(CRM) for finding the Maximum Tolerated Dose (MTD) in Phase I clinical trials is described in detail in this section. The CRM, introduced originally by O’Quigley et al. (1990), assumes a-priori a monotonically increasing single-parameter dose-toxicity curve (DTC) and a desired toxicity rate p_T . The estimated DTC is updated after each patient’s toxicity outcome is observed, so that each patient’s dose level is based on information about how previous patients tolerated the treatment.

N.2.1 Model

Y_j is the binary toxicity outcome observed in the j th patient recruited to the trial, with $Y_j = 1$ denoting a DLT.

d_1, \dots, d_k are the doses

p_1, \dots, p_k are the true unknown probabilities of toxicities for dose levels d_1, \dots, d_k

θ is the unknown parameter specifying the DTC

$\psi(d_i, \theta)$ is the functional form of the DTC, with $Prob(Y_i = 1) = \psi(d_i, \theta)$. There are three different forms of the DTC considered in East:

1. The Power Model

$$\psi(d_i, \theta) = d_i^\theta, \text{ for } \theta > 0$$

2. The Hyperbolic Tangent Model

$$\psi(d_i, \theta) = \left(\frac{\tanh d_i + 1}{2} \right)^\theta, \text{ for } \theta > 0, \text{ and}$$

3. The single-parameter Logistic Model

$$\psi(d_i, \theta) = \frac{e^{c+\theta d_i}}{1 + e^{c+\theta d_i}} \text{ for } \theta > 0, \text{ and } c \text{ fixed}$$

A Bayesian approach is implemented by placing a prior distribution, $\pi(\theta)$ on the model parameter. The adaptive nature of the CRM arises from choosing the dose for the next patient based on the posterior distribution from the currently recruited patients which is

$$\pi(\theta|y_1, \dots, y_n) \propto L(\theta; y_1, \dots, y_r)\pi(\theta),$$

where $L(\theta; y_1, \dots, y_r) = \prod_{j=1}^r \psi(d_i, \theta)^{y_j} (1 - \psi(d_i, \theta))^{1-y_j}$ and r is the number of subjects for which responses are observed.

Prior distributions

N Dose Escalation - Theory

The choice of a prior distribution for the parameter θ depends on the choice of a DTC. In particular

Power and Hyperbolic Tangent Models: θ is a-priori distributed as a Gamma random variable,

$$\pi(\theta) = \frac{\theta^{\alpha-1} \exp(-\theta\beta)\beta^\alpha}{\Gamma(\alpha)}, \text{ for } \theta > 0, \alpha, \beta > 0$$

Single-parameter Logistic Model: θ is a-priori distributed as a log-normal random variable,

$$\pi(\theta) = \frac{\exp\left(-\frac{(\ln\theta-\mu)^2}{2\sigma^2}\right)}{\theta\sigma\sqrt{2\pi}}, \text{ for } \theta > 0, \mu \in \mathcal{R}, \sigma > 0$$

N.2.2 Dose Escalation rules

The dose to be assigned to the next patient, or cohort of patients is the one that has posterior probability of being closest to the target toxicity probability p_T and simultaneously below an upper limit of the toxicity probability denoted by p_{UL} . In particular, the next cohort of patients is assigned to dose $d_i = \operatorname{argmin}_i(\hat{p}_{ir} - p_T)$ where \hat{p}_{ir} is the posterior probability of toxicity after r subject responses.

By default East uses in its dose escalation rules, the modification in the original CRM proposed by Goodman et al. based on which any given dose escalation cannot increase by more than one level, although dose de-escalations can be large. In addition a dose escalation is not allowed if the previous subject experienced a DLT. Both restrictions can be lifted by selecting the corresponding “Dose Skipping Options” in the “Design Parameters” tab.

N.3 The Modified Toxicity Probability Interval Design

N.3.1 Dosing Intervals

N.3.2 Dose Escalation Rules

N.3.3 Computation of the MTD

This section describes the modified Toxicity Probability Intervals (mTPI) proposed by Yuan Ji et al.(2010). The mTPI is a model-based design and it consists of 3 components:

1. Three dosing intervals,
2. a beta/binomial Bayesian model, and
3. a dose-assignment rule based on Unit Probability Mass (UPM).

Following the notation of Section N.2, we let p_1, \dots, p_k denote the toxicity probabilities for doses d_1, \dots, d_k where k is the total number of candidate doses in the trial. The observed data include n_i , the number of patients treated at dose i , and x_i , the

number of patients experiencing a toxicity. The likelihood function for data $\{(x_i, n_i), i = 1, \dots, k\}$ is a product of binomial densities. The estimates of these toxicity probabilities p_i are sequentially updated and are used to decide if some of the doses studied would be close to the true MTD. This is achieved through Bayes' Theorem. Each p_i is a-priori distributed as a Beta-random variable $Beta(\alpha, \beta)$ and a-posteriori is $Beta(\alpha + x_i, \beta + n_i - x_i)$.

N.3.1 Dosing Intervals

The mTPI design employs a simple beta-binomial hierarchic model. Decision rules are based on calculating the unit probability mass (UPM) of three intervals corresponding to underdosing, proper dosing, and overdosing in terms of toxicity. More specifically, the underdosing interval is defined as $(0, p_T - \epsilon_1)$, the overdosing interval as $(p_T + \epsilon_2, 1)$ and the proper dosing interval as $(p_T - \epsilon_1, p_T + \epsilon_2)$ where ϵ_i are small fractions, such as 0.05, to account for the uncertainty around the true target toxicity. These three dosing intervals are associated with three different dose-escalation decisions. The underdosing interval corresponds to a dose escalation (E), overdosing corresponds to a dose de-escalation (D), and proper dosing corresponds to staying at the current dose (S).

N.3.2 Dose Escalation Rules

Given an interval and a probability distribution, the UPM of that interval is defined as the probability of the interval divided by the length of the interval. The mTPI design calculates the UPMs for the three dosing intervals, and the one with the largest UPM implies the corresponding dose-finding decision. That decision provides the dose level to be used for future patients. In particular, the algorithm proceeds as follows:

1. Compute the posterior probability of excessive toxicity at the current tried dose, i.e., $Prob(p_i > p_T | x_i)$ which is a function of the cumulative Beta distribution $Beta(\alpha + x_i, \beta + n_i - x_i)$. Using a threshold for early stopping for safety, ξ , the current and all higher doses are excluded from the trial due to excessive toxicity if $Prob(p_i > p_T | x_i) > \xi$
2. If $Prob(p_i > p_T | x_i) < \xi$ we compute the UPM for each of the three toxicity probability intervals described in section N.3.1 as follows:

(a)

$$UPM(D)_{d_i} = \frac{Prob(p_i > (p_T + \epsilon_2) | x_i)}{1 - (p_T + \epsilon_2)}$$

(b)

$$UPM(S)_{d_i} = \frac{Prob((p_T - \epsilon_1) \leq p_i \leq (p_T + \epsilon_2) | x_i)}{\epsilon_2 - \epsilon_1}$$

N Dose Escalation - Theory

(c)

$$UPM(E)_{d_i} = \frac{Prob(p_i < (p_T - \epsilon_1)|x_i)}{p_T - \epsilon_1}$$

3. Select one of the following actions: E, S or D corresponding to the highest UPM of each toxicity interval provided that the resulting dose level was not excluded in Step 1.
4. If the selected action is 'E' and the current tried dose is the highest dose, then stop the trial. Similarly,
5. if the selected action is 'D' and the current tried dose is the lowest dose, then stop the trial.

N.3.3 Computation of the MTD

Once all the N toxicity responses are observed, we compute the MTD by using all the observed data. To compute the MTD, follow the steps as given below:

1. Using the accumulated information about x_i and n_i for $i = 1, \dots, k$ compute the posterior mean and variance for all the dose levels.
2. Compute isotonic regression estimates of the posterior mean by using the PAVA method with the inverse of the posterior variances of p_i as the weights to obtain isotonically transformed posterior means denoted by say, p_i^* .
3. Among all the tried doses i for which $Prob(p_i > p_T|x_i) < \xi$, select the estimated MTD as the dose with the smallest difference $p_T - p_i^*$.
4. In case of a tie (i.e. two or more doses have the smallest difference),
 - (a) If all the tied doses have the probability of toxicity above the target, select the lower dose as the MTD.
 - (b) Else select the higher dose level as MTD.

N.4 The Bayesian Logistic Regression Model

N.4.1 Prior distribution specification

N.4.2 Dosing Intervals and Selection

N.4.3 Posterior Calculations

This section describes the Bayesian Logistic Regression model as proposed by Neuenschwander et. al. (2009). We follow the notation of Section N.2 and consider a bivariate DTC of the form

$$\psi(d_i, \alpha, \beta) = \frac{e^{\ln \alpha + \beta d_i^*}}{1 + e^{\ln \alpha + \beta d_i^*}} \text{ for } \alpha, \beta > 0, \text{ and } d_i^* = \ln(d_i/d_R), \quad (\text{N.1})$$

where d_R is a reference dose, determined in a way so that $\ln \alpha$ becomes the log-odds of toxicity when $d_i = d_R$.

N.4.1 Prior distribution specification

The vector $\theta = (\ln \alpha, \ln \beta)'$ follows a-priori a bivariate normal distribution with mean vector μ_θ and variance-covariance matrix Σ . Determining the prior distribution parameters can be done in two ways; directly or indirectly.

Direct Prior Elicitation

Using the direct prior elicitation approach involves incorporating information about α and β directly. The parametrization of Equation (N.1) allows for the interpretation of the parameters as follows:

1. $\ln \alpha$ is the log-odds of a toxicity when $d_i = d_R$. As such, the normal distribution of $\ln \alpha$ would represent prior information for this dose. As stated in Neuenschwander et al (2008), if one sets the reference dose d_R to the a-priori anticipated MTD, the mean of $\ln \alpha$ would follow from the target probability p_T and an additional quantile would be needed to obtain the prior standard deviation.
2. For two doses d_i and d_j , the parameter β is the log-odds ratio of a DTL, i.e.,

$$\beta = \frac{\text{logit}(\psi(d_j)) - \text{logit}(\psi(d_i))}{\ln(d_j/d_i)}.$$

As an example, the parameters of the normal distribution of $\ln \beta$ can be obtained by specifying two quantiles for the change in the odds of a DLT if the dose is doubled.

Indirect Prior Elicitation

The indirect prior elicitation approach results in an uninformative prior specification for $\theta = (\ln \alpha, \ln \beta)'$. The following steps are used for this prior distribution specification:

1. Using preclinical data, one can calculate the starting dose and predicted MTD for the study. Median DLT rates are assigned for this two doses, e.g., 0.05 and 0.33 respectively.
2. The remaining doses are assumed to be linear in log-odds in one the $\ln(d/d_R)$ scale and lead to estimated median DLT rates for doses of interest.
3. At each dose level, a minimally informative Beta prior for the probability of a DLT is set and the 2.5% and 97.5% quantiles for each distribution are calculated.
4. The parameters of the bivariate normal distribution of θ are tuned so that the difference between the 2.5% and 97.5% quantiles of this distribution and the targeted values from the Beta distributions is minimized.

N Dose Escalation - Theory

N.4.2 Dosing Intervals and Selection

The probability of a DLT is classified into four categories: underdosing ($c_0 = 0, c_1$], targeted toxicity, (c_1, c_2], excessive toxicity (c_2, c_3] and unacceptable toxicity ($c_3, c_4 = 1$].

Dose selection proceeds with one of the two following methods:

Bayes Risk Minimization

A formal loss function is introduced, quantifying the penalty of ending up in each of the four aforementioned intervals:

$$L((\alpha, \beta), d^*) = \begin{cases} l_1 & \text{if } \text{Prob}((\alpha, \beta | \text{data}, d^*) \in (c_0, c_1]) \\ l_2 & \text{if } \text{Prob}((\alpha, \beta | \text{data}, d^*) \in (c_1, c_2]) \\ l_3 & \text{if } \text{Prob}((\alpha, \beta | \text{data}, d^*) \in (c_2, c_3]) \\ l_4 & \text{if } \text{Prob}((\alpha, \beta | \text{data}, d^*) \in (c_3, c_4]) \end{cases}$$

leading to an estimated Bayes risk of $\sum_{i=1}^4 l_i \{ \text{Prob}((\alpha, \beta | \text{data}, d^*) \in (c_{i-1}, c_i]) \}$. The dose minimizing the Bayes risk is selected as the next dose.

Escalation With Overdose Control (EWOC) Babb et al.(1998) proposed to select the dose for each patient as the one that maximizes the probability of targeted toxicity, i.e., $\text{Prob}((\alpha, \beta | \text{data}, d^*) \in (c_1, c_2])$ subject to the constraint that the probability of overdosing (i.e., excessive and unacceptable toxicity) does not exceed a predefined threshold α_T , say 0.25, called “the feasibility bound”. That is, choose the dose level subject to the constraint $\text{Prob}((\alpha, \beta | \text{data}, d^*) \in (c_2, c_4]) \leq \alpha_T$.

N.4.3 Posterior Calculations

The dose selection process described in Section N.4.2 depends in the calculation of the posterior probability

$$\text{Prob}((\alpha, \beta | \text{data}, d^*) \in (c_{i-1}, c_i]), \tag{N.2}$$

for $i = 1, 2, 3, 4$ which is calculated with respect to

$$\pi(\theta | \mathbf{y}, d_i^*) \propto \frac{e^{\sum_{j=1}^r (\ln \alpha + \beta d_i^*) y_j}}{\prod_{j=1}^r \{1 + e^{(\ln \alpha + \beta d_i^*)}\}} \times \pi(\theta) \tag{N.3}$$

As this bivariate posterior distribution is not a standard known distribution we calculate (N.2) by employing two different sampling-based methods.

Metropolis-Hastings The Metropolis-Hastings algorithm for obtaining samples from (N.3) proceeds as follows:

1. Given a starting value of $\theta = \theta^{(0)}$, generate a candidate value $\theta^* = \theta^{(0)} + \sigma\epsilon$, where $\epsilon \sim N_2(\mathbf{0}, \mathbf{I}_2)$.

2. Calculate

$$\rho = \min \left\{ \frac{\pi(\theta^* | \mathbf{y}, d_i^*)}{\pi(\theta^{(0)} | \mathbf{y}, d_i^*)}, 1 \right\}$$

3. Draw randomly $v \sim Unif(0, 1)$
4. If $v \leq \rho$ then set $\theta^{(1)} = \theta^*$, otherwise retain $\theta^{(1)} = \theta^{(0)}$
5. Repeat steps 1-4 until convergence

Direct sampling The second sampling method from the posterior distribution in (N.3) is a block sampling method. It involves discretizing $\ln \alpha$ and $\ln \beta$ values along with their probability of occurrence. The likelihood of each support point $(\ln \alpha, \ln \beta)$ is computed in this discrete prior. A block of values for $(\ln \alpha, \ln \beta)$ is sampled by first sampling a value $\ln \alpha$ from its discrete marginal distribution and then a value of $\ln \beta$ from the discrete conditional distribution of $\ln \beta | \ln \alpha$ using the inverse cumulative distribution method.

N.5 Bayesian Logistic Regression Model for Combination of Two Agents

This section describes the BLRM design for a combination of two active agents.

Prior Distribution

The vector $\theta = \begin{pmatrix} \log \alpha \\ \log \beta \end{pmatrix}$ of model parameters for each active agent a priori follows a bivariate normal distribution as follows:

$$\theta = \begin{pmatrix} \log \alpha \\ \log \beta \end{pmatrix} \sim BVN \left[\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{bmatrix} \right]$$

$$\sigma_{\alpha\beta} = \rho\sigma_\alpha\sigma_\beta$$

Where μ_α refers to prior mean of $\log \alpha$, μ_β refers to prior mean of $\log \beta$, σ_α refers to prior SD for $\log \alpha$, σ_β refers to prior SD for $\log \beta$ and ρ refers to correlation between $\log \alpha$ and $\log \beta$. The interaction parameter (η) a priori follows a normal distribution as follows:

$$\eta \sim N(\mu_\eta, \sigma_\eta^2)$$

N Dose Escalation - Theory

where μ_η denotes the prior mean of η and σ_η denotes the prior SD of η .

Model Definition

The proposed model has the following properties:

- (a) It has three components which stands for
 - Single-agent 1 toxicity, represented by parameters α_1, β_1
 - Single-agent 2 toxicity, represented by parameters α_2, β_2
 - Interaction, represented by parameter η .
- (b) If one of the doses is 0, $d_2 = 0$, say, the model should simplify to the single-agent model with parameters α_1, β_1 .

Single-agent probabilities of DLT:

Probability of DLT, given Agent1: π_{d_1}

Probability of DLT, given Agent2: π_{d_2} .

π_{d_1} and π_{d_2} are vectors of Probability of DLT at each dose of Agent 1 and Agent 2 respectively. In the special case of no interaction the single-agent parameters fully determine the risk of a DLT. For dose combination (d_1, d_2) a patient's probability to have no DLT is $(1 - \pi_{d_1})(1 - \pi_{d_2})$. Hence, Probability of DLT under no interaction is

$$\pi_{d_1, d_2}^0 = 1 - (1 - \pi_{d_1})(1 - \pi_{d_2}) = \pi_{d_1} + \pi_{d_2} - \pi_{d_1} \pi_{d_2}$$

On the odds scale this is equivalent to

$$odds_{d_1, d_2}^0 = odds_{d_1} + odds_{d_2} + odds_{d_1} odds_{d_2}$$

Interaction parameter (η) has the interpretation of an odds-multiplier, as follows:

$$odds_{d_1, d_2} = odds_{d_1, d_2}^0 \times g(\eta, d_1, d_2)$$

The odds-multiplier g should fulfill the constraints $g(\eta, 0, d_2) = g(\eta, d_1, 0) = 1$, since if one of the doses is 0, it should result in the single-agent odds. Hence, $g(\eta, d_1, d_2)$ is defined as $g(\eta, d_1, d_2) = \exp(\eta, d_1, d_2)$.

We will use same interaction for all dose combination and hence we can simply use $\exp(\eta)$. Typically $\eta > 0$, but not necessarily

$\eta = 0$: No interaction, the drug combination produces a toxic effect whose magnitude is **equal** to that obtained if the drugs act independently in the body.

$\eta < 0$: Protective, the drug combination produces a toxic effect whose magnitude is **less** than that obtained if the drugs act independently in the body.

$\eta > 0$: Synergistic, the drug combination produces a toxic effect whose magnitude is **greater** than that obtained if the drugs act independently in the body.

Likelihood

$$L(\theta_1, \theta_2, \eta | d_1, d_2, y) = \prod_{p=1}^S \left[(\pi_{d_{1p}, d_{2p}})^{y_p} \times (1 - \pi_{d_{1p}, d_{2p}})^{1-y_p} \right]$$

S = Observed sample size

d_{1p} = Dose of Agent1 assigned to patient p

d_{2p} = Dose of Agent2 assigned to patient p

$\pi_{d_{1p}, d_{2p}}$ = Probability of DLT with interaction (η) for patient p

y_p = Binary response (0 or 1) of patient p .

Prior Distribution for Two Parameter Logistic Model

$$\text{Prior distribution } \pi(\theta_i) \propto \frac{1}{\sigma_{\alpha_i} \sigma_{\beta_i} \sqrt{1 - \rho_i^2}} e^{\frac{-z_i}{2(1 - \rho_i^2)}}$$

$$Z = \frac{(\log \alpha_i - \mu_{\alpha_i})^2}{\sigma_{\alpha_i}^2} + \frac{(\log \beta_i - \mu_{\beta_i})^2}{\sigma_{\beta_i}^2} - \frac{2\rho_i(\log \alpha_i - \mu_{\alpha_i})(\log \beta_i - \mu_{\beta_i})}{\sigma_{\alpha_i} \sigma_{\beta_i}}$$

$i = 1$ for Agent 1 and $i = 2$ for Agent 2.

Prior distribution for interaction parameter η .

$$\text{Prior distribution } \pi(\eta) \propto \frac{1}{\sigma_{\eta}} e^{-Z} \text{ where } Z = \frac{(\eta - \mu_{\eta})^2}{2\sigma_{\eta}^2}.$$

Posterior Distribution

$$\pi(\theta_1, \theta_2, \eta | y) \propto L(\theta_1, \theta_2, \eta | d_1, d_2, y) \times \pi(\theta_1) \times \pi(\theta_2) \times \pi(\eta).$$

N Dose Escalation - Theory

Posterior Sampling Method : Metropolis Hastings

Step 1 : Initialize $\theta_1 = (\log \alpha_1^0, \log \beta_1^0), \theta_2 = (\log \alpha_2^0, \log \eta_2^0), \eta = \eta^0$ and $Sim = 1$.

Step 2 : Generate a new candidate for Agent1, $\theta_1^* = \theta_1 + RW\sigma_1 * \epsilon_1$ where $\epsilon_1 \sim BVN(0, 1)$.

Step 3 : Calculate ratio $R_1 = \min \left(\frac{\pi(\theta_1^*, \theta_2, \eta | y)}{\pi(\theta_1, \theta_2, \eta | y)}, 1 \right)$.

Step 4 : Draw a random number $v_1 \sim U(0, 1)$ and if $v_1 < R_1$ then accept the new candidate θ_1^* and set $\theta_1 = \theta_1^*$.

Step 5 : Generate a new candidate for Agent2, $\theta_2^* = \theta_2 + RW\sigma_2 * \epsilon_2$ where $\epsilon_2 \sim BVN(0, 1)$.

Step 6 : Calculate ratio $R_2 = \min \left(\frac{\pi(\theta_1, \theta_2^*, \eta | y)}{\pi(\theta_1, \theta_2, \eta | y)}, 1 \right)$.

Step 7 : Draw a random number $v_2 \sim U(0, 1)$ and if $v_2 < R_2$ then accept the new candidate θ_2^* and set $\theta_2 = \theta_2^*$.

Step 8 : Generate a new candidate for interaction, $\eta^* = \eta + RW\sigma_\eta * \epsilon_3$ where $\epsilon_3 \leftarrow BVN(0, 1)$.

Step 9 : Calculate ratio $R_3 = \min \left(\frac{\pi(\theta_1, \theta_2, \eta^* | y)}{\pi(\theta_1, \theta_2, \eta | y)}, 1 \right)$.

Step 10 : Draw a random number $v_3 \leftarrow U(0, 1)$ and if $v_3 < R_3$ then accept the new candidate η^* and set $\eta = \eta^*$.

Step 11 : Store the value in parameter θ_1, θ_2 and η for simulation Sim .

Step 12 : Go to next simulation, $Sim = Sim + 1$. If $Sim > Sim_{MH} + Burnin_{MH}$ then Stop else Go to Step 2.

Dose Finding Method

1. Compute posterior samples using the Metropolis Hastings method.
2. Compute posterior probability of DLT for every dose pair using steady state samples and Model definition for Combination of Two Agents.
3. Compute probability of being in each toxicity interval for every dose pair as follows,
 - (a) Count number of steady state simulations for which posterior probability of DLT lies within each interval
 - (b) Divide the count for each interval by number of steady state simulations.
4. Exclude the dose pairs which doesn't satisfy the following EWOC principle, Probability of being in overdosing interval $< EWOC$ threshold.
5. If all dose pairs are excluded then stop the trial due to overdosing else go to next step
6. If user has selected any stopping rule(s) to determine MTD early in the trial then check the rules as follows,
 - (a) Consider only the dose pairs which are not excluded due to overdosing
 - (b) Select the dose pair which has maximum probability in the target interval and minimum probability in the overdosing interval. In case the ties still exist, then select the largest of dose pairs based on the dose indices. (See the Note below for this change)
 - (c) Min SS Rule: Check whether the total number of subjects observed in the trial is \geq user specified threshold
 - (d) Allocation Rule: Check whether the total number of subjects observed on the selected dose pair is \geq user specified threshold
 - (e) Target Rule: Check whether the probability of being in targeted toxicity interval for the selected dose pair is \geq user specified threshold.
7. Stop the trial if MTD is determined in the Step6 else go to next step
8. Compute the next dose pair to be assigned to the next group of subjects as follows,
 - (a) Consider only the dose pairs which are not excluded due to overdosing
 - (b) Filter the dose pairs which satisfies the selected Dose Skipping Option and the requirement of whether to increase dose of both agents at the same time
 - (c) Select the highest dose pair which has maximum probability of being in targeted toxicity interval as the next dose.
9. Compute MTD for the final analysis as follows,
 - (a) Consider only the tried dose pairs which are not excluded due to overdosing

N Dose Escalation - Theory

- (b) Select the highest dose pair which has maximum probability of being in targeted toxicity interval as MTD.

N.6 The Product of Independent Beta Probabilities Escalation Design

The Product of Independent Beta Probabilities Escalation design is a Bayesian dose finding method for a combination therapy with two active agents. This method allows for the specification of prior risk of toxicity for all dose combinations and uses posterior probabilities from all proposed dose combinations for dose escalation. The aim is to design a dual agent dose escalation trial targeting a MTD contour such that the risk of toxicity for all dose combinations on this contour is the pre-specified target toxicity level p_T .

Prior and Posterior Distributions

Let $d_i A$ denote the i -th dose level of drug A and $d_j B$ denote the j -th dose level of drug B where doses increase with i and j and $i = 1, \dots, I$ and $j = 1, \dots, J$. We assume that the probabilities of toxicity at every dose combination follow an independent Beta distribution i.e. $\pi_{ij} | a_{ij}, b_{ij} \rightarrow \text{Beta}(a_{ij}, b_{ij}) \forall i, j$. Prior distribution can be specified in two formats:

1. Prior median of P(DLT) π_{ij} and prior sample size SS_{ij} for each dose combination d_{ij} .
2. Prior parameters a_{ij} and b_{ij} of the Beta distribution for each dose combination d_{ij} .

If the prior is specified in format (a), it is internally converted into the format (b) by the software. Suppose $Y^{(m)} = \{r_{ij}^{(m)}, n_{ij}^{(m)}, i = 1, \dots, I, j = 1, \dots, J\}$: Data up to the end of m^{th} cohort. Such that we have observed $r_{ij}^{(m)}$ DLTs from $n_{ij}^{(m)}$ patients for the dose combination d_{ij} . Then because of conjugacy and prior independence of the π_{ij} , the posterior distribution of π_{ij} is also a Beta distribution given by

$$(\pi_{ij} | Y^{(m)} a_{ij}, b_{ij}) \leftarrow \text{Beta}(a_{ij} + r_{ij}^{(m)}, b_{ij} + n_{ij}^{(m)} - r_{ij}^{(m)}) \forall i, j.$$

We assume that the toxicity risk increases with increasing dose, i.e.

$$\begin{aligned} \pi_{ij} &< \pi_{(i+1)j}, I = 1, \dots, I-1, \forall j \text{ and} \\ \pi_{ij} &< \pi_{i(j+1)}, j = 1, \dots, J-1, \forall i, j = 1, \dots, J-1. \end{aligned}$$

Maximum Tolerated Contour

The Maximum Tolerated Contour is formed by the dose combinations that have a posterior mean of DLT rate equal to the targeted toxicity risk. The PIPE design method targets the MTC corresponding to the pre-specified target probability of toxicity (p_T) to recommend the dose level for the next cohorts. Let us denote this MTC as MTC_θ . It is estimated by the line partitioning the dose combination space into toxicity risks above θ or less than θ . MTC_θ must be such that it does not contradict the assumption of monotonicity.

Dose Escalation Rules

For dose escalation, the PIPE method begins by identifying the set of admissible dose combination based on one of the following three criteria: adjacent to MTC_θ , closest to MTC_θ , and those lying in the interval of fixed pre-specified length around the targeted toxicity probability p_T . “Adjacent” doses are the dose levels that lie adjacent to the current estimated MTC_θ . The “closest” doses are defined as those adjacent doses below/above the contour that cannot move up (for below) or down (for above) by one dose level without crossing the contour. Hence “closest” doses must be a subset of adjacent doses. The “Interval” criteria picks up the dose levels having probability of DLT within the interval $(p_T - \epsilon, p_T + \epsilon)$ where ϵ is the pre-specified margin.

Dose Skipping Rules

Dose skipping during escalation can be achieved by using one of the following criteria: Neighborhood constraint, non- neighborhood constraint. Under the neighborhood constraint, the admissible doses for the next cohort further reduces to a set of doses that are a maximum of one dose level higher or lower than the current experimented dose, both for agents A and B. Hence any dose combination can be chosen up to one dose level above or below current drug A and drug B levels including the current dose combination. Under the non-neighborhood constraint, all the previous doses administered are considered, and to allow dose skipping, the constraint allows any dose that is a single-dose level higher in both agents A and B than any previously administered dose combination. The option related to diagonal dose escalation allows escalating levels of both agents at the same time.

Dose Selection

The dose combination for the next cohort is selected from the admissible dose set. This can be done in two possible ways. One is to select the next dose to be the admissible dose with the smallest current sample size. Here sample size is defined as the sum of the prior sample size and the sample size observed in the trial.

N Dose Escalation - Theory

That is, we select a dose d_{ij} where $(i, j) = \arg \min_{\xi \in \Omega^{(m)}} S_{\xi}^{(m)}$ where

$S_{ij}^{(m)} = n_{ij}^{(m)} + a_{ij} + b_{ij}$. The other possible dose selection method is based on a weighted randomization, where the selection of the admissible doses is weighted by the inverse of their sample size.

$P(\text{cohort } m \text{ is allocated } d_{ij} | (i, j) \in \Omega^{(m)}) = \frac{S_{ij}^{-1(m)}}{\sum_{\xi \in \Omega^{(m)}} S_{\xi}^{-1(m)}}$ and the dose combination with the highest probability is chosen.

At the end of the trial, the MTD is selected as the dose closest to the estimated MTC_{θ} from below.

O *R Functions*

O.1 Introduction

For [East6](#), in simulation module we will provide the user the opportunity to perform various tasks using R. In this chapter, we list all tasks for which R functions can be used. We will provide syntax and suggested format for various functions. We have divided functions in various categories.

1. Function for initialization
2. Functions for data generation
3. Functions for test statistic and perform test computations
4. Functions for performing basic simulations
5. Functions for re-estimating sample size in adaptive simulations
6. Function for selecting treatment in multi-arm combining p-values design

O.2 Initialization Function

This function will be optional. If provided, this function will be executed before executing any of the other user defined functions. User can use this function for various reasons. Below we list some of these.

1. Setting seed for R environment
2. Setting working directory for R
3. Initializing global variables.

For more details of uses of this function please see section [O.12](#).

The following table provides details about Initialization function.

O R Functions

Table O.1: Initialization Function

| Suggested Name of the function | <i>Init()</i> | | | | |
|---------------------------------------|--|----------|-------------|------|--|
| Description | Performs Initialization for all simulations | | | | |
| Syntax | <i>Init(Seed)</i> | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>Seed</td> <td>Seed to be set at the beginning of all simulations</td> </tr> </tbody> </table> | Argument | Description | Seed | Seed to be set at the beginning of all simulations |
| Argument | Description | | | | |
| Seed | Seed to be set at the beginning of all simulations | | | | |
| Return Value Type | Integer (Optional). This function may return Error Code (optional) | | | | |
| Suggested format | <pre>Init ← function(Seed) { Error = 0 set.seed(Seed) # User may use other options in set.seed like setting the random # number generator. User may also initialize global variables or # set up the working directory etc. # Do the error handling. Modify Error appropriately return (as.integer(Error)) }</pre> | | | | |

0.3 Data Generation Functions

0.3.1 Generating Arrival Times

0.3.2 Generating Censor indicator

0.3.3 Generating Dropout Times

0.3.4 Randomizing Subjects to Treatments

0.3.5 Randomizing Subjects to Groups

0.3.6 Randomizing Subjects to Populations

Following points are applicable to all functions used for Data Generation described in this document.

1. This document provides suggested name for each function.
2. Argument names and Argument Type for each function are compulsory but the order of the arguments is not. Input argument names are case sensitive.
3. User can have additional input arguments in the function but he must make sure that appropriate values will be available for those additional arguments during function call. For details please see section [0.13](#).
4. Function will return a list. The Identifier Names (Case Insensitive) and Type (we strongly advice that user should type cast the output elements) mentioned for outputs in a list for a particular function are compulsory while their order in the list is not. User can have additional outputs in the list. If user wants to print the arrays (Same size as number of subjects) in the Simulation CSV file then he has to provide identifier for those arrays. These identifiers will be the columns names in output. Any repeated identifiers (column names) will be ignored.
5. We suggest that the return List contain an identifier "ErrorCode". If specified, it has to be of Type Integer. Its values are classified as follows.

0: No Error

Positive integer: Non Fatal Error - Particular Simulation will be aborted but Next Simulation will be performed.

Negative Integer: Fatal Error - No further simulation will be attempted.

We suggest that user should classify error in these categories depending on the context.

O R Functions

0.3.1 Generating Arrival Times

Table O.2: Function for Generating Arrival Times

| Suggested Name of the function | <i>GenArrTimes()</i> | | | | | | | | | | | | |
|---------------------------------------|--|---------|--|---------------------|-------------|--------|--------------------|-------------------------------------|---------------------------|----------|---|----------|--|
| Description | Generates arrival times for a specified number of subjects. Start time and accrual rate (one per period) for each period is provided. | | | | | | | | | | | | |
| Syntax | <i>GenArrTimes(NumSub, NumPrd, PrdStart, AccrRate)</i> | | | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Compulsory Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>NumSub</td> <td>Number of Subjects</td> </tr> <tr> <td>NumPrd</td> <td>Number of Accrual Periods</td> </tr> <tr> <td>PrdStart</td> <td>Array of start times of specified periods</td> </tr> <tr> <td>AccrRate</td> <td>Array of accrual rates (one rate per period)</td> </tr> </tbody> </table> | | | Compulsory Argument | Description | NumSub | Number of Subjects | NumPrd | Number of Accrual Periods | PrdStart | Array of start times of specified periods | AccrRate | Array of accrual rates (one rate per period) |
| Compulsory Argument | Description | | | | | | | | | | | | |
| NumSub | Number of Subjects | | | | | | | | | | | | |
| NumPrd | Number of Accrual Periods | | | | | | | | | | | | |
| PrdStart | Array of start times of specified periods | | | | | | | | | | | | |
| AccrRate | Array of accrual rates (one rate per period) | | | | | | | | | | | | |
| Return Value Type | R List The must identifiers in this list are <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>ArrivalTime</td> <td>An array of generated arrival times</td> <td>Double.</td> </tr> </tbody> </table> | | | Identifier | Description | Type | ArrivalTime | An array of generated arrival times | Double. | | | | |
| Identifier | Description | Type | | | | | | | | | | | |
| ArrivalTime | An array of generated arrival times | Double. | | | | | | | | | | | |
| Suggested format | <pre>GenArrTimes ← function(NumSub, NumPrd, PrdStart, AccrRate) { Error = 0 # Write the actual code here. # Store the generated accrual times in an array called retval. # Use appropriate error handling and modify the # Error appropriately. return(list(ArrivalTime = as.double(retval), ErrorCode = as.integer(Error))) }</pre> | | | | | | | | | | | | |

0.3.2 Generating Censor indicator

Table O.3: Generating Censor Indicator (Normal and Binary)

| Suggested Name of the function | <i>GenCensorInd()</i> | | | | | | | | |
|---------------------------------------|---|---------|--|---------------------|-------------|--------|--------------------|--|-------------------------|
| Description | Generates Censor Indicator (Subject has dropped out (0) or not (1)) for a specified number of subjects. | | | | | | | | |
| Syntax | <i>GenCensorInd (NumSub, ProbDrop)</i> | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Compulsory Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>NumSub</td> <td>Number of Subjects</td> </tr> <tr> <td>ProbDrop</td> <td>Probability of Drop out</td> </tr> </tbody> </table> | | | Compulsory Argument | Description | NumSub | Number of Subjects | ProbDrop | Probability of Drop out |
| Compulsory Argument | Description | | | | | | | | |
| NumSub | Number of Subjects | | | | | | | | |
| ProbDrop | Probability of Drop out | | | | | | | | |
| Return Value Type | R List The must identifiers in this list are <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>CensorInd</td> <td>An array of censor indicator values 0 (Drop out) and 1(No Drop out)</td> <td>Integer</td> </tr> </tbody> </table> | | | Identifier | Description | Type | CensorInd | An array of censor indicator values 0 (Drop out) and 1(No Drop out) | Integer |
| Identifier | Description | Type | | | | | | | |
| CensorInd | An array of censor indicator values 0 (Drop out) and 1(No Drop out) | Integer | | | | | | | |
| Suggested format | <pre> GenCensorInd ← function(function(NumSub, ProbDrop) { Error = 0 # Write the actual code here. # Store the generated censor indicator values in an # array called retval. # Use appropriate error handling and modify the # Error appropriately. return(list(CensorInd = as.integer(retval), ErrorCode = as.integer(Error))) } </pre> | | | | | | | | |

O R Functions

0.3.3 Generating Dropout Times

Table O.4: Generating Dropout Times (Survival)

| Suggested Name of the function | <i>GenDropTimes()</i> | | | | | | | | | | | | | |
|--|---|--|----------|-------------|--------|--------------------|--------|---|-------------|--|------------|--|--------|---------------------------|
| Description | Generates dropout times for a specified number of subjects for survival end point. | | | | | | | | | | | | | |
| Syntax for one arm test | <i>GenDropTimes (NumSub, DropMethod, NumPrd, PrdTime, DropParam)</i> | | | | | | | | | | | | | |
| Syntax for more than one arm test | <i>GenDropTimes (NumSub, NumArm, TreatmentID, DropMethod, NumPrd, PrdTime, DropParam)</i> | | | | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>NumSub</td> <td>Number of Subjects</td> </tr> <tr> <td>NumArm</td> <td>Number of Arms in the trial (including placebo/control)</td> </tr> <tr> <td>TreatmentID</td> <td>Array specifying indexes of arms to which subjects are allocated (one arm index per subject) Index for placebo / control is 0. For other arms, indexes are consecutive positive numbers starting with 1. Thus if the trial has 4 arms (1 placebo + 3 treatment arms), arm indexes will be 0, 1, 2 and 3.</td> </tr> <tr> <td>DropMethod</td> <td>Input method for specifying dropout parameters. 1 - Hazard rates 2 - Probability of Dropouts</td> </tr> <tr> <td>NumPrd</td> <td>Number of dropout periods</td> </tr> </tbody> </table> | | Argument | Description | NumSub | Number of Subjects | NumArm | Number of Arms in the trial (including placebo/control) | TreatmentID | Array specifying indexes of arms to which subjects are allocated (one arm index per subject) Index for placebo / control is 0. For other arms, indexes are consecutive positive numbers starting with 1. Thus if the trial has 4 arms (1 placebo + 3 treatment arms), arm indexes will be 0, 1, 2 and 3. | DropMethod | Input method for specifying dropout parameters. 1 - Hazard rates 2 - Probability of Dropouts | NumPrd | Number of dropout periods |
| Argument | Description | | | | | | | | | | | | | |
| NumSub | Number of Subjects | | | | | | | | | | | | | |
| NumArm | Number of Arms in the trial (including placebo/control) | | | | | | | | | | | | | |
| TreatmentID | Array specifying indexes of arms to which subjects are allocated (one arm index per subject) Index for placebo / control is 0. For other arms, indexes are consecutive positive numbers starting with 1. Thus if the trial has 4 arms (1 placebo + 3 treatment arms), arm indexes will be 0, 1, 2 and 3. | | | | | | | | | | | | | |
| DropMethod | Input method for specifying dropout parameters. 1 - Hazard rates 2 - Probability of Dropouts | | | | | | | | | | | | | |
| NumPrd | Number of dropout periods | | | | | | | | | | | | | |

Table O.5: Generating Dropout Times (Survival) (Continued)

| Arguments | Argument | Description | | | | | | | |
|--------------------------|---|--|--|------------|-------------|------|-------------|-------------------------------------|--------|
| | PrdTime | <p>Array of times used to specify dropout parameters.</p> <p>If DropMethod is 1, then this array specifies the starting times of dropout periods.</p> <p>If DropMethod is 2, then this array specifies the times at which the probabilities of dropout are specified.</p> | | | | | | | |
| | DropParam | <p>2-D array of parameters uses to generate dropout times.</p> <p>Number of rows = Number of Dropout Periods</p> <p>Number of Columns = Number of Arms including Control/Placebo</p> <p>If DropMethod is 1, the DropParam array specifies arm by arm hazard rates (one rate per arm per piece). Thus DropParam [i, j] specifies hazard rate in ith piece for jth arm.</p> <p>If DropMethod is 2, the DropParams array specifies arm by arm probabilities of dropout (one value of probability of dropout per arm per piece). Thus DropParams [i, j] specifies probability of dropout in ith piece for jth arm.</p> | | | | | | | |
| Return Value Type | <p>R List</p> <p>The must identifiers in this list are</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 25%;">Identifier</th> <th style="width: 50%;">Description</th> <th style="width: 25%;">Type</th> </tr> </thead> <tbody> <tr> <td>DropOutTime</td> <td>An array of generated dropout times</td> <td>Double</td> </tr> </tbody> </table> | | | Identifier | Description | Type | DropOutTime | An array of generated dropout times | Double |
| Identifier | Description | Type | | | | | | | |
| DropOutTime | An array of generated dropout times | Double | | | | | | | |

O R Functions

Table O.6: Generating Dropout Times (Survival) (Continued)

| | |
|-------------------------|--|
| Suggested format | <pre>GenDropTimes ← function(NumSub, NumArm, TreatmentID, DropMethod, NumPrd, PrdTime, DropParam) { Error = 0 If(DropMethod == 1) { # Write the actual code for method 1 here. # Store the generated dropouts times in an array called retval. } If(DropMethod == 2) { # Write the actual code for method 2 here. # Store the generated dropout times in an array called retval. } # Use appropriate error handling and modify the # Error in each of the methods appropriately. return(list(DropOutTime = as.double(retval), ErrorCode = as.integer(Error)) } Please note that ErrorCode is optional for this function.</pre> |
|-------------------------|--|

0.3.4 Randomizing Subjects to Treatments

Table O.7: Treatment Randomization

| Suggested Name of the function | <i>GenTreatID()</i> | | | | | | | | | | |
|---------------------------------------|--|---------|--|------------|-------------|--------|---------------------------------|--|-----------------------------------|------------|--|
| Description | Randomizes subjects to specified arms. The function should produce 0-based indexes of arms to which the subjects are allocated. The treatment arms have consecutive positive arm indices starting with 1. | | | | | | | | | | |
| Syntax | <i>randomize(NumSub, NumArm, AllocRatio)</i> | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>NumSub</td> <td>Number of Subjects to randomize</td> </tr> <tr> <td>NumArm</td> <td>Total Number of arms in the trial</td> </tr> <tr> <td>AllocRatio</td> <td>Array of size (NumArm-1) specifying expected allocation ratios for the treatment arms (Allocation ratios are relative to placebo.)</td> </tr> </tbody> </table> | | | Argument | Description | NumSub | Number of Subjects to randomize | NumArm | Total Number of arms in the trial | AllocRatio | Array of size (NumArm-1) specifying expected allocation ratios for the treatment arms (Allocation ratios are relative to placebo.) |
| Argument | Description | | | | | | | | | | |
| NumSub | Number of Subjects to randomize | | | | | | | | | | |
| NumArm | Total Number of arms in the trial | | | | | | | | | | |
| AllocRatio | Array of size (NumArm-1) specifying expected allocation ratios for the treatment arms (Allocation ratios are relative to placebo.) | | | | | | | | | | |
| Return Value Type | R List The must identifiers in this list are <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>TreatmentID</td> <td>An array of generated allocation indices for all subjects. Placebo = 0</td> <td>Integer</td> </tr> </tbody> </table> | | | Identifier | Description | Type | TreatmentID | An array of generated allocation indices for all subjects. Placebo = 0 | Integer | | |
| Identifier | Description | Type | | | | | | | | | |
| TreatmentID | An array of generated allocation indices for all subjects. Placebo = 0 | Integer | | | | | | | | | |
| Suggested format | <pre> GenTreatID ← function(NumSub, NumArm, AllocRatio) { Error = 0 # Write the actual code here. Store the generated treatment indices # in an array called retval. Use error handling and modify the error appropriately. return(list(TreatmentID = as.integer(retval), ErrorCode = as.integer(Error))) } </pre> | | | | | | | | | | |

O R Functions

0.3.5 Randomizing Subjects to Groups

Table O.8: Group Randomization

| Suggested Name of the function | <i>GenGroupID()</i> | | | | | | | | | | |
|---------------------------------------|---|---------|--|------------|-------------|--------|---------------------------------|--|--------------------------------|------------|--|
| Description | Randomizes subjects to specified groups. The function should produce 0-based indexes of groups to which the subjects are allocated. The groups have consecutive positive group indices starting with 1. The first group will have index 0. | | | | | | | | | | |
| Syntax | <i>GenGroupID (NumSub, NumGrp, AllocRatio)</i> | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>NumSub</td> <td>Number of Subjects to randomize</td> </tr> <tr> <td>NumGrp</td> <td>Number of Groups in the trial.</td> </tr> <tr> <td>AllocRatio</td> <td>Array of size (NumGrp-1) specifying expected allocation ratios for the Groups (Allocation ratios are relative to first Group.)</td> </tr> </tbody> </table> | | | Argument | Description | NumSub | Number of Subjects to randomize | NumGrp | Number of Groups in the trial. | AllocRatio | Array of size (NumGrp-1) specifying expected allocation ratios for the Groups (Allocation ratios are relative to first Group.) |
| Argument | Description | | | | | | | | | | |
| NumSub | Number of Subjects to randomize | | | | | | | | | | |
| NumGrp | Number of Groups in the trial. | | | | | | | | | | |
| AllocRatio | Array of size (NumGrp-1) specifying expected allocation ratios for the Groups (Allocation ratios are relative to first Group.) | | | | | | | | | | |
| Return Value Type | R List The must identifiers in this list are <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>GroupID</td> <td>An array of generated allocation indices for all subjects.</td> <td>Integer</td> </tr> </tbody> </table> | | | Identifier | Description | Type | GroupID | An array of generated allocation indices for all subjects. | Integer | | |
| Identifier | Description | Type | | | | | | | | | |
| GroupID | An array of generated allocation indices for all subjects. | Integer | | | | | | | | | |
| Suggested format | <pre>GenGroupID ← function(NumSub, NumGrp, AllocRatio) { Error = 0 # Write the actual code here. Store the generated group indices # in an array called retval. Use appropriate error handling # and modify the Error appropriately. return(list(GroupID = as.integer(retval), ErrorCode = as.integer(Error))) }</pre> | | | | | | | | | | |

0.3.6 Randomizing Subjects to Populations

Table O.9: Population Randomization

| Suggested Name of the function | <i>GenPopulationID()</i> | | | | | | | | | | |
|---------------------------------------|---|---------|--|------------|-------------|--------|---------------------------------|---|-------------------------------------|-----------|--|
| Description | Randomizes subjects to specified populations. Used only for Trend in R ordered proportions test. The function should produce 0-based indices of populations to which the subjects are allocated. The populations have consecutive positive population indices starting with 1. The first population will have index 0. | | | | | | | | | | |
| Syntax | <i>GenPopulationID (NumSub, NumPop, AllocFrac)</i> | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>NumSub</td> <td>Number of Subjects to randomize</td> </tr> <tr> <td>NumPop</td> <td>Number of populations in the trial.</td> </tr> <tr> <td>AllocFrac</td> <td>Array of size (NumPop) specifying expected allocation fractions for the populations.</td> </tr> </tbody> </table> | | | Argument | Description | NumSub | Number of Subjects to randomize | NumPop | Number of populations in the trial. | AllocFrac | Array of size (NumPop) specifying expected allocation fractions for the populations. |
| Argument | Description | | | | | | | | | | |
| NumSub | Number of Subjects to randomize | | | | | | | | | | |
| NumPop | Number of populations in the trial. | | | | | | | | | | |
| AllocFrac | Array of size (NumPop) specifying expected allocation fractions for the populations. | | | | | | | | | | |
| Return Value Type | R List The must identifiers in this list are <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>PopulationID</td> <td>An array of generated populations indices for all subjects.</td> <td>Integer</td> </tr> </tbody> </table> | | | Identifier | Description | Type | PopulationID | An array of generated populations indices for all subjects. | Integer | | |
| Identifier | Description | Type | | | | | | | | | |
| PopulationID | An array of generated populations indices for all subjects. | Integer | | | | | | | | | |
| Suggested format | <pre>GenPopulationID ← function(NumSub, NumPop, AllocFrac) { Error = 0 # Write the actual code here. Store the generated population # indices in an array called retval. Use appropriate error handling # and modify the Error appropriately. return(list(PopulationID = as.integer(retval), ErrorCode = as.integer(Error))) }</pre> | | | | | | | | | | |

O *R Functions*

O.4 *Generating Continuous Response*

O.4.1 Response for Single Mean Test

O.4.2 Response for Mean of Paired Differences Test

O.4.3 Response for Difference of Means Test

O.4.4 Response for Mean of Paired Ratio Test

O.4.5 Generating Response for Ratio of Means Test

O.4.6 Generating Binary Response Values

O.4.7 Generating Categorical Response Values

O.4.8 Generating Survival Times

In this section we describe various functions for generating continuous response for various tests in East as well as SiZ.

O.4.1 *Response for Single Mean Test*

Table O.10: Generating response for Single Mean Test

| Suggested Name of the function | <i>GenRespSingleMean()</i> | | | | | | | | | | |
|---|---|--------|--|------------|-------------|--------|--------------------|---|--|--------|---|
| Description for a specified number of subjects. | Generates response values for Single Mean Test | | | | | | | | | | |
| Syntax | <i>GenRespSingleMean (NumSub, Mean, StdDev)</i> | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>NumSub</td> <td>Number of Subjects</td> </tr> <tr> <td>Mean</td> <td>Array (Size 1) specifying mean response value.</td> </tr> <tr> <td>StdDev</td> <td>Array (Size 1) specifying standard deviation.</td> </tr> </tbody> </table> | | | Argument | Description | NumSub | Number of Subjects | Mean | Array (Size 1) specifying mean response value. | StdDev | Array (Size 1) specifying standard deviation. |
| Argument | Description | | | | | | | | | | |
| NumSub | Number of Subjects | | | | | | | | | | |
| Mean | Array (Size 1) specifying mean response value. | | | | | | | | | | |
| StdDev | Array (Size 1) specifying standard deviation. | | | | | | | | | | |
| Return Value Type | R List The must identifiers in this list are <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>Response</td> <td>An array of generated response for all subjects</td> <td>Double</td> </tr> </tbody> </table> | | | Identifier | Description | Type | Response | An array of generated response for all subjects | Double | | |
| Identifier | Description | Type | | | | | | | | | |
| Response | An array of generated response for all subjects | Double | | | | | | | | | |
| Suggested format | <pre>GenRespSingleMean ← function(NumSub, Mean, StdDev) { Error = 0 # Write the actual code here. # Store the generated response values in an # array called retval. # Use appropriate error handling and modify the # Error appropriately. return(list(Response = as.double(retval), Error- Code = as.integer(Error))) } Please note that ErrorCode is optional for this function.</pre> | | | | | | | | | | |

0.4.2 Response for Mean of Paired Differences Test

O R Functions

Table O.11: Generating response for Mean of Paired Differences test

| Suggested Name of the function | <i>GenRespPairedDiff()</i> | | | | | | | | | | | | | | | | | |
|---------------------------------------|--|--------|--|------------|-------------|--------|--------------------|---|---|--------|--|--|-------|--|--------|-------|--|--------|
| Description | Generates response values for Mean of Paired Differences Test for a specified number of subjects. | | | | | | | | | | | | | | | | | |
| Syntax | <i>GenRespPairedDiff (NumSub, Mean, SigmaD)</i> | | | | | | | | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>NumSub</td> <td>Number of Subjects</td> </tr> <tr> <td>Mean</td> <td>Array (Size 2) specifying mean response value for Control (First element) and Treatment (second element) Arm.</td> </tr> <tr> <td>SigmaD</td> <td>Array (Size 1) specifying Standard Deviation of Paired Difference.</td> </tr> </tbody> </table> | | | Argument | Description | NumSub | Number of Subjects | Mean | Array (Size 2) specifying mean response value for Control (First element) and Treatment (second element) Arm. | SigmaD | Array (Size 1) specifying Standard Deviation of Paired Difference. | | | | | | | |
| Argument | Description | | | | | | | | | | | | | | | | | |
| NumSub | Number of Subjects | | | | | | | | | | | | | | | | | |
| Mean | Array (Size 2) specifying mean response value for Control (First element) and Treatment (second element) Arm. | | | | | | | | | | | | | | | | | |
| SigmaD | Array (Size 1) specifying Standard Deviation of Paired Difference. | | | | | | | | | | | | | | | | | |
| Return Value Type | <p>R List</p> <p>The must identifiers in this list are</p> <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>DiffResp</td> <td>An array of Difference of generated response values on Treatment and Control arm.</td> <td>Double</td> </tr> <tr> <td colspan="3" style="text-align: center;">OR</td> </tr> <tr> <td>RespC</td> <td>An array of generated Control response values for all subjects</td> <td>Double</td> </tr> <tr> <td>RespT</td> <td>An array of generated Treatment response values for all subjects</td> <td>Double</td> </tr> </tbody> </table> <p>Note - If "DiffResp" is found in output list then "RespC" and "RespT" will be optional identifiers otherwise they will be mandatory identifiers</p> | | | Identifier | Description | Type | DiffResp | An array of Difference of generated response values on Treatment and Control arm. | Double | OR | | | RespC | An array of generated Control response values for all subjects | Double | RespT | An array of generated Treatment response values for all subjects | Double |
| Identifier | Description | Type | | | | | | | | | | | | | | | | |
| DiffResp | An array of Difference of generated response values on Treatment and Control arm. | Double | | | | | | | | | | | | | | | | |
| OR | | | | | | | | | | | | | | | | | | |
| RespC | An array of generated Control response values for all subjects | Double | | | | | | | | | | | | | | | | |
| RespT | An array of generated Treatment response values for all subjects | Double | | | | | | | | | | | | | | | | |

0.4.3 Response for Difference of Means Test

The following table provides details of the functions for generating response for difference of means test.

Table O.12: Generating response for Mean of Paired Differences test (Contd.)

| | |
|--------------------------------|--|
| <p>Suggested format</p> | <p>Format1 GenRespPairedDiff ← function(NumSub, Mean, SigmaD) { Error = 0 # Write the actual code here. # Store the generated difference of response values in an # array called retval. # Use appropriate error handling and modify the # Error appropriately. return(list(DiffResp = as.double(retval), ErrorCode = as.integer(Error))) } Format2 GenRespPairedDiff ← function(NumSub, Mean, SigmaD) { Error = 0 # Write the actual code here. # Store the generated Responses on control arm values in an # array called retval1. # Store the generated Responses on treatment arm values in an # array called retval2. # Use appropriate error handling and modify the # Error appropriately. return(list(RespC = as.double(retval1), RespT = as.double(retval2), ErrorCode = as.integer(Error))) }</p> |
|--------------------------------|--|

O R Functions

Table O.13: Generating Response for Difference of Mean Test

| Suggested Name of the function | <i>GenRespDiffofMeans()</i> | | | | | | | | | | | | |
|---------------------------------------|---|--------|--|------------|-------------|--------|--------------------|---------------------------------------|---|------|--|--------|--|
| Description | Generates response values for Difference of Means test for a specified number of subjects. | | | | | | | | | | | | |
| Syntax | <i>GenRespDiffofMeans (NumSub, TreatmentID, Mean, StdDev)</i> | | | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>NumSub</td> <td>Number of Subjects</td> </tr> <tr> <td>TreatmentID</td> <td>Array specifying indexes of arms to which subjects are allocated (one arm index per subject). Index for placebo / control is 0.</td> </tr> <tr> <td>Mean</td> <td>Array (size 2) specifying mean response values for control (first element) and treatment (second element) arms</td> </tr> <tr> <td>StdDev</td> <td>Array (of size 2) specifying standard deviations for control (first element) and treatment (second element) arm.</td> </tr> </tbody> </table> | | | Argument | Description | NumSub | Number of Subjects | TreatmentID | Array specifying indexes of arms to which subjects are allocated (one arm index per subject). Index for placebo / control is 0. | Mean | Array (size 2) specifying mean response values for control (first element) and treatment (second element) arms | StdDev | Array (of size 2) specifying standard deviations for control (first element) and treatment (second element) arm. |
| Argument | Description | | | | | | | | | | | | |
| NumSub | Number of Subjects | | | | | | | | | | | | |
| TreatmentID | Array specifying indexes of arms to which subjects are allocated (one arm index per subject). Index for placebo / control is 0. | | | | | | | | | | | | |
| Mean | Array (size 2) specifying mean response values for control (first element) and treatment (second element) arms | | | | | | | | | | | | |
| StdDev | Array (of size 2) specifying standard deviations for control (first element) and treatment (second element) arm. | | | | | | | | | | | | |
| Return Value Type | R List The must identifiers in this list are <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>Response</td> <td>An array of response for all subjects</td> <td>Double</td> </tr> </tbody> </table> | | | Identifier | Description | Type | Response | An array of response for all subjects | Double | | | | |
| Identifier | Description | Type | | | | | | | | | | | |
| Response | An array of response for all subjects | Double | | | | | | | | | | | |
| Suggested format | <pre>GenRespDiffofMeans ← function (NumSub, TreatmentID, Mean, StdDev) { Error = 0 # Write the actual code here. Store the generated continuous # response values in an array called retval. # Use appropriate error handling and modify the # Error appropriately. return(list(Response = as.double(retval), ErrorCode = as.integer(Error))) } Please note that ErrorCode is optional for this function.</pre> | | | | | | | | | | | | |

0.4.4 Response for Mean of Paired Ratio Test

Table O.14: Generating Response for Mean of Paired Ratio Test

| Suggested Name of the function | <i>GenRespPairedRatio()</i> | | | | | | | | | | | | | | | | | |
|---------------------------------------|---|--------|--|------------|-------------|--------|--------------------|--|---|----------------|--|--|-------|--|--------|-------|--|--------|
| Description | Generates response values for Mean of Paired Ratios Test for a specified number of subjects. | | | | | | | | | | | | | | | | | |
| Syntax | <i>GenRespPairedRatio(NumSub, Mean, StdDevLogRatio)</i> | | | | | | | | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>NumSub</td> <td>Number of Subjects</td> </tr> <tr> <td>Mean</td> <td>Array (Size 2) specifying mean response values (i.e. means of corresponding Log Normal distribution) for Control (first element) and Treatment (second element)Arm.</td> </tr> <tr> <td>StdDevLogRatio</td> <td>Array (Size 1) specifying Standard Deviation of Log of Ratio of Response of Treatment and Control.</td> </tr> </tbody> </table> | | | Argument | Description | NumSub | Number of Subjects | Mean | Array (Size 2) specifying mean response values (i.e. means of corresponding Log Normal distribution) for Control (first element) and Treatment (second element)Arm. | StdDevLogRatio | Array (Size 1) specifying Standard Deviation of Log of Ratio of Response of Treatment and Control. | | | | | | | |
| Argument | Description | | | | | | | | | | | | | | | | | |
| NumSub | Number of Subjects | | | | | | | | | | | | | | | | | |
| Mean | Array (Size 2) specifying mean response values (i.e. means of corresponding Log Normal distribution) for Control (first element) and Treatment (second element)Arm. | | | | | | | | | | | | | | | | | |
| StdDevLogRatio | Array (Size 1) specifying Standard Deviation of Log of Ratio of Response of Treatment and Control. | | | | | | | | | | | | | | | | | |
| Return Value Type | <p>R List The must identifiers in this list are</p> <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>RatioResp</td> <td>An array of Ratio of generated response values on treatment and control arm.</td> <td>Double</td> </tr> <tr> <td colspan="3" style="text-align: center;">OR</td> </tr> <tr> <td>RespC</td> <td>An array of generated control response values for all subjects</td> <td>Double</td> </tr> <tr> <td>RespT</td> <td>An array of generated treatment response values for all subjects</td> <td>Double</td> </tr> </tbody> </table> <p>Note - If "RatioResp" is found in output list then "RespC" and "RespT" will be optional identifiers otherwise they will be mandatory identifiers</p> | | | Identifier | Description | Type | RatioResp | An array of Ratio of generated response values on treatment and control arm. | Double | OR | | | RespC | An array of generated control response values for all subjects | Double | RespT | An array of generated treatment response values for all subjects | Double |
| Identifier | Description | Type | | | | | | | | | | | | | | | | |
| RatioResp | An array of Ratio of generated response values on treatment and control arm. | Double | | | | | | | | | | | | | | | | |
| OR | | | | | | | | | | | | | | | | | | |
| RespC | An array of generated control response values for all subjects | Double | | | | | | | | | | | | | | | | |
| RespT | An array of generated treatment response values for all subjects | Double | | | | | | | | | | | | | | | | |

O R Functions

| | |
|--------------------------------|--|
| <p>Suggested format</p> | <p>Format1 GenRespPairedRatio ← function(NumSub, Mean, StdDevLogRatio) { Error = 0 # Write the actual code here. # Store the generated ratio of response values in an # array called retval. # Use appropriate error handling and modify the # Error appropriately. return(list(RatioResp = as.double(retval), ErrorCode = as.integer(Error))) } Format2 GenRespPairedRatio ← function(NumSub, Mean, StdDevLogRatio) { Error = 0 # Write the actual code here. # Store the generated Responses on control arm values in an # array called retval1. # Store the generated Responses on treatment arm values in an # array called retval2. # Use appropriate error handling and modify the # Error appropriately. return(list(RespC = as.double(retval1), RespT = as.double(retval2), ErrorCode = as.integer(Error))) }</p> |
|--------------------------------|--|

0.4.5 Generating Response for Ratio of Means Test

Table O.15: Generating response for Ratio of Means Test

| Suggested Name of the function | <i>GenRespRatioofMeans()</i> | | | | | | | | | | | | |
|---------------------------------------|--|--------|--|------------|-------------|--------|--------------------|---|--|------|---|----|--|
| Description | Generates response values for Ratio of Means test for a specified number of subjects. | | | | | | | | | | | | |
| Syntax | <i>GenRespRatioofMeans (NumSub, TreatmentID, Mean, CV)</i> | | | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>NumSub</td> <td>Number of Subjects</td> </tr> <tr> <td>TreatmentID</td> <td>Array specifying indexes of arms to which subjects are allocated (one arm index per subject). Index for placebo / control is 0.</td> </tr> <tr> <td>Mean</td> <td>Array (size 2) specifying mean response values (i.e. means of corresponding Log Normal distribution) for control (first element) and treatment (second element) arms.</td> </tr> <tr> <td>CV</td> <td>Array (size 2) specifying Coefficient of Variation for control (first element) and treatment (second element) arm.</td> </tr> </tbody> </table> | | | Argument | Description | NumSub | Number of Subjects | TreatmentID | Array specifying indexes of arms to which subjects are allocated (one arm index per subject). Index for placebo / control is 0. | Mean | Array (size 2) specifying mean response values (i.e. means of corresponding Log Normal distribution) for control (first element) and treatment (second element) arms. | CV | Array (size 2) specifying Coefficient of Variation for control (first element) and treatment (second element) arm. |
| Argument | Description | | | | | | | | | | | | |
| NumSub | Number of Subjects | | | | | | | | | | | | |
| TreatmentID | Array specifying indexes of arms to which subjects are allocated (one arm index per subject). Index for placebo / control is 0. | | | | | | | | | | | | |
| Mean | Array (size 2) specifying mean response values (i.e. means of corresponding Log Normal distribution) for control (first element) and treatment (second element) arms. | | | | | | | | | | | | |
| CV | Array (size 2) specifying Coefficient of Variation for control (first element) and treatment (second element) arm. | | | | | | | | | | | | |
| Return Value Type | R List The must identifiers in this list are <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>Response</td> <td>An array of generated response (from Log Normal Distribution)</td> <td>Double</td> </tr> </tbody> </table> | | | Identifier | Description | Type | Response | An array of generated response (from Log Normal Distribution) | Double | | | | |
| Identifier | Description | Type | | | | | | | | | | | |
| Response | An array of generated response (from Log Normal Distribution) | Double | | | | | | | | | | | |

O R Functions

Table O.16: Generating response for Ratio of Means Test (Contd)

| | |
|-------------------------|---|
| Suggested format | <pre> GenRespRatioofMeans ← function(NumSub,TreatmentID, Mean, CV) { Error = 0 # Write the actual code here. Store the generated response # values in an array called retval. # Use appropriate error handling and modify the # Error appropriately. return(list(Response = as.double(retval), ErrorCode = as.integer(Error))) } Please note that ErrorCode is optional for this function. </pre> |
|-------------------------|---|

0.4.6 Generating Binary Response Values

The following table provides details of generating binary response values.

Table O.17: Generating Binary Response Values

| | |
|---|--|
| Suggested Name of the function | <i>GenBinResp()</i> |
| Description | Generates Binary response (Two categories 0 (Non-Responder) and 1 (Responder) values for a specified number of subjects. |
| Syntax for one arm test | <i>GenBinResp (NumSub, PropResp)</i> |
| Syntax for more than one arm test | <i>GenBinResp (NumSub, NumArm, TreatmentID, PropResp)</i> |
| Syntax only for Trend in R Ordered Proportions | <i>GenTrendResp (NumSub, NumPop, PopulationID, PropResp)</i> |

Table O.18: Generating Binary Response Values (Contd)

| <p>Arguments</p> | <table border="1"> <thead> <tr> <th data-bbox="539 355 689 389">Argument</th> <th data-bbox="689 355 1189 389">Description</th> </tr> </thead> <tbody> <tr> <td data-bbox="539 389 689 423">NumSub</td> <td data-bbox="689 389 1189 423">Number of Subjects</td> </tr> <tr> <td data-bbox="539 423 689 491">NumArm</td> <td data-bbox="689 423 1189 491">Number of arms in the trial (including placebo / control)</td> </tr> <tr> <td data-bbox="539 491 689 722">TreatmentID</td> <td data-bbox="689 491 1189 722">Array specifying indexes of arms to which subjects are allocated (one arm index per subject). Index for placebo / control is 0. For other arms, indexes are consecutive positive numbers starting with 1. Thus if the trial has 4 arms (1 placebo + 3 treatment arms), arm indexes will be 0, 1, 2 and 3.</td> </tr> <tr> <td data-bbox="539 722 689 953">PopulationID</td> <td data-bbox="689 722 1189 953">Array specifying indexes of populations to which subjects are allocated (one population index per subject). Index for first population is 0. For other populations, indexes are consecutive positive numbers starting with 1. Thus if the trial has 4 populations, their indices will be 0, 1, 2 and 3.</td> </tr> <tr> <td data-bbox="539 953 689 1021">PropResp</td> <td data-bbox="689 953 1189 1021">An array specifying expected proportions of responders on each arm/Population.</td> </tr> </tbody> </table> | | | Argument | Description | NumSub | Number of Subjects | NumArm | Number of arms in the trial (including placebo / control) | TreatmentID | Array specifying indexes of arms to which subjects are allocated (one arm index per subject). Index for placebo / control is 0. For other arms, indexes are consecutive positive numbers starting with 1. Thus if the trial has 4 arms (1 placebo + 3 treatment arms), arm indexes will be 0, 1, 2 and 3. | PopulationID | Array specifying indexes of populations to which subjects are allocated (one population index per subject). Index for first population is 0. For other populations, indexes are consecutive positive numbers starting with 1. Thus if the trial has 4 populations, their indices will be 0, 1, 2 and 3. | PropResp | An array specifying expected proportions of responders on each arm/Population. |
|---------------------------------|--|--------|--|------------|-------------|--------|--------------------|--|---|-------------|---|--------------|---|----------|--|
| Argument | Description | | | | | | | | | | | | | | |
| NumSub | Number of Subjects | | | | | | | | | | | | | | |
| NumArm | Number of arms in the trial (including placebo / control) | | | | | | | | | | | | | | |
| TreatmentID | Array specifying indexes of arms to which subjects are allocated (one arm index per subject). Index for placebo / control is 0. For other arms, indexes are consecutive positive numbers starting with 1. Thus if the trial has 4 arms (1 placebo + 3 treatment arms), arm indexes will be 0, 1, 2 and 3. | | | | | | | | | | | | | | |
| PopulationID | Array specifying indexes of populations to which subjects are allocated (one population index per subject). Index for first population is 0. For other populations, indexes are consecutive positive numbers starting with 1. Thus if the trial has 4 populations, their indices will be 0, 1, 2 and 3. | | | | | | | | | | | | | | |
| PropResp | An array specifying expected proportions of responders on each arm/Population. | | | | | | | | | | | | | | |
| <p>Return Value Type</p> | <p>R List - The must identifiers in this list are</p> <table border="1"> <thead> <tr> <th data-bbox="539 1072 689 1106">Identifier</th> <th data-bbox="689 1072 1096 1106">Description</th> <th data-bbox="1096 1072 1189 1106">Type</th> </tr> </thead> <tbody> <tr> <td data-bbox="539 1106 689 1175">Response</td> <td data-bbox="689 1106 1096 1175">An array of generated Binary response for all subjects</td> <td data-bbox="1096 1106 1189 1175">Double</td> </tr> </tbody> </table> | | | Identifier | Description | Type | Response | An array of generated Binary response for all subjects | Double | | | | | | |
| Identifier | Description | Type | | | | | | | | | | | | | |
| Response | An array of generated Binary response for all subjects | Double | | | | | | | | | | | | | |
| <p>Suggested format</p> | <pre> GenBinResp ← function(NumSub, NumArm, TreatmentID, PropResp) {Error = 0 # Write the actual code here. Store the generated binary response # values in an array called retval. # Use appropriate error handling and modify the # Error appropriately. return(list(Response = as.double(retval), ErrorCode = as.integer(Error))) } </pre> | | | | | | | | | | | | | | |

O R Functions

0.4.7 Generating Categorical Response Values

Table O.19: Generating Categorical Response Values

| Suggested Name of the function | <i>GenCatResp()</i> | | | | | | | | | | | |
|--|--|--|----------|-------------|--------|--------------------|--------|--------------------------------|---------|---|--------------------|--|
| Description | Generates Categorical response values (0 to (Number of categories-1)) for a specified number of subjects. Binary response is a special case of this when number of categories is 2. | | | | | | | | | | | |
| Syntax for one group test | <i>GenCatResp(NumSub, NumCat, PropResp)</i> | | | | | | | | | | | |
| Syntax for more than one group test | <i>GenCatResp (NumSub, NumGrp, GroupID, NumCat, PropResp)</i> | | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>NumSub</td> <td>Number of Subjects</td> </tr> <tr> <td>NumGrp</td> <td>Number of groups in the trial.</td> </tr> <tr> <td>GroupID</td> <td>Array specifying indices of groups to which subjects are allocated.</td> </tr> <tr> <td>NumCat PropResp</td> <td>Number of categories of response. 2-D array specifying expected proportions of responders in each category and on each group. PropResp[i, j] specifies expected proportion of responders in the jth category and on the ith group.</td> </tr> </tbody> </table> | | Argument | Description | NumSub | Number of Subjects | NumGrp | Number of groups in the trial. | GroupID | Array specifying indices of groups to which subjects are allocated. | NumCat PropResp | Number of categories of response. 2-D array specifying expected proportions of responders in each category and on each group. PropResp[i, j] specifies expected proportion of responders in the jth category and on the ith group. |
| Argument | Description | | | | | | | | | | | |
| NumSub | Number of Subjects | | | | | | | | | | | |
| NumGrp | Number of groups in the trial. | | | | | | | | | | | |
| GroupID | Array specifying indices of groups to which subjects are allocated. | | | | | | | | | | | |
| NumCat PropResp | Number of categories of response. 2-D array specifying expected proportions of responders in each category and on each group. PropResp[i, j] specifies expected proportion of responders in the jth category and on the ith group. | | | | | | | | | | | |

Table O.20: Generating Categorical Response Values (Contd)

| | | | |
|--------------------------|--|---|--------|
| Return Value Type | R List The must identifiers in this list are | | |
| | Identifier | Description | Type |
| | CatID | An array of generated categorical response (0,1,2,...,(NumCat-1)) for all subjects. | Double |
| Suggested format | <pre> GenCatResp ← function(NumSub, NumGrp, GroupID, Num- Cat, PropResp) { Error = 0 # Write the actual code here. # Store the generated multinomial response values in an # array called retval. # Use appropriate error handling and modify the # Error appropriately. return(list(CatID = as.double(retval), ErrorCode = as.integer(Error))) } Please note that ErrorCode is optional for this function. </pre> | | |

O R Functions

0.4.8 Generating Survival Times

Table O.21: Generating Survival Times (Time to Response)

| Suggested Name of the function | <i>GenSurvTime()</i> | | | | | | | | | | | | | | | |
|--|--|--|----------|-------------|--------|--------------------|--------|------------------------------|-------------|--|------------|--|--------|-----------------------------|---------|---|
| Description | Generates survival times for a specified number of subjects. | | | | | | | | | | | | | | | |
| Syntax for one arm test | <i>GenSurvTime (NumSub, SurvMethod, NumPrd, PrdTime, SurvParam)</i> | | | | | | | | | | | | | | | |
| Syntax for more than one arm test | <i>GenSurvTime (NumSub, NumArm, TreatmentID, SurvMethod, NumPrd, PrdTime, SurvParam)</i> | | | | | | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>NumSub</td> <td>Number of Subjects</td> </tr> <tr> <td>NumArm</td> <td>Number of Arms in the trial.</td> </tr> <tr> <td>TreatmentID</td> <td>Array specifying indexes of arms to which subjects are allocated (one arm index per subject). Index for placebo / control is 0. For other arms, indexes are consecutive positive numbers starting with 1.</td> </tr> <tr> <td>SurvMethod</td> <td>Input method. 1 - Hazard rates. 2 - Cumulative % survival rates. 3 - Median Survival Times.</td> </tr> <tr> <td>NumPrd</td> <td>Number of survival periods.</td> </tr> <tr> <td>PrdTime</td> <td>Array of times used to specify survival parameters. If SurvMethod is 1, this array specifies the starting times of hazard pieces. If SurvMethod is 2, this array specifies the times at which the cumulative % survivals are specified. If SurvMethod is 3, the period time is 0 by default.</td> </tr> </tbody> </table> | | Argument | Description | NumSub | Number of Subjects | NumArm | Number of Arms in the trial. | TreatmentID | Array specifying indexes of arms to which subjects are allocated (one arm index per subject). Index for placebo / control is 0. For other arms, indexes are consecutive positive numbers starting with 1. | SurvMethod | Input method. 1 - Hazard rates. 2 - Cumulative % survival rates. 3 - Median Survival Times. | NumPrd | Number of survival periods. | PrdTime | Array of times used to specify survival parameters. If SurvMethod is 1, this array specifies the starting times of hazard pieces. If SurvMethod is 2, this array specifies the times at which the cumulative % survivals are specified. If SurvMethod is 3, the period time is 0 by default. |
| Argument | Description | | | | | | | | | | | | | | | |
| NumSub | Number of Subjects | | | | | | | | | | | | | | | |
| NumArm | Number of Arms in the trial. | | | | | | | | | | | | | | | |
| TreatmentID | Array specifying indexes of arms to which subjects are allocated (one arm index per subject). Index for placebo / control is 0. For other arms, indexes are consecutive positive numbers starting with 1. | | | | | | | | | | | | | | | |
| SurvMethod | Input method. 1 - Hazard rates. 2 - Cumulative % survival rates. 3 - Median Survival Times. | | | | | | | | | | | | | | | |
| NumPrd | Number of survival periods. | | | | | | | | | | | | | | | |
| PrdTime | Array of times used to specify survival parameters. If SurvMethod is 1, this array specifies the starting times of hazard pieces. If SurvMethod is 2, this array specifies the times at which the cumulative % survivals are specified. If SurvMethod is 3, the period time is 0 by default. | | | | | | | | | | | | | | | |

Table O.22: Generating Survival Times (Time to Response) (Contd)

| <p>Arguments</p> | <table border="1"> <thead> <tr> <th data-bbox="539 363 686 397">Argument</th> <th data-bbox="686 363 1196 397">Description</th> </tr> </thead> <tbody> <tr> <td data-bbox="539 397 686 787">SurvParam</td> <td data-bbox="686 397 1196 787"> 2-D array of parameters uses to generate time of events. If SurvMethod is 1, this array specifies arm by arm hazard rates (one rate per arm per piece). Thus SurvParam [i, j] specifies hazard rate in ith period for jth arm. If SurvMethod is 2, this array specifies arm by arm Cum % Survivals (one value per arm per piece). Thus SurvParam [i, j] specifies Cum % Survivals in ith period for jth arm. If SurvMethod is 3, this will be a 1 x 2 array with median survival times on each arms. </td> </tr> </tbody> </table> | Argument | Description | SurvParam | 2-D array of parameters uses to generate time of events. If SurvMethod is 1, this array specifies arm by arm hazard rates (one rate per arm per piece). Thus SurvParam [i, j] specifies hazard rate in ith period for jth arm. If SurvMethod is 2, this array specifies arm by arm Cum % Survivals (one value per arm per piece). Thus SurvParam [i, j] specifies Cum % Survivals in ith period for jth arm. If SurvMethod is 3, this will be a 1 x 2 array with median survival times on each arms. | | |
|---------------------------------|--|------------|-------------|-----------|---|---|--------|
| Argument | Description | | | | | | |
| SurvParam | 2-D array of parameters uses to generate time of events. If SurvMethod is 1, this array specifies arm by arm hazard rates (one rate per arm per piece). Thus SurvParam [i, j] specifies hazard rate in ith period for jth arm. If SurvMethod is 2, this array specifies arm by arm Cum % Survivals (one value per arm per piece). Thus SurvParam [i, j] specifies Cum % Survivals in ith period for jth arm. If SurvMethod is 3, this will be a 1 x 2 array with median survival times on each arms. | | | | | | |
| <p>Return Value Type</p> | <p>R List - The must identifiers in this list are</p> <table border="1"> <thead> <tr> <th data-bbox="539 838 686 872">Identifier</th> <th data-bbox="686 838 1096 872">Description</th> <th data-bbox="1096 838 1196 872">Type</th> </tr> </thead> <tbody> <tr> <td data-bbox="539 872 686 941">SurvivalTime</td> <td data-bbox="686 872 1096 941">An array of generated time to response values for each subject.</td> <td data-bbox="1096 872 1196 941">Double</td> </tr> </tbody> </table> | Identifier | Description | Type | SurvivalTime | An array of generated time to response values for each subject. | Double |
| Identifier | Description | Type | | | | | |
| SurvivalTime | An array of generated time to response values for each subject. | Double | | | | | |
| <p>Suggested format</p> | <pre> GenSurvTime ← function(NumSub, NumArm, TreatmentID, SurvMethod, NumPrd, PrdTime, SurvParam) {Error = 0 If(SurvMethod == 1) { # Write the actual code for SurvMethod 1here. Store the generated survival times in an array called retval. } If(SurvMethod ==2) { # Write the actual code for SurvMethod 2here. # Store the generated survival times in an array called retval. } # Use appropriate error handling and modify the # Error appropriately. return(list(SurvivalTime = as.double(retval), ErrorCode = as.integer(Error))) } </pre> | | | | | | |

O R Functions

O.5 Enhanced Simulations

O.5.1 Input Arguments for One Look Test

O.5.2 Input Arguments for Multi Look Test

O.5.3 Output from R function

User will provide an R function for computing test statistic as well as for performing test for the current look in current simulation. Name of this R function is not mandatory.

O.5.1 Input Arguments for One Look Test

This section describes input arguments for R function for one look test for computing test statistic or perform test for One Look as well as Multi Look tests.

For One Look Test, R function will have following two mandatory named arguments

1. **SimData** - R Data frame which consists of data generated in current simulation (Case Data). This data frame will have headers indicating the names of the columns. These names will be same as those used in Data Generation. User should access the variables using headers for ex. `SimData$ArrivalTime` and not order. Optional outputs from Data Generation will also be available.
2. **DesignParam** - R List which consists of Design and Simulation Parameters which user may need to compute test statistic and perform test. User should access the variables using names for eg. `DesignParam$SideType` and not order. For details of this list please see below.

Table O.23: Input Table for One Look Test

| Argument Name | Description | Applicability | Codes |
|---------------|--|--|---|
| Alpha | Type I Error | Multi Look Enabled One Sided and Two Sided Symmetric Tests | |
| LowerAlpha | Lower Type I Error | Multi Look Enabled Two Sided Asymmetric Tests | |
| UpperAlpha | Upper Type I Error | Multi Look Enabled Two Sided Asymmetric | |
| TrialType | Type of the Trial | All Tests | 0 - Superiority 1 - Non-Inferiority 2 - Equivalence |
| TestType | Type of Test | All Tests | 0 - One Side 1 - Two Sided 2 - Two Sided Asymmetric |
| TailType | Nature of Critical Region | One Sided Tests | 0 - Left Tailed 1 - Right Tailed |
| AllocInfo | Array of the ratios of the treatment group sample sizes to control group sample size | Multi Arms Tests | |
| | Population Fractions | Trend Test | |

O R Functions

Table O.24: Input Table for One Look Test (Contd)

| Argument Name | Description | Applicability | Codes |
|--------------------|------------------------------|--------------------------------------|--|
| CriticalPoint | Critical value | Single Look One Sided Tests | |
| UpperCriticalPoint | Upper Critical value | Single Look Two Sided Tests | |
| LowerCriticalPoint | Lower Critical value | Single Look Two Sided Tests | |
| SampleSize | Sample Size | All | |
| MaxCompleters | Maximum Number of Completers | All Non Survival Tests | |
| RespLag | Response Lag | All non survival tests | |
| LookFixOption | Time/Events Based Flag | All Survival Tests | 0 - Event Based 1 - Time Based |
| MaxEvents | Maximum Events | All Survival Event Based Tests | |
| MaxStudyDur | Maximum Study Duration | All Survival Time Based Tests | |
| FollowUpType | Follow Up Type | All Survival Tests | 0 - Until End of the Study 1 - For Fixed Period |
| FollowUpDur | Follow Up Duration | All Survival Tests | |
| TestStatType | Test Statistic Type | All Normal Tests | 3 - Z test 4 - t Test |
| | | All Survival Tests | 0 - Log Rank 1 - Wilcoxon Gehan 2 - Harrington Fleming |
| | | Ratio of Proportions Non-Inferiority | 5 - Wald 6 - Score |

Table O.25: Input Table for One Look Test (Contd)

| Argument Name | Description | Applicability | Codes |
|----------------|--|--|---|
| HFParam | Harrington Fleming Parameter | Survival Tests | |
| VarType | Variance Type | t Test Diff of Prop Ratio of Prop Single Prop Ratio of Proportions Score | 4 - Equal 5 - UnEqual 0 - Pooled 1 - UnPooled 2 - Null 3 - Empirical |
| SigmaD | Standard Deviation of Paired Difference | Mean of Paired Difference Z test | |
| SigmaLogRatio | Standard Deviation of Log Ratio | Mean of Paired Ratios Z test | |
| CoeffVar | Coefficient of Variation | Ratio of Means Z test | |
| Sigma | Standard Deviation | All other Normal Z tests | |
| TrtEffNull | Treatment Effect under Null on natural scale | All Single Arms Test and Non-Inferiority Trials in Two Arms Tests | |
| UpperEquiLimit | Upper Equivalence Limit on Natural Scale | All Continuous Tests with Equivalence Trial Type | |
| LowerEquiLimit | Lower Equivalence Limit on Natural Scale | All Continuous Tests with Equivalence Trial Type | |
| EquiMargin | Equivalence Margin | Difference of Proportions Test with Equivalence Trial Type | |
| MuC | Mean for the Control Arm | Multi Look Enabled Normal Two Arms Tests | |

O R Functions

Table O.26: Input Table for One Look Test (Contd)

| Argument Name | Description | Applicability | Codes |
|---------------|---|--|---|
| PiC | Proportion for Control Arm | Multi Look Enabled Binary Two Arms Tests | |
| NumHzrdPrd | Number of Hazard Pieces | All Survival Tests | |
| PrdAt | Array of Starting Value of Each Period | All Survival Tests | |
| LambdaC | Array of Control Hazard Rates for each period | All Survival Tests | |
| TestID | Test ID | | Single Mean - 101 Mean of Paired Diff. - 105 Diff. of Means - 102 Single Prop. - 301 Diff. of Prop. - 303 Ratio of Prop. - 304 Ratio of Prop. FM - 305 Odds Ratio - 306 Survival Given study Durn. - 401 Survival Given Accrual Rates - 410 Ratio of Means test= 103 Mean of Paired Ratios= 106 Diff of Prop Equivalence = 309 Trend in R ordered Proportions = 310 Chisquare test for specified proportions in C categories = 201 Two Group Chi square for Proportions in C Categories = 202 Chi Square for Proportions in RxC tables= 203 Chi Square for Proportions in Rx2 tables = 314 |

Table O.27: Input Table for One Look Test (Contd)

| Argument Name | Description | Applicability | Codes |
|---------------|--|---|-------|
| Scores | Array of Scores | Trend in R proportions Test | |
| NumPop | Number of Populations | Trend in R proportions Test | |
| NumGrp | Number of Groups | Chi Square Tests | |
| NumCat | Number of Categories | Chi Square Tests | |
| CatPropNull | Array of category wise Proportions under Null Hypothesis | Chi-Square for Specified Proportions in C Categories Test | |

0.5.2 Input Arguments for Multi Look Test

For Multi Look Test, R function will have following three mandatory named arguments

1. SimData - Same as for One Look Test
2. DesignParam - Same as for One Look Test
3. LookInfo - R List which consists of Design and Simulation

Parameters related to multi looks which user may need to compute test statistic and perform test. User should access the variables using names for ex. LookInfo\$SideType and not order. For details of this list please see below.

O R Functions

Table O.28: Input Table for Multi Look Tests

| Argument Name | Description | Applicability | Codes |
|---------------|---------------------------------------|--------------------------------|--|
| NumLooks | Number of Looks | All Tests | |
| CurrLookIndex | Current Look Index (1- Based) | All Tests | |
| InfoFrac | Array of Information Fraction | All Tests | |
| CumAlpha | Array of cumulative alpha spent | One Sided Tests | |
| CumAlphaUpper | Array of Upper cumulative alpha spent | Two Sided Tests | |
| CumAlphaLower | Array of Lower cumulative alpha spent | Two Sided Tests | |
| CumCompleters | Array of Cumulative Completers | All Non Survival-Tests | |
| CumEvents | Array of Cumulative Events | All Survival Event Based tests | |
| LookTime | Array of Look Times on Calendar Scale | All Survival Time Based tests | |
| RejType | Rejection Type | All Tests | 1 Sided Efficacy Upper = 0 1 Sided Futility Upper = 1 1 Sided Efficacy Lower = 2 1 Sided Futility Lower = 3 1 Sided Efficacy Upper Futility Lower = 4 1 Sided Efficacy Lower Futility Upper = 5 2 Sided Efficacy Only = 6 2 Sided Futility Only = 7 2 Sided Efficacy Futility = 8 Equivalence = 9 |

Table O.29: Input Table for Multi Look Tests (Contd)

| Argument Name | Description | Applicability | Codes |
|---------------|--|--|---|
| EffBdryScale | Efficacy Boundary Scale | All Tests | 0 - Z Scale 1 - p value scale |
| EffBdry | Array of Efficacy Boundaries | One Sided Tests | |
| EffBdryUpper | Array of Upper Efficacy Boundary | Two Sided Tests | |
| EffBdryLower | Array of Lower Efficacy Boundary | Two Sided Tests | |
| FutBdryScale | Futility Boundary Scale | All Tests | 0 - Z scale 1 - p Value scale 2 - Delta Scale 3 - CP Scale |
| CPDeltaOption | Option of using Design or Estimated Delta for CP Computation | Tests with Futility Boundary on CP Scale | 0 - Design Delta Option 1 - Estimated Delta Option |
| FutBdry | Array of Futility Boundaries | One Sided Tests | |
| FutBdryUpper | Array of Upper Futility Boundary | Two Sided Tests | |
| FutBdryLower | Array of Lower Futility Boundary | Two Sided Tests | |
| BindingType | Binding Type | All Tests | 0 - Non Binding 1 - Binding |

O R Functions

0.5.3 Output from R function

R function will return a list. The Identifier Names (Case Insensitive) and Type (we suggest user type casts the output) mentioned for outputs are compulsory while their order in the list is not. User can have additional outputs (scalars) in the list.

If user wants to print scalars in the Simulation CSV file then he has to provide identifier for those scalars. These identifiers will be the columns names in output. Any repeated identifiers (column names) will be ignored.

User can either return identifier 'Decision' in which case other identifiers will become optional. If 'Decision' is not returned then other identifiers will become mandatory.

We suggest that the return List contain an identifier "ErrorCode". If specified, it has to be of Type Integer. Its values are classified as follows.

1. **0:** No Error
2. **Positive Integer:** Non Fatal Error - Particular Simulation will be terminated but Next Simulation will be performed.
3. **Negative Integer:** Fatal Error - No further simulation will be attempted

We suggest that user should classify error in these categories depending on the context.

Table O.30: Output from R function (Decision Only)

| Identifier | Description | Type | Applicability |
|------------|---|---------|---------------|
| Decision | Decision Code 0 - No Boundary Crossed 1 - Lower Efficacy Boundary Crossed 2 - Upper Efficacy Boundary Crossed 3 - Futility Boundary Crossed 4 - Equivalence Boundary Crossed | Integer | All Tests |

Table O.31: Output from R function (without 'Decision')

| Identifier | Description | Type | Applicability |
|-------------------------------|---|---------|--|
| TestStat | Value of appropriate Test Statistic. Regardless of the Efficacy or Futility Boundary Scale (ex. Delta or p Value or CP Scale) R function should return Test Statistic on Wald (Z) Scale | Double | All Tests except Equivalence Trial Type |
| TestStatLeft TestStatRight | Left and Right Test Statistic on Wald Scale Corresponding to Two Hypotheses | Double | All test with Equivalence Trial Type |
| Delta | Estimate of Delta | Double | Futility Boundary Scale is Delta or CP |
| CtrlCompleters | Number of Completers on Control Arm | Integer | Endpoint is Binomial and FutBdryScale is CP and Delta option is estimated. |
| TrmtCompleters | Number of Completers on Treatment Arm | Integer | Endpoint is Binomial and FutBdryScale is CP and delta options is estimated |
| CtrlPi | Proportion on Control Arm | Double | Endpoint is Binomial and FutBdryScale is CP and delta options is estimated |

O R Functions

O.6 Suggested Formats

O.6.1 Test Stat for One Look

O.6.2 Performing Test for One Look Tests

O.6.3 Computing Test Statistic for Multi Look Tests

O.6.4 Performing Test for Multi Look Tests

O.6.1 Test Stat for One Look

Suggested format for computing test statistic for one look tests is `ComputeTestStat` ← `function(SimData, DesignParam)`

```
{  
Error = 0  
# Write the actual code here.  
# Store the computed test statistic value in retval.  
# Use appropriate error handling and modify the Error appropriately.  
return(list(TestStat = as.double(retval), ErrorCode = as.integer(Error)))  
}
```

Please note that `ErrorCode` is optional for this function. You can also return quantities of interest (scalar) (like estimates) in the output list. Provide identifiers for such outputs and they will be displayed in Output of East6.1

O.6.2 Performing Test for One Look Tests

Suggested format for performing test for one look tests is `PerformDecision` ← `function(SimData, DesignParam)`

```
{  
Error = 0  
# Write the actual code here.  
# compute the test statistic value and store the decision in retval.  
# Use appropriate error handling and modify the Error appropriately.  
return(list(Decision = as.integer(retval), ErrorCode = as.integer(Error)))  
}
```

Please note that `ErrorCode` is optional for this function. You can also return quantities of interest (scalar) (like estimates) in the output list. Provide identifiers for such outputs and they will be displayed in Output of East6.1

O.6.3 Computing Test Statistic for Multi Look Tests

`ComputeTestStat` ← `function(SimData, DesignParam, LookInfo)`

```
{  
Error = 0  
# Write the actual code here.  
# Store the computed test statistic value in retval.  
# Use appropriate error handling and modify the Error appropriately.  
return(list(TestStat = as.double(retval), ErrorCode = as.integer(Error)))  
}
```

Please note that `ErrorCode` is optional for this function. You can also return quantities

of interest (scalar) (like estimates) in the output list. Provide identifiers for such outputs and they will be displayed in Output of East6.1

0.6.4 Performing Test for Multi Look Tests

```
PerformDecision ← function(SimData, DesignParam, LookInfo)
{
Error = 0
# Write the actual code here.
# Compute the test statistic value and store the decision
# value (appropriate code) in retval.
# Use appropriate error handling and modify the Error appropriately.
return(list(Decision = as.integer(retval), ErrorCode = as.integer(Error)))
}
```

Please note that ErrorCode is optional for this function. You can also return quantities of interest (scalar) (like estimates) in the output list. Provide identifiers for such outputs and they will be displayed in Output of East6.1

0.7 Basic Simulation

0.7.1 Input Arguments for One Look Test

0.7.2 Input Arguments for Multi Look Tests

User can perform basic simulation in East6.1 using R function. This option will be available if user performs simulation for Difference of Means Z Test and generates data using Difference of Means option. In this case R function will directly generate test statistic.

0.7.1 Input Arguments for One Look Test

For One Look Test, R function for basic simulation will have only one mandatory named argument

DesignParam - R List which consists of Design and Simulation Parameters which user may need to compute test statistic and perform test. User should access the variables using names for e.g. DesignParam\$SideType and not order.

0.7.2 Input Arguments for Multi Look Tests

For Multi Look Test, R function will have following two mandatory named arguments

1. **DesignParam** - Same as for One Look Test
2. **LookInfo** - R List which consists of Design and Simulation Parameters related to multi looks which user may need to compute test statistic and perform test. User should access the variables using names for ex. LookInfo\$SideType and not order.

O R Functions

O.8 Output from R function

R function for basic simulation will return a list. The Identifier Names (Case Insensitive) and Type (we suggest user type casts the output) mentioned for outputs are compulsory while their order in the list is not. User can have additional outputs (scalars) in the list. If user wants to print scalars in the Simulation CSV file then user has to provide identifier for those scalars. These identifiers will be the columns names in output. Any repeated identifiers (column names) will be ignored.

The must identifier(s) in this list are

| Identifier | Description | Type |
|------------|---|---------|
| Decision | Decision Code 0 - No Boundary Crossed 1 - Lower Efficacy Boundary Crossed 2 - Upper Efficacy Boundary Crossed 3 - Futility Boundary Crossed 4 - Equivalence Boundary Crossed | Integer |
| | OR | |
| TestStat | Test Statistic Value | Double |

We suggest that the return List contain an identifier "ErrorCode". If specified, it has to be of Type Integer. Its values are classified as follows

1. **0:** No Error
2. **Positive Integer:** Non Fatal Error - Particular Simulation will be aborted but Next Simulation will be performed.
3. **Negative Integer:** Fatal Error - No further simulation will be attempted

We suggest that user should classify error in these categories depending on the context.

O.9 Suggested Formats**O.9.1 Test Stat for One Look****O.9.2 Performing Test for One Look Tests****O.9.3 Test Statistic for Multi Look Tests****O.9.4 Performing Test for Multi Look Tests****O.9.1 Test Stat for One Look**

Suggested format for computing test statistic for one look tests is

```
ComputeBasicTestStat ← function(DesignParam)
```

```
{
```

```
Error = 0
```

```
# Write the actual code here.
```

```
# Store the computed test statistic value in retval.
```

```
# Use appropriate error handling and modify the Error appropriately.
```

```
return(list(TestStat = as.double(retval), ErrorCode = as.integer(Error)))
```

```
}
```

Please note that ErrorCode is optional for this function. You can also return quantities of interest (scalar) (like estimates) in the output list. Provide identifiers for such outputs and they will be displayed in Output of East6.1

O.9.2 Performing Test for One Look Tests

Suggested format for performing test for one look tests is PerformDecision ←

```
function(DesignParam)
```

```
{
```

```
Error = 0
```

```
# Write the actual code here.
```

```
# compute the test statistic value and store the decision in retval.
```

```
# Use appropriate error handling and modify the Error appropriately.
```

```
return(list(Decision = as.integer(retval), ErrorCode = as.integer(Error)))
```

```
}
```

Please note that ErrorCode is optional for this function. You can also return quantities of interest (scalar) (like estimates) in the output list. Provide identifiers for such outputs and they will be displayed in Output of East6.1

O.9.3 Test Statistic for Multi Look Tests

```
ComputeBasicTestStat ← function(DesignParam, LookInfo)
```

```
{
```

```
Error = 0
```

```
# Write the actual code here.
```

```
# Store the computed test statistic value in retval.
```

```
# Use appropriate error handling and modify the Error appropriately.
```

```
return(list(TestStat = as.double(retval), ErrorCode = as.integer(Error)))
```

```
}
```

Please note that ErrorCode is optional for this function. You can also return quantities

O R Functions

of interest (scalar) (like estimates) in the output list. Provide identifiers for such outputs and they will be displayed in Output of East6.1

0.9.4 Performing Test for Multi Look Tests

```
PerformDecision ← function(DesignParam, LookInfo)
{
Error = 0
# Write the actual code here.
# Compute the test statistic value and store the decision
# value (appropriate code) in retval.
# Use appropriate error handling and modify the Error appropriately.
return(list(Decision = as.integer(retval), ErrorCode = as.integer(Error)))
}
```

Please note that ErrorCode is optional for this function. You can also return quantities of interest (scalar) (like estimates) in the output list. Provide identifiers for such outputs and they will be displayed in Output of East6.1

O.10 Treatment Selection Function

Treatment Selection can be performed in combining p-values difference of means and difference of proportions designs using R. This section provides details on this functionality.

This function will be called once in each simulation after first look if trial is not terminated.

The use of error codes in this R function is similar to that explained in other R functions.

This function has following inputs

1. **SimData** - R Data frame which consists of data generated in current simulation (Case Data). This data frame will have headers indicating the names of the columns. These names will be same as those used in Data Generation. User should access the variables using headers for ex. `SimData$TreatmentID` and not order.
2. **DesignParam** - R List which consists of Design parameters which user may need to perform treatment selection. User should access the variables using names for ex. `DesignParam$SideType` and not order. For details of this list please see appropriate table in this section
3. **LookInfo** - R List which consists of Design Parameters related to two looks which user may need to perform treatment selection. User should access the variables using names for ex. `LookInfo$NumLooks` and not order. For details of this list please see appropriate table in this section

O R Functions

Table O.32: Function for treatment selection

| Suggested Name of the function | <i>TreatmentSelection()</i> | | | | | | | | | |
|--|--|---------------------|-------------|---------|----------------|-----------------------------------|----------------------|------------|-------------------------------|---------|
| Description | Performs treatment selection for combining p-values designs. This function is called once in each simulation after first look. | | | | | | | | | |
| Syntax | <i>TreatmentSelection(SimData, DesignParam, LookInfo)</i> | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Compulsory Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>SimData</td> <td>Simulated Data</td> </tr> <tr> <td>DesignParam</td> <td>Parameters of design</td> </tr> <tr> <td>LookInfo</td> <td>Look-wise information</td> </tr> </tbody> </table> | Compulsory Argument | Description | SimData | Simulated Data | DesignParam | Parameters of design | LookInfo | Look-wise information | |
| Compulsory Argument | Description | | | | | | | | | |
| SimData | Simulated Data | | | | | | | | | |
| DesignParam | Parameters of design | | | | | | | | | |
| LookInfo | Look-wise information | | | | | | | | | |
| Return Value Type | <p>R List</p> <p>The must identifiers in this list are</p> <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>TreatmentID</td> <td>An array of treatment identifiers</td> <td>Integer.</td> </tr> <tr> <td>AllocRatio</td> <td>An array of allocation ratios</td> <td>Double.</td> </tr> </tbody> </table> | Identifier | Description | Type | TreatmentID | An array of treatment identifiers | Integer. | AllocRatio | An array of allocation ratios | Double. |
| Identifier | Description | Type | | | | | | | | |
| TreatmentID | An array of treatment identifiers | Integer. | | | | | | | | |
| AllocRatio | An array of allocation ratios | Double. | | | | | | | | |
| Suggested format and additional information | <pre>TreatmentSelection ← function(SimData, DesignParam, Look- Info) { Error = 0 # Write the actual code here. # TreatmentID must contain values 1, 2, ... (No.ofTreatment - 1) # Allocation ratios are with respect to control # East expects TreatmentIDs sorted according to preference of treatment selection # Use appropriate error handling and modify the # Error appropriately. return(list(TreatmentID = as.integer(retval1), AllocRatio = as.double(retval2), ErrorCode = as.integer(Error))) }</pre> | | | | | | | | | |

Table O.33: DesignParam for Treatment Selection

| Argument Name | Description | Codes |
|----------------|---|---|
| Alpha | Type I Error | |
| Trial Type | Type of the trial | 0 - Superiority 1 - Non-inferiority 2 - Equivalence |
| Taile Type | Nature of critical region | 0 - Left tailed 1 - Right Tailed |
| SampleSize | Total Sample Size | |
| TestStatType | Test Statistic Type | 3 - Z-stat 4 - t-Stat |
| VarType | Variance Type | 4 - equal 5 - Un-equal 0 - Pooled 1 - Un-pooled |
| MultAdjMethod | Multiplicity adjustment method | 0 - Bonferonni 1 - Sidak 2 - Simes 3 - Dunnett |
| PValCombMethod | P-Value Combination Method | 0 - Inverse Normal |
| Sigma | Common Standard deviation or standard deviation array | |
| w1 | Weight for stage 1 | |
| w2 | Weight for stage 2 | |
| TestID | Test ID | 418 - DOM 419 - DOP |
| NumTreatments | Number of treatments including control | |

O R Functions

Table O.34: LookInfo for Treatment Selection

| Argument Name | Description | Codes |
|---------------|--------------------------------|---|
| NumLooks | Number of Looks | |
| CurrLookIndex | Current Look Index (1-based) | |
| InfoFrac | Array of Information fractions | |
| EffBdry | Array of Efficacy Boundaries | |
| FutBdryScale | Futility Boundary Scale | 1 - p-value scale 2 - Delta/Sigma Scale (DOM) or Delta Scale (DOP) |

***0.11 Functions
for Adaptive
Simulations***

This section describes details of various R functions for adaptive simulations.

R function can be used for performing sample size re-estimation. R function can be used along with CHW or CDL simulation but not with Muller and Schafer simulation.

R function assumes that Promising Zone scale is 'Conditional Power'. For Survival endpoint, R function can be used to re-estimate events. Whereas for Normal and Binary endpoints, R function can be used to re-estimate completers.

Even with R function, East will not allow reduction of planned events or completers and will not allow exceeding maximum feasible number of events or completers.

R function can also be used for computing cumulative Wald statistic in adaptive survival simulations.

O R Functions

Table O.35: Function for Re-estimating events

| Suggested Name of the function | <i>PerformSSR()</i> | | | | | | | | | | |
|--|--|---------------------|-------------|--------|--|---------------------|---|-------|---|-----------|-------------------------|
| Description | Performs re-estimation of events at adapt look in survival simulation. | | | | | | | | | | |
| Syntax | <i>PerformSSR(OrigCP, CPmin, CPmax, DesEvents)</i> | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Compulsory Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>OrigCP</td> <td>CP computed with design number of events</td> </tr> <tr> <td>CPmin</td> <td>Minimum CP threshold for promising zone</td> </tr> <tr> <td>CPmax</td> <td>Minimum CP threshold for promising zone</td> </tr> <tr> <td>DesEvents</td> <td>Design Number of Events</td> </tr> </tbody> </table> | Compulsory Argument | Description | OrigCP | CP computed with design number of events | CPmin | Minimum CP threshold for promising zone | CPmax | Minimum CP threshold for promising zone | DesEvents | Design Number of Events |
| Compulsory Argument | Description | | | | | | | | | | |
| OrigCP | CP computed with design number of events | | | | | | | | | | |
| CPmin | Minimum CP threshold for promising zone | | | | | | | | | | |
| CPmax | Minimum CP threshold for promising zone | | | | | | | | | | |
| DesEvents | Design Number of Events | | | | | | | | | | |
| Return Value Type | <p>R List</p> <p>The must identifiers in this list are</p> <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>ReEstEvents</td> <td>Re-estimated events</td> <td>Integer.</td> </tr> </tbody> </table> | Identifier | Description | Type | ReEstEvents | Re-estimated events | Integer. | | | | |
| Identifier | Description | Type | | | | | | | | | |
| ReEstEvents | Re-estimated events | Integer. | | | | | | | | | |
| Suggested format and additional information | <pre>PerformSSR ← function(OrigCP, CPmin, CPmax, DesEvents) { Error = 0 # Write the actual code here. # Use appropriate error handling and modify the # Error appropriately. return(list(ReEstEvents = as.integer(retval), ErrorCode = as.integer(Error))) }</pre> | | | | | | | | | | |

Table O.36: Function for Re-estimating Completers

| Suggested Name of the function | <i>PerformSSR()</i> | | | | | | | | | | |
|--|--|---------------------|-------------|--------|--|-------------------------|---|-------|---|---------------|-----------------------------|
| Description | Performs re-estimation of completers at adapt look in Normal and Binary adaptive simulations. | | | | | | | | | | |
| Syntax | <i>PerformSSR(OrigCP, CPmin, CPmax, DesCompleters)</i> | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Compulsory Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>OrigCP</td> <td>CP computed with design number of completers</td> </tr> <tr> <td>CPmin</td> <td>Minimum CP threshold for promising zone</td> </tr> <tr> <td>CPmax</td> <td>Minimum CP threshold for promising zone</td> </tr> <tr> <td>DesCompleters</td> <td>Design Number of Completers</td> </tr> </tbody> </table> | Compulsory Argument | Description | OrigCP | CP computed with design number of completers | CPmin | Minimum CP threshold for promising zone | CPmax | Minimum CP threshold for promising zone | DesCompleters | Design Number of Completers |
| Compulsory Argument | Description | | | | | | | | | | |
| OrigCP | CP computed with design number of completers | | | | | | | | | | |
| CPmin | Minimum CP threshold for promising zone | | | | | | | | | | |
| CPmax | Minimum CP threshold for promising zone | | | | | | | | | | |
| DesCompleters | Design Number of Completers | | | | | | | | | | |
| Return Value Type | R List The must identifiers in this list are <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>ReEstCompleters</td> <td>Re-estimated completers</td> <td>Integer.</td> </tr> </tbody> </table> | Identifier | Description | Type | ReEstCompleters | Re-estimated completers | Integer. | | | | |
| Identifier | Description | Type | | | | | | | | | |
| ReEstCompleters | Re-estimated completers | Integer. | | | | | | | | | |
| Suggested format and additional information | <pre>PerformSSR ← function(OrigCP, CPmin, CPmax, DesCompleters) { Error = 0 # Write the actual code here. # Use appropriate error handling and modify the # Error appropriately. return(list(ReEstCompleters = as.integer(retval1), ErrorCode = as.integer(Error))) }</pre> | | | | | | | | | | |

O R Functions

Table O.37: Computing cumulative Wald Statistic in Survival adaptive Simulations

| Suggested Name of the function | <i>CumWaldAdapt()</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------------------------------------|--|---------------------|-------------|---------|------------------|---------------------------|-------------------|-----------|-----------------------|------------|---------------------|-------------|------|----------|-------------------------|---------|---------------|-------------------------------------|----------|---------------|---|---------|---------------|---|---------|-------------|--------------------------------|--------|-----------------|-------------------------------|--------|
| Description | Computes cumulative Wald statistics at each look in CHW or CDL survival simulations. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Syntax | <i>CumWaldAdapt(SimData, DesignParam, LookInfo, AdaptParam)</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Arguments | <table border="1"> <thead> <tr> <th>Compulsory Argument</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>SimData</td> <td>Simulated Data</td> </tr> <tr> <td>DesignParam</td> <td>Design Parameters</td> </tr> <tr> <td>LookInfo</td> <td>Look-wise Information</td> </tr> <tr> <td>AdaptParam</td> <td>Adaptive Parameters</td> </tr> </tbody> </table> | Compulsory Argument | Description | SimData | Simulated Data | DesignParam | Design Parameters | LookInfo | Look-wise Information | AdaptParam | Adaptive Parameters | | | | | | | | | | | | | | | | | | | | |
| Compulsory Argument | Description | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| SimData | Simulated Data | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| DesignParam | Design Parameters | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| LookInfo | Look-wise Information | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| AdaptParam | Adaptive Parameters | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Return Value Type | <p>R List</p> <p>The must identifiers in this list are</p> <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>CumWaldStatistic</td> <td>Cumulative Wald Statistic</td> <td>Double.</td> </tr> <tr> <td>CumEvents</td> <td>Cumulative Events</td> <td>Integer.</td> </tr> </tbody> </table> <p>Optional identifiers in this list are</p> <table border="1"> <thead> <tr> <th>Identifier</th> <th>Description</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>LookTime</td> <td>Look Time for each Look</td> <td>Double.</td> </tr> <tr> <td>CumSampleSize</td> <td>Cumulative sample size at each Look</td> <td>Integer.</td> </tr> <tr> <td>CumEventsCtrl</td> <td>Cumulative Events on Control Arm at each Look</td> <td>Integer</td> </tr> <tr> <td>CumEventsTrmt</td> <td>Cumulative Events on Treatment Arm at each Look</td> <td>Integer</td> </tr> <tr> <td>AvgFollowUp</td> <td>Average Follow up at each Look</td> <td>Double</td> </tr> <tr> <td>AccrualDuration</td> <td>Accrual Duration at each Look</td> <td>Double</td> </tr> </tbody> </table> | Identifier | Description | Type | CumWaldStatistic | Cumulative Wald Statistic | Double. | CumEvents | Cumulative Events | Integer. | Identifier | Description | Type | LookTime | Look Time for each Look | Double. | CumSampleSize | Cumulative sample size at each Look | Integer. | CumEventsCtrl | Cumulative Events on Control Arm at each Look | Integer | CumEventsTrmt | Cumulative Events on Treatment Arm at each Look | Integer | AvgFollowUp | Average Follow up at each Look | Double | AccrualDuration | Accrual Duration at each Look | Double |
| Identifier | Description | Type | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CumWaldStatistic | Cumulative Wald Statistic | Double. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CumEvents | Cumulative Events | Integer. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Identifier | Description | Type | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| LookTime | Look Time for each Look | Double. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CumSampleSize | Cumulative sample size at each Look | Integer. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CumEventsCtrl | Cumulative Events on Control Arm at each Look | Integer | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CumEventsTrmt | Cumulative Events on Treatment Arm at each Look | Integer | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| AvgFollowUp | Average Follow up at each Look | Double | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| AccrualDuration | Accrual Duration at each Look | Double | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Table O.38: Computing cumulative Wald Statistic in Survival adaptive Simulations (Continued)

| | |
|--|---|
| Suggested format and additional information | <pre>CumWaldAdapt ← function(SimData, DesignParam, LookInfo, AdaptParam) { Error = 0 # Write the actual code here. # Use appropriate error handling and modify the # Error appropriately. return(list(CumWaldStatistic = as.double(retval1), CumEvents = as.integer(retval2), ErrorCode = as.integer(Error))) }</pre> |
|--|---|

O R Functions

O.12 Use of Initialization Function

O.12.1 Setting Seed

O.12.2 Setting Working Directory

O.12.3 Initialize Global Variable

This appendix provides more information on **Init(Seed)** function. This function will be optional. If provided, this function will be executed before executing any of the other user defined functions. User can use this function for various reasons. Below we list some of these.

O.12.1 Setting Seed

If user wants repeatability of the results for a run of simulations, he can set the seed using **set.seed** command inside this function. He can also choose the Random Number Generator as well as the method for Normal method generation. The default random number generator is "Mersenne-Twister" in R.

Example 1

Default random number generator will be used.

```
Init(Seed)
{
Error = 0
set.seed(seed = Seed)
return(as.integer(Error))
}
```

Example2

Wichmann Hill random number generator will be used.

```
Init(Seed)
{
Error = 0
set.seed(seed = Seed, kind = "Wichmann-Hill")
return(as.integer(Error))
}
```

O.12.2 Setting Working Directory

User can set the working directory. User may want to source the files he intends to use.

Example 1

```
Init(Seed)
{ Error = 0
setwd("E:\\Work\\East6.1")
source('ConstantsFile.R')
return(as.integer(Error))
}
```

0.12.3 Initialize Global Variable

User can initialize the global variables which may be used by his other R functions

Example 1

```

Init(Seed)
{ Error = 0
Tolerance ;;- 1e-6
NoIntervals ;;- 3
return(as.integer(Error))
}

```

0.13 Additional Arguments

Suppose for a user defined function f, the mandatory named arguments are **Arg1** and **Arg2**.

This function will be called as follows **f(Arg1 = Val1, Arg2 = Val2)** where Val1 and Val2 will be appropriately passed. Now user can have additional arguments for this function f, for example suppose he has additional arguments Arg3 and Arg4. The syntax for this function is `f ← function(Arg1, Arg3, Arg2, Arg4)`

```

{
Body of the function
}

```

Note that in the call to this function; only appropriate values will be passed to mandatory named arguments hence it is important that user initializes the other arguments.

Some of the ways to do this are

Initialize in the Definition

```

f ← function(Arg1, Arg3 = 2, Arg2, Arg4 = 5)
{
Body of the function
}

```

Initialize using Global Variables initialized in Init function.

```

f ← function(Arg1, Arg3 = Tolerance, Arg2, Arg4 = NoIntervals)
{
Body of the function
}

```

P *East 5.x to East 6.4 Import Utility*

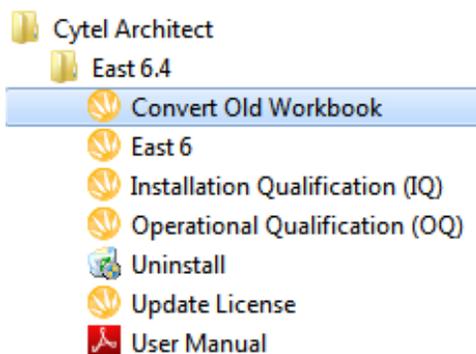
P.1 Import capabilities

This document serves the purpose of providing a step-by-step procedure as well as describing the scope of the **East 5.x to East 6.4 Import** Utility provided by Cytel to the East 6.4 Users. The Utility has been developed with a view to facilitate importing and converting the workbooks created in the earlier versions of East, namely the Microsoft Excel based **East 5.x** to the new architect based version of East namely, **East 6.4**. With the help of this Utility provided in the **All Program menu**, the East 6.4 user can now import the older workbooks and continue working on the imported designs for further development. For example, monitoring the design at subsequent interim looks or simulating the design is possible within the East 6.4 environment.

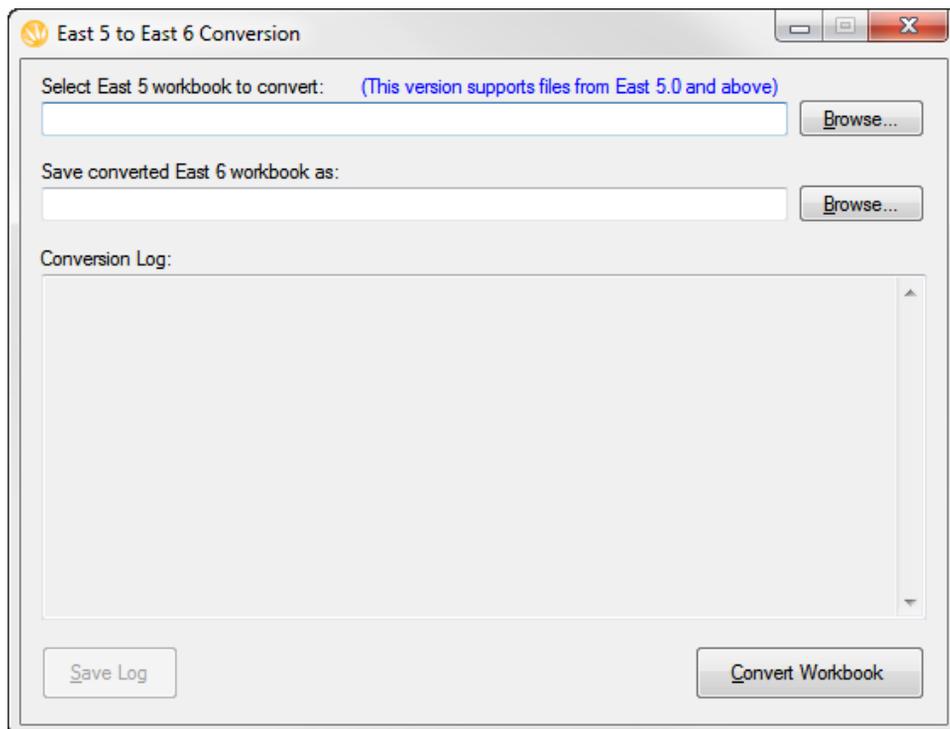
In order to open a workbook with the **.es5** extension given by East 5.x version, it must first be converted to a file with the **.cywx** extension that will be recognized by **East 6.4**.

From the **Start** Menu select:

All Programs → **Cytel Architect** → **East 6.4** → **Convert Old Workbook**



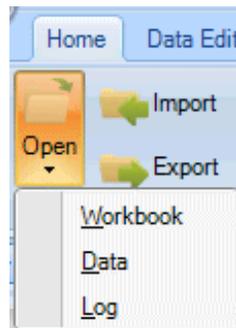
We can see the following window which accepts **East5.x** workbook as input and outputs a workbook of **East 6**. Click the **Browse** buttons to choose the **East 5.x** file to be converted and the file to be saved with **.cywx** extension of **East 6** version. To start the conversion click **Convert Workbook**:



The default location for the converted **East 6** workbook will be the same as that of the old workbook. You may select a different location of your choice for saving the same. While the conversion is in process, you will see a detailed log being displayed about the progress of the workbook creation. After completion of the conversion, you can save the log at the location of your choice.

Once complete, the file can be opened as a workbook in **East 6.4** as usual as shown below:

P East 5.x to East 6.4 Import Utility



When user imports an **East 5.x** workbook into **East 6.4**, East 6.4 will retain the input parameters and re-compute all output and make it available in the **6.4 (.cywx)** format. Since there have been major computational improvements from earlier versions of **East** to this version, some results may not match with those computed in **East 5.x**. In some rare situations, **East 6.4** will give a message that the input parameters are too extreme and it won't be able to import the workbooks. In general, user should be able to import any workbook created in **East 5.x** using any supported version of Excel into **East 6.4**. The list includes workbooks containing single look designs, group sequential designs, interim monitoring sheets, simulations etc. All supported locales will work including English (US/UK), French, Spanish, Japanese etc. However, there are some exceptions to the **Convert Old Workbook** functionality. These are described below:

1. **East 6.4** will not support importing of the following:
Direct monitoring, Basic simulations, Enhanced simulations with information scale, Adaptive worksheets for two-sided tests, expected sample size under H1/2, graph sheets and scratch sheets, interim monitoring sheets for single look designs.
2. Adaptive worksheets (like CHW simulations) for the odds ratio (OR) test from **East 5.0** will not be imported into **East 6.4** as **East 6.4** does not have adaptive features for this test yet. If user tries to import, **East 6.4** will display the following message: "CHW simulations are not available for this test in this version of **East**."
3. **East 5.x** allowed user to input floating point sample size / events value while

computing power of a design. If it is a group sequential design, **East 6.4** uses the option "Do not round sample size/events" to deal with the specified floating point value.

However, in case of some designs which are necessarily fixed look designs only, such as ratio of means, crossover designs, difference of means designs using t statistic etc, the option of using floating point input is not amenable by East 6.4. For such designs, East 6.4 will round down the sample size to the nearest integer for computing the power.

4. **East 6.4** won't import group sequential designs from **East 5.x** for the following tests:

- Linear Regression, Single Slope
- Linear Regression for Comparing Two Slopes
- Repeated Measures for Comparing Two Slopes

If user tries to import, East will display the following message:

"Group sequential option is not available for this test in this version of East."

East 6.4 supports only fixed sample (single look) designs for these tests.

5. **East 6.4** will not import **East 5.x** designs of the following type as these are not available in **East 6.4**:

- Logistic regression
- Cox proportional hazards regression

If user tries to import, East will display the following message:

"Unable to convert workbook as this test is not implemented in this version."

6. **East 6.4** will not import **East 5.x** designs with spending functions of type "Power Family" as these spending functions are not available in **East 6.4**. If user tries to import, East will display the following message:
"Power family spending function is not supported in this version."

7. Definition of treatment effect and effect size has been changed from **East 5.x** to **East 6.4** in the following cases:

In these cases, corresponding changes will be observed in the workbook after importing.

8. Muller and Schafer adaptive simulations performed with SWACI method in **East 5.x** workbooks will be run with BWCI method of estimation instead of SWACI

P East 5.x to East 6.4 Import Utility

Table P.1: Treatment effect in non-inferiority trials

| Test | East 5.x | East 6.4 |
|---|--|--|
| Difference of Means for Independent Data | $\delta = \mu_c - \mu_t$ | $\delta = \mu_t - \mu_c$ |
| Difference of Proportion for Independent Data | $\delta = \pi_c - \pi_t$ | $\delta = \pi_t - \pi_c$ |
| Odds Ratio of Proportion for Independent Data | $\psi = \frac{\pi_c(1 - \pi_t)}{\pi_t(1 - \pi_c)}$ | $\psi = \frac{\pi_t(1 - \pi_c)}{\pi_c(1 - \pi_t)}$ |

Table P.2: Longrank Test

| Test | East 5.x | East 6.4 |
|-----------------------------|---|--|
| Effect Size in Logrank Test | $\delta = -\ln\left(\frac{\lambda_t}{\lambda_c}\right)$ | $\delta = \ln\left(\frac{\lambda_t}{\lambda_c}\right)$ |

while importing to **East 6.4**. This is because **East 6.4** has replaced the SWACI method with the BWCI method as the latter is more advanced.

- East 6.4** will not import exact paired difference design from **East 5.x** as this design is not yet available in **East 6.4**. The **East 5.x** design is for the exact unconditional test for matched pairs whereas the design in **East 6.4** is for the exact McNemar's test which is a conditional test. If user tries to import the **East 5.x** design, **East 6.4** will display the following message:

"The exact unconditional test for matched pairs is not available in the current version of East. This workbook cannot be imported."

- While importing survival designs from **East 5.x**, **East 6.4** will convert input method to hazard rates if the **East 5.x** design was created with any other **input**

method.

11. In case of Logrank test with accrual rates and accrual duration, **East** first computes a range for the target accrual and when user specifies the committed accrual, East computes the study duration and other outputs. Because of computational improvements from **East 5.x** to **East 6.4**, the target accrual range in **East 6.4** could be a little different for the same design compared to **East 5.x**. If user has an **East 5.x** workbook where the committed accrual is equal or very close to the minimum, this workbook may not be imported in **East 6.4** as specified committed accrual may be less than the minimum accrual computed by **East 6.4**.
12. **East 6.4** will not import an **East 5.x** workbook if its file name contains the single quote (') character.

For technical support, please call us on 617-661-2011 or send a fax on 617-661-4405, or send email to support@cytel.com. Visit our website www.cytel.com for more information.

Q

Technical Reference and Formulas: Single Look Designs

Q.1 Introduction

In this Appendix, we provide theory used in the computation of single look designs in **East** and formulas used for computing sample size N (total number of subjects on the treatment arm in case of single arm studies, total number of pairs of subjects included in the study in paired designs and total number of subjects on the treatment and control arms both in case of two sample studies).

We begin with introducing common notations. The general method of computing sample size is solving the power equation for 'N' given other parameters such as δ , α , σ^2 . In a few cases, the procedure resorts to a closed form formula for the sample size. In rest of the cases, such a closed form expression for sample size is not possible. As a result, it requires use of an iterative method for computing the sample size for given power, starting with a sensible initial solution for N . In this Appendix, we describe the closed form solution wherever possible and in other cases state the initial solution for N along with the power equation used to derive the solution for N .

Q.2 Common Notation

Below we give notation which will be used throughout this chapter.

Common Notation

μ : Unknown mean of a single population

μ_0 : Mean response under Null hypothesis

S : Sample standard deviation

\bar{X} : Sample Mean

D : Difference variable of treatment and control when the response is continuous

\bar{D} : Sample Mean of D

σ_D : Population standard deviation of D

S_D : Sample standard deviation of D

λ : Median of the difference variable

μ_t : Unknown mean of treatment group

μ_c : Unknown mean of control group

σ : Population standard deviation

δ : Effect size, for example, difference of means, difference of proportions, log hazard ratio etc

SE : Standard Error

δ_0 : Non-inferiority margin for difference

ρ_0 : Non-inferiority margin for ratio

δ_L : Lower equivalence limit for difference

δ_U : Upper equivalence limit for difference

ρ_L : Lower equivalence limit for ratio

ρ_U : Upper equivalence limit for ratio
 $\phi(x)$: density function of standard normal variable, evaluated at x
 $\Phi(x)$: Distribution function of standard normal variable, evaluated at x
 Z_α : Upper α percent point of standard normal distribution
 $\tau_\nu(x)$: Distribution function of a student's t distribution, with ν degrees of freedom evaluated at x
 $\tau_\nu(x|\Omega)$: Distribution function of a non-central t distribution with ν degrees of freedom and non-centrality parameter Ω , evaluated at x
 $t_{\alpha,\nu}$: Upper α percent point of a student's t-distribution with ν degrees of freedom

**Q.3 Sample Size :
Continuous**

Q.3.1 Single:Sup:Normal

Q.3.2 Single:Sup:t

Q.3.3 Paired:Diff:Sup:
Normal

Q.3.4 Paired:Diff:Sup:t

Q.3.5 Paired:Diff:Noninf:
Normal

Q.3.6 Paired:Diff:
Noninf:t

Q.3.7 Paired:Diff:Equiv:t

Q.3.8 Paired:Ratios:
Sup:Normal

Q.3.9 Paired:Ratios:Sup:t

Q.3.10 Paired:Ratios:
Noninf:Normal

Q.3.11 Paired:Ratios:
Noninf:t

Q.3.12 Paired:Ratios:
Equiv:t

**Q.3.1 Single Arm Design : Single Mean : Superiority: Test Statistic
Distribution: Normal**

$$\sigma = \mu - \mu_0$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$N = \frac{\sigma^2}{\delta^2} (Z_\alpha + Z_\beta)^2$$

$$Power = 1 - \Phi \left(Z_\alpha - \frac{|\delta| \sqrt{N}}{\sigma} \right)$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2}{\delta^2} (Z_{\alpha/2} + Z_\beta)^2$$

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{\delta \sqrt{N}}{\sigma} \right)$$

Q Technical Reference and Formulas: Single Look Designs

$$+ \Phi \left(-Z_{\alpha/2} - \frac{\delta\sqrt{N}}{\sigma} \right)$$

- Two sided asymmetric (both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2}{\delta^2} (Z_{\alpha/2} + Z_{\beta})^2$$

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$\begin{aligned} \text{Power} &= 1 - \Phi \left(Z_{\alpha_u} - \frac{\delta\sqrt{N}}{\sigma} \right) \\ &+ \Phi \left(-Z_{\alpha_l} - \frac{\delta\sqrt{N}}{\sigma} \right) \end{aligned}$$

Q.3.2 Single Arm Design : Single Mean : Superiority: Test Statistic Distribution: t

$$\delta = \mu - \mu_0$$

- One sided (both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2}{\delta^2} (Z_{\alpha} + Z_{\beta})^2 + \frac{Z_{\alpha}^2}{2}$$

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$\text{Power} = 1 - \tau_{N-1} \left(t_{\alpha, N-1} \left| \frac{|\delta| \sqrt{N}}{\sigma} \right. \right)$$

- Two sided symmetric (both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2}{\delta^2} (Z_{\alpha} + Z_{\beta})^2 + \frac{Z_{\alpha}^2}{2}$$

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$\begin{aligned} Power &= 1 - \tau_{N-1} \left(t_{\frac{\alpha}{2}, N-1} \left| \frac{\delta\sqrt{N}}{\sigma} \right. \right) \\ &+ \tau_{N-1} \left(-t_{\frac{\alpha}{2}, N-1} \left| \frac{\delta\sqrt{N}}{\sigma} \right. \right) \end{aligned}$$

Q.3.3 Paired Design: Superiority: Test Statistic Distribution: Normal: Mean of paired differences

$$\delta = \mu_t - \mu_c$$

- One sided (both $\delta > 0$ and $\delta < 0$)

$$N = \frac{\sigma_D^2}{\delta^2} (Z_\alpha + Z_\beta)^2$$

$$Power = 1 - \Phi \left(Z_\alpha - \frac{|\delta|\sqrt{N}}{\sigma_D} \right)$$

- Two sided symmetric (both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma_D^2}{\delta^2} (Z_{\alpha/2} + Z_\beta)^2$$

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$\begin{aligned} Power &= 1 - \Phi \left(Z_{\frac{\alpha}{2}} - \frac{\delta\sqrt{N}}{\sigma_D} \right) \\ &+ \Phi \left(-Z_{\frac{\alpha}{2}} - \frac{\delta\sqrt{N}}{\sigma_D} \right) \end{aligned}$$

Q Technical Reference and Formulas: Single Look Designs

- Two sided asymmetric (both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma_D^2}{\delta^2} (Z_{\alpha/2} + Z_{\beta})^2$$

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$\begin{aligned} \text{Power} &= 1 - \Phi \left(Z_{\alpha_{\mu}} - \frac{\delta\sqrt{N}}{\sigma_D} \right) \\ &+ \Phi \left(-Z_{\alpha_l} - \frac{\delta\sqrt{N}}{\sigma_D} \right) \end{aligned}$$

Q.3.4 Paired Design: Superiority: Test Statistic Distribution: t

$$\delta = \mu_t - \mu_c$$

- One sided (both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2}{\delta^2} (Z_{\alpha} + Z_{\beta})^2 + \frac{Z_{\alpha}^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$\text{Power} = 1 - \tau_{N-1} \left(t_{\alpha, N-1} \left| \frac{|\delta|\sqrt{N}}{\sigma_D} \right. \right)$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2}{\delta^2} (Z_{\alpha} + Z_{\beta})^2 + \frac{Z_{\alpha}^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$\text{Power} = 1 - \tau_{N-1} \left(t_{\frac{\alpha}{2}, N-1} \left| \frac{\delta\sqrt{N}}{\sigma} \right. \right)$$

$$+ \tau_{N-1} \left(-t_{\frac{\alpha}{2}, N-1} \left| \frac{\delta \sqrt{N}}{\sigma} \right. \right)$$

Q.3.5 Paired Design : Non-inferiority: Test Statistic Distribution: Normal

$$\delta = \mu_t - \mu_c$$

- One sided (for both $\delta > \delta_0$ and $\delta < \delta_0$)

$$N = \frac{\sigma_D^2 (Z_\alpha + Z_\beta)^2}{(\delta - \delta_0)^2}$$

$$Power = 1 - \Phi \left(Z_\alpha - \frac{|\delta - \delta_0| \sqrt{N}}{\sigma_D} \right)$$

- Two sided symmetric (for both $\delta > \delta_0$ and $\delta < \delta_0$)
Start with the initial solution as

$$N = \frac{\sigma_D^2 (Z_{\alpha/2} + Z_\beta)^2}{(\delta - \delta_0)^2}$$

and solve using an iterative procedure the following equation, so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{(\delta - \delta_0) \sqrt{N}}{\sigma_D} \right) + \Phi \left(-Z_{\alpha/2} - \frac{(\delta - \delta_0) \sqrt{N}}{\sigma_D} \right)$$

- Two sided asymmetric (for both $\delta > \delta_0$ and $\delta < \delta_0$)
Start with the initial solution as

$$N = \frac{\sigma_D^2 (Z_{\alpha/2} + Z_\beta)^2}{(\delta - \delta_0)^2}$$

and solve using an iterative procedure the following equation, so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha_u} - \frac{(\delta - \delta_0) \sqrt{N}}{\sigma_D} \right) + \Phi \left(-Z_{\alpha_l} - \frac{(\delta - \delta_0) \sqrt{N}}{\sigma_D} \right)$$

Q Technical Reference and Formulas: Single Look Designs

Q.3.6 Paired Design : Non-inferiority: Test Statistic Distribution: t

$$\delta = \mu_t - \mu_c$$

- One sided (for both $\delta > 0$ and $\delta < 0$) Start with the initial solution as

$$N = \frac{\sigma^2}{(\delta - \delta_0)^2} (Z_\alpha + Z_\beta)^2 + \frac{Z_\alpha^2}{2}$$

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{N-1} \left(t_{\alpha, N-1} \left| \frac{|\delta - \delta_0| \sqrt{N}}{\sigma_D} \right. \right)$$

Q.3.7 Paired Design : Equivalence: Test Statistic Distribution: t

$$\delta = \mu_t - \mu_c$$

- Solve using an iterative procedure the following equation, so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{N-1} \left(t_{\alpha, N-1} \left| \frac{(\delta - \delta_L) \sqrt{N}}{\sigma_D} \right. \right) + \tau_{N-1} \left(-t_{\alpha, N-1} \left| \frac{(\delta - \delta_U) \sqrt{N}}{\sigma_D} \right. \right)$$

Q.3.8 Paired Design: Superiority: Test Statistic Distribution: Normal: Mean of Paired Ratios

$$\delta = \ln \left(\frac{\mu_t}{\mu_c} \right)$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$N = \frac{\sigma_D^2}{\delta^2} (Z_\alpha + Z_\beta)^2$$

$$Power = 1 - \Phi \left(Z_{\alpha} - \frac{|\delta| \sqrt{N}}{\sigma_D} \right)$$

Where σ_D = standard deviation of log ratios

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma_D^2}{\delta^2} (Z_{\alpha/2} + Z_{\beta})^2$$

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{\delta \sqrt{N}}{\sigma_D} \right) + \Phi \left(-Z_{\alpha/2} - \frac{\delta \sqrt{N}}{\sigma_D} \right)$$

Where σ_D = standard deviation of log ratios

- Two sided asymmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma_D^2}{\delta^2} (Z_{\alpha/2} + Z_{\beta})^2$$

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha_u} - \frac{\delta \sqrt{N}}{\sigma_D} \right) + \Phi \left(-Z_{\alpha_l} - \frac{\delta \sqrt{N}}{\sigma_D} \right)$$

Where σ_D = standard deviation of log ratios.

Q.3.9 Paired Design: Superiority: Test Statistic Distribution: t

Mean of paired ratios: $\delta = \ln \left(\frac{\mu_t}{\mu_c} \right)$

- One sided (for both $\delta > 0$ and $\delta < 0$) Start with the initial solution as

$$N = \frac{\sigma^2}{\delta^2} (Z_{\alpha} + Z_{\beta})^2 + \frac{Z_{\alpha}^2}{2}$$

Q Technical Reference and Formulas: Single Look Designs

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{N-1} \left(t_{\alpha, N-1} \left| \frac{|\delta| \sqrt{N}}{\sigma_D} \right. \right)$$

where σ_D = standard deviation of log ratios

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$).
Start with the initial solution as

$$N = \frac{\sigma^2}{\delta^2} (Z_\alpha + Z_\beta)^2 + \frac{Z_\alpha^2}{2}$$

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{N-1} \left(t_{\frac{\alpha}{2}, N-1} \left| \frac{\delta \sqrt{N}}{\sigma_D} \right. \right) + \tau_{N-1} \left(-t_{\frac{\alpha}{2}, N-1} \left| \frac{\delta \sqrt{N}}{\sigma_D} \right. \right)$$

where σ_D = standard deviation of log ratios.

Q.3.10 Paired Design : Non-inferiority: Test Statistic Distribution: Normal

$$\delta = \ln \left(\frac{\mu_t}{\mu_c} \right)$$

- One sided (for both $\delta > \delta_0$ and $\delta < \delta_0$)

$$N = \frac{\sigma_D^2 (Z_\alpha + Z_\beta)^2}{(\delta - \delta_0)^2}$$

$$Power = 1 - \Phi \left(Z_\alpha - \frac{|\delta - \delta_0| \sqrt{N}}{\sigma_D} \right)$$

where $\delta_0 = \log(\rho_0)$ and σ_D = standard deviation of log ratios

Q.3.11 Paired Design : Non-inferiority: Test Statistic Distribution: t

$$\delta = \ln \left(\frac{\mu_t}{\mu_c} \right)$$

- One sided (for both $\delta > 0$ and $\delta < 0$) Start with the initial solution as

$$N = \frac{\sigma^2}{(\delta - \delta_0)^2} (Z_\alpha + Z_\beta)^2 + \frac{Z_\alpha^2}{2}$$

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{N-1} \left(t_{\alpha, N-1} \left| \frac{|\delta - \delta_0| \sqrt{N}}{\sigma_D} \right. \right)$$

where $\delta_0 = \ln(\rho_0)$ and $\sigma_D =$ standard deviation of log ratios

Q.3.12 Paired Design : Equivalence: Test Statistic Distribution: t

$$\delta = \ln \left(\frac{\mu_t}{\mu_c} \right)$$

- Solve using an iterative procedure the following equation, so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{N-1} \left(t_{\alpha, N-1} \left| \frac{(\delta - \delta_L) \sqrt{N}}{\sigma_D} \right. \right) + \tau_{N-1} \left(-t_{\alpha, N-1} \left| \frac{(\delta - \delta_U) \sqrt{N}}{\sigma_D} \right. \right)$$

where $\delta_0 = \ln(\rho_0)$ and $\sigma_D =$ standard deviation of log ratios

Q Technical Reference and Formulas: Single Look Designs

Q.4 Sample Size : Continuous:Two Samples

Q.4.1 Diff:Sup:Normal

Q.4.2 Diff:Sup:t:Var Equal

Q.4.3 Diff:Sup:t:Var
Unequal

Q.4.4 Diff:Noninf:Normal

Q.4.5 Diff:Noninf:t:Var
Equal

Q.4.6 Diff:Noninf:t:Var
Unequal

Q.4.7 Diff:Equiv:t

Q.4.8 Ratios:Sup:Normal

Q.4.9 Ratios:Sup:t:Var
Equal

Q.4.10 Ratios:Noninf:Normal

Q.4.11 Ratios: Noninf:t

Q.4.12 Ratios:Equiv:t

Q.4.13 Wilcoxon Mann
Whitney Test

Q.4.1 Two Independent Samples:Superiority:Test Statistic Dist: Normal

$$\delta = \mu_t - \mu_c, TF = \frac{N_t}{N}, \sigma = \text{common s.d.}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$N = \frac{\sigma^2 (Z_\alpha + Z_\beta)^2}{\delta^2 * TF * (1 - TF)}$$

$$Power = 1 - \Phi \left(Z_\alpha - \frac{\delta \sqrt{N * TF * (1 - TF)}}{\sigma} \right)$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2 (Z_{\frac{\alpha}{2}} + Z_\beta)^2}{\delta^2 * TF * (1 - TF)}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{\delta \sqrt{N * TF * (1 - TF)}}{\sigma} \right) + \Phi \left(-Z_{\alpha/2} - \frac{\delta \sqrt{N * TF * (1 - TF)}}{\sigma} \right)$$

- Two sided asymmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2 (Z_{\frac{\alpha}{2}} + Z_\beta)^2}{\delta^2 * TF * (1 - TF)}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha_u} - \frac{\delta \sqrt{N * TF * (1 - TF)}}{\sigma} \right)$$

$$+\Phi\left(-Z_{\alpha_t}-\frac{\delta\sqrt{N*TF*(1-TF)}}{\sigma}\right)$$

Q.4.2 Two Independent Samples: Superiority: Test Statistic Distribution: t: Variance : Equal

$$\delta = \mu_t - \mu_c, TF = \frac{N_t}{N}, \sigma = \text{common s.d.}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2 TF^2}{\sigma^2(TF - 1)}(Z_{\alpha} + Z_{\beta})^2 + \frac{Z_{\alpha}^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t+n_c-2}\left(t_{\alpha, n_t+n_c-2} \left| \frac{|\delta| \sqrt{N(TF - 1)}}{\sigma * TF} \right.\right)$$

- Two sided (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2 TF^2}{\sigma^2(TF - 1)}(Z_{\alpha/2} + Z_{\beta})^2 + \frac{Z_{\alpha/2}^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t+n_c-2}\left(t_{\frac{\alpha}{2}, n_t+n_c-2} \left| \frac{|\delta| \sqrt{N(TF - 1)}}{\sigma * TF} \right.\right) + \tau_{n_t+n_c-2}\left(-t_{\frac{\alpha}{2}, n_t+n_c-2} \left| \frac{|\delta| \sqrt{N(TF - 1)}}{\sigma * TF} \right.\right)$$

Q Technical Reference and Formulas: Single Look Designs

Q.4.3 Two Independent Samples: Superiority: Test Statistic Distribution: t Variance : Unequal

$$\delta = \mu_t - \mu_c, TF = \frac{N_t}{N},$$

σ_t and σ_c are s.d.s for treatment and control respectively

- One sided (for both $\delta > 0$ and $\delta < 0$)
Start with a relevant initial solution and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_\nu \left(t_{\alpha, \nu} \left| \frac{|\delta|}{\sqrt{\frac{\sigma_t^2}{n_t} + \frac{\sigma_c^2}{n_c}}} \right. \right)$$

where the d.f.v. are given by:

$$\nu = \frac{\left(\frac{\sigma_t^2}{n_t} + \frac{\sigma_c^2}{n_c}\right)^2}{\frac{\left(\frac{\sigma_t^2}{n_t}\right)^2}{n_t - 1} + \frac{\left(\frac{\sigma_c^2}{n_c}\right)^2}{n_c - 1}}$$

Q.4.4 Two Independent Samples: Non-inferiority : Test Statistic Distribution: Normal

- One sided (for both $\delta > \delta_0$ and $\delta < \delta_0$)

$$N = \frac{\sigma^2(Z_\alpha + Z_\beta)^2}{(\delta - \delta_0)^2 * TF * (1 - TF)}$$

Q.4.5 Two Independent Samples: Non-inferiority : Test Statistic Distribution: t Variance : Equal

- One sided (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2 TF^2}{(\delta - \delta_0)^2 (TF - 1)} (Z_\alpha + Z_\beta)^2 + \frac{Z_\alpha^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t + n_c - 2} \left(t_{\alpha, n_t + n_c - 2} \left| \frac{|\delta - \delta_0| \sqrt{N(TF - 1)}}{\sigma * TF} \right| \right)$$

Q.4.6 Two Independent Samples: Non-inferiority : Test Statistic Distribution: t: Variance : Unequal

- One sided (for both $\delta > 0$ and $\delta < 0$)
Start with a relevant initial solution and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{\nu} \left(t_{\alpha, \nu} \left| \frac{|\delta - \delta_0|}{\sqrt{\frac{\sigma_t^2}{n_t} + \frac{\sigma_c^2}{n_c}}} \right| \right)$$

where d.f.is given by:

$$\nu = \frac{\left(\frac{\sigma_t^2}{n_t} + \frac{\sigma_c^2}{n_c} \right)^2}{\frac{\left(\frac{\sigma_t^2}{n_t} \right)^2}{n_t - 1} + \frac{\left(\frac{\sigma_c^2}{n_c} \right)^2}{n_c - 1}}$$

Q.4.7 Two Independent Samples: Equivalence: Test Statistic Distribution: t

$$\delta = \mu_t - \mu_c$$

Solve using an iterative procedure the following equation, so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t + n_c - 2} \left(t_{\alpha, n_t + n_c - 2} \left| \frac{|\delta - \delta_L| \sqrt{N(TF - 1)}}{\sigma * TF} \right| \right) + \tau_{n_t + n_c - 2 - 1} \left(-t_{\alpha, n_t + n_c - 2} \left| \frac{|\delta - \delta_U| \sqrt{N(TF - 1)}}{\sigma * TF} \right| \right)$$

Q Technical Reference and Formulas: Single Look Designs

Q.4.8 Two Independent Samples: Superiority: Test Statistic Distribution: Normal: Variance: Equal

$$\delta = \ln(\mu_t/\mu_c), TF = \frac{N_t}{n}$$

cv = Coefficient of variation of the original data is the input.

σ = common standard deviation of log ratios = $\sqrt{\ln(CV^2) + 1}$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$N = \frac{\sigma^2(Z_\alpha + Z_\beta)^2}{\delta^2 * TF * (1 - TF)}$$

$$Power = 1 - \Phi\left(Z_\alpha - \frac{|\delta| \sqrt{N * TF * (1 - TF)}}{\sigma}\right)$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)

Start with the initial solution as

$$N = \frac{\sigma^2(Z_{\alpha/2} + Z_\beta)^2}{\delta^2 * TF * (1 - TF)}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi\left(Z_{\alpha/2} - \frac{\delta \sqrt{N * TF * (1 - TF)}}{\sigma}\right)$$

$$+ \Phi\left(-Z_{\alpha/2} - \frac{\delta \sqrt{N * TF * (1 - TF)}}{\sigma}\right)$$

- Two sided asymmetric (for both $\delta > 0$ and $\delta < 0$)

Start with the initial solution as

$$N = \frac{\sigma^2(Z_{\alpha/2} + Z_\beta)^2}{\delta^2 * TF * (1 - TF)}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha_u} - \frac{\delta \sqrt{N * TF * (1 - TF)}}{\sigma} \right) + \Phi \left(-Z_{\alpha_l} - \frac{\delta \sqrt{N * TF * (1 - TF)}}{\sigma} \right)$$

Q.4.9 Two Independent Samples: Superiority: Test Statistic Distribution: t:Variance : Equal

$$\delta = \ln(\mu_t / \mu_c), TF = \frac{N_t}{N}$$

CV = Coefficient of variation of the original data is the input.

$$\sigma = \text{common standard deviation of log ratios} = \sqrt{\ln(CV^2) + 1}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2 TF^2}{\delta^2 (TF - 1)} (Z_{\alpha} + Z_{\beta})^2 + \frac{Z_{\alpha}^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t + n_c - 2} \left(t_{\alpha, n_t + n_c - 2} \left| \frac{|\delta| \sqrt{N(TF - 1)}}{\sigma * TF} \right. \right)$$

- Two sided (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2 TF^2}{\delta^2 (TF - 1)} (Z_{\alpha/2} + Z_{\beta})^2 + \frac{Z_{\alpha/2}^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t + n_c - 2} \left(t_{\alpha/2, n_t + n_c - 2} \left| \frac{\delta \sqrt{N(TF - 1)}}{\sigma * TF} \right. \right) +$$

Q Technical Reference and Formulas: Single Look Designs

$$\tau_{n_t+n_c-2} \left(-t_{\alpha/2, n_t+n_c-2} \left| \frac{\delta \sqrt{N(TF-1)}}{\sigma * TF} \right. \right)$$

Q.4.10 Two Independent Samples: Non-inferiority : Test Statistic Distribution: Normal

$$\delta = \ln(\mu_t/\mu_c), TF = \frac{N_t}{N}$$

CV = Coefficient of variation of the original data is the input.

$$\sigma = \text{common standard deviation of log ratios} = \sqrt{\ln(CV^2) + 1}$$

- One sided (for both $\delta > \delta_0$ and $\delta < \delta_0$)

$$N = \frac{\sigma^2(Z_\alpha + Z_\beta)^2}{(\delta - \delta_0)^2 * TF * (1 - TF)}$$

where $\delta_0 = \ln(\rho_0)$ and σ = standard deviation of log ratios

Q.4.11 Two Independent Samples: Non-inferiority : Test Statistic Distribution: t

$$\delta = \ln(\mu_t/\mu_c), TF = \frac{N_t}{N}$$

CV = Coefficient of variation of the original data is the input.

$$\sigma = \text{common standard deviation of log ratios} = \sqrt{\ln(CV^2) + 1}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

Start with the initial solution as

$$N = \frac{\sigma^2 TF^2}{(\delta - \delta_0)^2 (1 - TF)} (Z_\alpha + Z_\beta)^2 + \frac{Z_\alpha^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$\text{Power} = 1 - \tau_{n_t+n_c-2} \left(t_{\alpha, n_t+n_c-2} \left| \frac{|\delta - \delta_0| \sqrt{N(1 - TF)}}{\sigma * TF} \right. \right)$$

where $\delta_0 = \ln(\rho_0)$

Q.4.12 Two Independent Samples: Equivalence : Test Statistic Distribution: t

$$\delta = \ln(\mu_t/\mu_c), TF = \frac{N_t}{N}$$

CV = Coefficient of variation of the original data is the input.

$$\sigma = \text{common standard deviation of log ratios} = \sqrt{\ln(CV^2) + 1}$$

Solve using an iterative procedure the following equation, so that the computed power matches with the desired power with 1.e-6 precision.

$$\begin{aligned} Power = 1 - \tau_{n_t+n_c-2} \left(t_{\alpha, n_t+n_c-2} \left| \frac{|\delta - \delta_L| \sqrt{N(1-TF)}}{\sigma * TF} \right. \right) + \\ \tau_{n_t+n_c-2-1} \left(-t_{\alpha, n_t+n_c-2} \left| \frac{|\delta - \delta_U| \sqrt{N(1-TF)}}{\sigma * TF} \right. \right) \end{aligned}$$

Q.4.13 Two Independent Samples: Wilcoxon Mann Whitney Test

x_1, x_2, \dots, x_{n_c} observations from Control

x_1, x_2, \dots, x_{n_t} observations from Treatment

$$r = \frac{n_t}{N}$$

θ = treatment effect

Test Statistic

$$U_1 = R_1 - \frac{n_c(n_c+1)}{2} \sim AN(\mu_U, \mu_U^2)$$

where

R1= Sum of ranks of control population in the combined sample

$$\mu_U = \frac{n_c n_t}{2} \text{ and } \mu_U^2 = \frac{n_c n_t (n_c + n_t + 1)}{12}$$

- One sided
 $H_0 : \theta = 0$ against $H_1 : \theta > 0$; Y observations tend to be larger than X observations
 Sample Size

$$N = \frac{(Z_\alpha + Z_\beta)^2}{12r(1-r)(p-0.5)^2}$$

Q Technical Reference and Formulas: Single Look Designs

where

$p = P(X < Y) = \Phi\left(\frac{\mu_t - \mu_c}{\sqrt{2}\sigma}\right)$ assuming that the observations come from Normal distributions with common standard deviation σ .

- Two sided
 $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$
 Sample Size

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{12r(1-r)(p-0.5)^2}$$

Q.5 Sample Size : Continuous : Crossover Designs : Two Samples

Q.5.1 Crossover:Sup:t

Q.5.2 Crossover:Noninf:t

Q.5.3 Crossover: Equiv:t

Q.5.4 Crossover:Sup:t

Q.5.5 Crossover: Noninf:t

Q.5.6 Crossover:Equiv:t

Q.5.1 Crossover Designs :Superiority : Test Statistic Distribution: t

$$\delta = \mu_t - \mu_c, TF = \frac{N_t}{N}, \sigma = \sqrt{MSE}$$

$\sigma_D = \sqrt{2}\sqrt{MSE}$ = s.d. of difference of treatment effects

- One sided (for both $\delta > 0$ and $\delta < 0$)
 Start with the initial solution as

$$N = \frac{\sigma^2 TF^2}{\delta^2 (TF - 1)} (Z_{\alpha} + Z_{\beta})^2 + \frac{Z_{\alpha}^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t+n_c-2} \left(t_{\alpha, n_t+n_c-2} \left| \frac{|\delta| \sqrt{2N(1-TF)}}{\sigma * TF} \right. \right)$$

- Two sided (for both $\delta > 0$ and $\delta < 0$)
 Start with the initial solution as

$$N = \frac{\sigma^2 TF^2}{2\delta^2 (1-TF)} (Z_{\alpha/2} + Z_{\beta})^2 + \frac{Z_{\alpha/2}^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t+n_c-2} \left(t_{\frac{\alpha}{2}, n_t+n_c-2} \left| \frac{|\delta| \sqrt{2N(1-TF)}}{\sigma * TF} \right. \right) +$$

$$\tau_{n_t+n_c-2} \left(-t_{\frac{\alpha}{2}, n_t+n_c-2} \left| \frac{|\delta| \sqrt{2N(1-TF)}}{\sigma * TF} \right. \right)$$

Q.5.2 Crossover Designs :Noninferiority : Test Statistic Distribution:t

$$\delta = \mu_t - \mu_c, TF = \frac{N_t}{N}, \sigma = \sqrt{MSE}$$

$$\sigma_D = \sqrt{2}\sqrt{MSE} = \text{s.d. of difference of treatment effects}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2 TF^2}{2(\delta - \delta_0)^2 (1 - TF)} (Z_\alpha + Z_\beta)^2 + \frac{Z_\alpha^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t+n_c-2} \left(t_{\alpha, n_t+n_c-2} \left| \frac{|\delta - \delta_0| \sqrt{2N(1-TF)}}{\sigma * TF} \right. \right)$$

Q.5.3 Crossover Designs :Equivalence : Test Statistic Distribution: t

$$\delta = \mu_t - \mu_c, TF = \frac{N_t}{N}, \sigma = \sqrt{MSE}$$

$$\sigma_D = \sqrt{2}\sqrt{MSE} = \text{s.d. of difference of treatment effects.}$$

Solve using an iterative procedure using the following equation, so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t+n_c-2} \left(t_{\alpha, n_t+n_c-2} \left| \frac{|\delta - \delta_L| \sqrt{2N(1-TF)}}{\sigma * TF} \right. \right) +$$

$$\tau_{n_t+n_c-2} \left(-t_{\alpha, n_t+n_c-2} \left| \frac{|\delta - \delta_U| \sqrt{N(1-TF)}}{\sigma * TF} \right. \right)$$

Q.5.4 Crossover Designs: Superiority: Test Statistic Distribution: t

$$\delta = \ln(\mu_t/\mu_c), TF = \frac{N_t}{N},$$

Q Technical Reference and Formulas: Single Look Designs

$$\sigma = \sqrt{MSE \log}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2 TF^2}{2\delta^2(1 - TF)} (Z_\alpha + Z_\beta)^2 + \frac{Z_\alpha^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t+n_c-2} \left(t_{\alpha, n_t+n_c-2} \left| \frac{|\delta| \sqrt{2N(1 - TF)}}{\sigma * TF} \right. \right)$$

- Two sided (for both $\delta > 0$ and $\delta < 0$)
Start the initial solution as

$$N = \frac{\sigma^2 TF^2}{2\delta^2(TF - 1)} (Z_{\alpha/2} + Z_\beta)^2 + \frac{Z_{\alpha/2}^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t+n_c-2} \left(t_{\frac{\alpha}{2}, n_t+n_c-2} \left| \frac{\delta \sqrt{2N(1 - TF)}}{\sigma * TF} \right. \right) + \tau_{n_t+n_c-2} \left(-t_{\frac{\alpha}{2}, n_t+n_c-2} \left| \frac{\delta \sqrt{2N(1 - TF)}}{\sigma * TF} \right. \right)$$

Q.5.5 Crossover Designs :Noninferiority : Test Statistic Distribution: t

$$\delta = \ln(\mu_t/\mu_c), TF = \frac{N_t}{N}$$

$$\sigma = \sqrt{MSE \log}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2 TF^2}{2(\delta - \delta_0)^2(1 - TF)} (Z_\alpha + Z_\beta)^2 + \frac{Z_\alpha^2}{2}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t+n_c-2} \left(t_{\alpha, n_t+n_c-2} \left| \frac{|\delta - \delta_0| \sqrt{2N(1 - TF)}}{\sigma * TF} \right. \right)$$

Where $\delta_0 = \ln(\rho_0)$

Q.5.6 Crossover Designs :Equivalence : Test Statistic Distribution: t

$$\delta = \ln(\mu_t/\mu_c), TF = \frac{N_t}{N}$$

$$\sigma = \sqrt{MSE \log}$$

Solve using an iterative procedure the following equation, so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \tau_{n_t+n_c-2} \left(t_{\alpha, n_t+n_c-2} \left| \frac{|\delta - \delta_L| \sqrt{2N(1 - TF)}}{\sigma * TF} \right. \right) \\ + \tau_{n_t+n_c-2-1} \left(-t_{\alpha, n_t+n_c-2} \left| \frac{|\delta - \delta_U| \sqrt{N(1 - TF)}}{\sigma * TF} \right. \right)$$

where $\delta_L = \ln(\rho_L)$ and $\delta_U = \ln(\rho_U)$.

Q Technical Reference and Formulas: Single Look Designs

Q.6 Sample Size : Continuous : Many Samples

Q.6.1 One Way ANOVA

Q.6.2 One Way Contrast:t

Q.6.3 One Way Repeated :
ANOVA

Q.6.4 One Way Repeated
Measures Contrast

Q.6.5 Two Way ANOVA

Q.6.6 Linear regression
single slope

Q.6.7 Linear Regression:
Diff. of slopes

Q.6.8 Repeated measures:
Diff. of slopes

Q.6.1 One Way ANOVA : Superiority: Test Statistic Distribution: F

σ = Common standard deviation

σ_m^2 = Variance of means

r = Number of groups.

Solve for n using an iterative procedure the following equation, so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = P_{\lambda}(F > F_{1,(r-1)(n-r),\alpha})$$

with non-centrality parameter

$$\lambda = \frac{n \sigma_m^2}{\sigma^2}$$

Q.6.2 One Way ANOVA : Single One Way Contrast: t

σ = Common standard deviation

σ_{mc}^2 = Variance of means

r = Number of groups

- One sided

Solve for n using an iterative procedure the following equation, so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = P_{\lambda_1}(t > t_{n-r,\alpha})$$

with non-centrality parameter

$$\lambda_1 = \sqrt{n} \frac{\sigma_{mc}}{\sigma}$$

- Two Sided

Solve for n using an iterative procedure the following equation, so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = P_{\lambda}(F > F_{1,(n-r),\alpha})$$

with non-centrality parameter

$$\lambda = \frac{n \sigma_{mc}^2}{\sigma^2}$$

Q.6.3 One Way Repeated Measures: ANOVA: Superiority: Constant Correlation

M=number of levels

μ_i = mean at level I

σ = standard deviation at each level

ρ = between level correlation

$\sigma_m^2 = \frac{\sum (\mu_i - \bar{\mu})^2}{M}$ = variance of means

Effective size = $\Delta = \frac{\sigma_m^2}{\sigma^2(1-\rho)}$

Power = $P_\lambda(F > F_{(M-1), (M-1)(n-1), \alpha})$ with noncentrality parameter $\lambda = nM\Delta$

Q.6.4 One Way Repeated Measures Contrast

M=number of levels

μ_i = mean at level i

σ = standard deviation at each level

ρ = between level correlation

Contrast $C = \sum C_i \mu_i$ such that $\sum C_i = 0$

$D = \sqrt{\sum C_i^2}$

Effective size = $\Delta = \frac{|C|}{\sigma D \sqrt{1-\rho}}$

Power = $P_\lambda(F > F_{1, (M-1)(n-1), \alpha})$ with noncentrality parameter $\lambda = n \Delta^2$

Q.6.5 Two Way ANOVA

r = number of factor A levels,

Q Technical Reference and Formulas: Single Look Designs

s = number of factor B levels,

μ = overall mean

σ = common s.d. in each of the groups

μ_i = mean across factor A levels for factor A level i

μ_j = mean across factor B levels for factor B level j

μ_{ij} = mean for factor A level i and factor B level j

V_A = Variance of the marginal means for factor A

$$V_A = \frac{\sum_i (\mu_i - \mu)^2}{r}$$

V_B = Variance of the marginal means for factor B

$$V_B = \frac{\sum_j (\mu_j - \mu)^2}{r}$$

V_{AB} = Variance of cell means for factor A and B

$$V_{AB} = \frac{\sum_i \sum_j (\mu_{ij} - \mu_i - \mu_j + \mu)^2}{rs}$$

$Power_A = P(F > F_{(r-1),rs,(n-1),\alpha})$ with non-centrality parameter $\lambda = nrs \frac{V_A}{\sigma^2}$

$Power_B = P(F > F_{(s-1),rs,(n-1),\alpha})$ with non-centrality parameter $\lambda = nrs \frac{V_B}{\sigma^2}$

$Power_{AB} = P(F > F_{(r-1)(s-1),rs,(n-1),\alpha})$ with non-centrality parameter

$$\lambda = nrs \frac{V_{AB}}{\sigma^2}$$

Q.6.6 Linear regression single slope

Use $\delta = \theta - \theta_0$ and $\sigma = \frac{s.d.of\ residual}{s.d.of X} = \frac{\sigma_\epsilon}{\sigma_x}$ throughout the computation of power, sample size .

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$N = \frac{\sigma^2}{\delta^2} (Z_\alpha + Z_\beta)^2$$

$$Power = 1 - \Phi \left(Z_\alpha - \frac{\delta \sqrt{N}}{\sigma} \right)$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2}{\delta^2} (Z_{\alpha/2} + Z_{\beta})^2$$

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{\delta\sqrt{N}}{\sigma} \right) + \Phi \left(-Z_{\alpha/2} - \frac{\delta\sqrt{N}}{\sigma} \right)$$

Q.6.7 Linear Regression : Difference of slopes

Use $\delta = \theta_t - \theta_c$, $TF = \frac{N_t}{N}$, $\sigma = \sigma_e \sqrt{\frac{(1-TF)*\sigma_{xc}^2 + (1-TF)*\sigma_{xt}^2}{\sigma_{xc}^2 \sigma_{xt}^2}}$

where

σ_{xc} = Std dev of X under control

σ_{xt} = Std dev of X under treatment

σ_e = Std dev of residuals

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$N = \frac{\sigma^2 (Z_{\alpha} + Z_{\beta})^2}{\delta^2 * TF * (1 - TF)}$$

$$Power = 1 - \Phi \left(Z_{\alpha} - \frac{|\delta| \sqrt{N * TF * (1 - TF)}}{\sigma} \right)$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$) Start with the initial solution as

$$N = \frac{\sigma^2 (Z_{\alpha/2} + Z_{\beta})^2}{\delta^2 * TF * (1 - TF)}$$

and solve using an iterative procedure following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{\delta\sqrt{N * TF * (1 - TF)}}{\sigma} \right) + \Phi \left(-Z_{\alpha/2} - \frac{\delta\sqrt{N * TF * (1 - TF)}}{\sigma} \right)$$

Q Technical Reference and Formulas: Single Look Designs

Q.6.8 Repeated measures: Difference of slopes

Use $\delta = \theta_t - \theta_c$, $TF = \frac{N_t}{N}$, $\sigma = \sqrt{\sigma_b^2 + \frac{12(M-1)\sigma_w^2}{M(M-1)S^2}}$ throughout the computation of power, sample size, alpha and delta.

Where

M = Number of measurements

S = Duration of follow up

σ_w = Within subject std. dev

σ_b = Between subject std. dev

σ_e = Std dev of residuals

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$N = \frac{\sigma^2(Z_\alpha + Z_\beta)^2}{\delta^2 * TF * (1 - TF)}$$

$$Power = 1 - \Phi \left(Z_\alpha - \frac{|\delta| \sqrt{N * TF * (1 - TF)}}{\sigma} \right)$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\sigma^2(Z_{\alpha/2} + Z_\beta)^2}{\delta^2 * TF * (1 - TF)}$$

and solve using an iterative procedure using the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{\delta \sqrt{N * TF * (1 - TF)}}{\sigma} \right) + \Phi \left(-Z_{\alpha/2} - \frac{\delta \sqrt{N * TF * (1 - TF)}}{\sigma} \right)$$

**Q.7 Sample Size :
Discrete**

Q.7.1 Single Prop:
Sup:Null

Q.7.2 Single Prop:
Sup:Empirical

Q.7.3 Paired:Sup:
McNemar

Q.7.1 Single Arm Design : Single Proportion : Superiority: Test Statistic Distribution: Normal:Variance: Under Null hypothesis

$$\delta = \pi_1 - \pi_0, \Delta = \sqrt{\frac{\pi_0(1 - \pi_0)}{\pi_1(1 - \pi_1)}}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$N = \frac{\pi_1(1 - \pi_1)}{\delta^2} (Z_\beta + \Delta Z_\alpha)^2$$

$$Power = 1 - \Phi \left(\left(Z_\alpha - \frac{|\delta| \sqrt{N}}{\sqrt{\pi_0(1 - \pi_0)}} \right) \Delta \right)$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\pi_1(1 - \pi_1)}{\delta^2} (Z_\beta + \Delta Z_{\alpha/2})^2$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(\left(Z_{\alpha/2} - \frac{|\delta| \sqrt{N}}{\sqrt{\pi_0(1 - \pi_0)}} \right) \Delta \right) + \Phi \left(\left(-Z_{\alpha/2} - \frac{|\delta| \sqrt{N}}{\sqrt{\pi_0(1 - \pi_0)}} \right) \Delta \right)$$

- Two sided asymmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\pi_1(1 - \pi_1)}{\delta^2} (Z_\beta + \Delta Z_{\alpha/2})^2$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(\left(Z_{\alpha_u} - \frac{|\delta| \sqrt{N}}{\sqrt{\pi_0(1 - \pi_0)}} \right) \Delta \right)$$

Q Technical Reference and Formulas: Single Look Designs

$$+ \Phi \left(\left(-Z_{\alpha_i} - \frac{|\delta| \sqrt{N}}{\sqrt{\pi_0(1 - \pi_0)}} \right) \Delta \right)$$

Q.7.2 Single Arm Design : Single Proportion : Superiority: Test Statistic Distribution: Normal: Variance: Empirical

$$\delta = \pi_1 - \pi_0, \Delta = \sqrt{\frac{\pi_0(1 - \pi_0)}{\pi_1(1 - \pi_1)}}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$N = \frac{\pi_1(1 - \pi_1)}{\delta^2} (Z_\beta + Z_\alpha)^2$$

$$Power = 1 - \Phi \left(Z_\alpha - \frac{|\delta| \sqrt{N}}{\sqrt{\pi_1(1 - \pi_1)}} \right)$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\pi_1(1 - \pi_1)}{\delta^2} (Z_\beta + Z_{\alpha/2})^2$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{|\delta| \sqrt{N}}{\sqrt{\pi_1(1 - \pi_1)}} \right) + \Phi \left(-Z_{\alpha/2} - \frac{|\delta| \sqrt{N}}{\sqrt{\pi_1(1 - \pi_1)}} \right)$$

- Two sided asymmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{\pi_1(1 - \pi_1)}{\delta^2} (Z_\beta + Z_{\alpha/2})^2$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha_u} - \frac{|\delta| \sqrt{N}}{\sqrt{\pi_1(1 - \pi_1)}} \right) + \Phi \left(-Z_{\alpha_l} - \frac{|\delta| \sqrt{N}}{\sqrt{\pi_1(1 - \pi_1)}} \right)$$

Q.7.3 Paired Design: McNemar’s Test: Superiority: Test Statistic Distribution: Normal

$$\delta = \mu_t - \mu_c$$

| Control | Experimental | Total Prob |
|-------------|-----------------------|-------------|
| No Response | π_{00} π_{01} | $1 - \pi_c$ |
| Response | π_{10} π_{11} | π_c |
| Total Prob | $1 - \pi_t$ π_t | 1 |

$$\hat{\xi} = \hat{\pi}_{01} + \hat{\pi}_{10}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$N = \frac{[\hat{\xi} - (\hat{\pi}_{01} - \hat{\pi}_{10})]^2}{(\hat{\pi}_{01} - \hat{\pi}_{10})^2} (Z_\beta + Z_\alpha)^2$$

$$Power = 1 - \Phi \left(Z_\alpha - \frac{|\hat{\pi}_{01} - \hat{\pi}_{10}| \sqrt{N}}{\sqrt{\hat{\xi} - (\hat{\pi}_{01} - \hat{\pi}_{10})^2}} \right)$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{[\hat{\xi} - (\hat{\pi}_{01} - \hat{\pi}_{10})]^2}{(\hat{\pi}_{01} - \hat{\pi}_{10})^2} (Z_\beta + Z_{\alpha/2})^2$$

Q Technical Reference and Formulas: Single Look Designs

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$\begin{aligned} Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{|\hat{\pi}_{01} - \hat{\pi}_{10}| \sqrt{N}}{\sqrt{\hat{\xi} - (\hat{\pi}_{01} - \hat{\pi}_{10})^2}} \right) \\ + \Phi \left(-Z_{\alpha/2} - \frac{|\hat{\pi}_{01} - \hat{\pi}_{10}| \sqrt{N}}{\sqrt{\hat{\xi} - (\hat{\pi}_{01} - \hat{\pi}_{10})^2}} \right) \end{aligned}$$

- Two sided asymmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{[\hat{\xi} - (\hat{\pi}_{01} - \hat{\pi}_{10})^2]^2}{(\hat{\pi}_{01} - \hat{\pi}_{10})^2} (Z_{\beta} + Z_{\alpha/2})^2$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$\begin{aligned} Power = 1 - \Phi \left(Z_{a_u} - \frac{|\hat{\pi}_{01} - \hat{\pi}_{10}| \sqrt{N}}{\sqrt{\hat{\xi} - (\hat{\pi}_{01} - \hat{\pi}_{10})^2}} \right) \\ + \Phi \left(-Z_{a_l} - \frac{|\hat{\pi}_{01} - \hat{\pi}_{10}| \sqrt{N}}{\sqrt{\hat{\xi} - (\hat{\pi}_{01} - \hat{\pi}_{10})^2}} \right) \end{aligned}$$

Q.8 Sample Size :Discrete : Two Samples

- Q.8.1 Diff:Sup:Unpooled
- Q.8.2 Diff:Sup:Pooled
- Q.8.3 Diff:Noninf
- Q.8.4 Diff:Equiv
- Q.8.5 Ratios:Sup:Unpooled
- Q.8.6 Ratios:Sup:Pooled
- Q.8.7 Ratios:Noninf:FM
- Q.8.8 Ratios:Noninf:Wald
- Q.8.9 Odds Ratio:Sup
- Q.8.10 Odds Ratio:noninf
- Q.8.11 Common Odds Ratio:Sup

Q.8.1 Two Independent Samples : Difference of Proportions: Superiority: Test Statistic Distribution: Normal:Variance:Unpooled estimate

$$\delta = \mu_t - \mu_c$$

$$Z = \frac{\hat{\pi}_t - \hat{\pi}_c}{\sqrt{\frac{\hat{\pi}_t(1-\hat{\pi}_t)}{n_t} + \frac{\hat{\pi}_c(1-\hat{\pi}_c)}{n_c}}}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$Power = 1 - \Phi \left(Z_{\alpha} - \frac{|\delta|}{\sqrt{\frac{\hat{\pi}_t(1-\hat{\pi}_t)}{n_t} + \frac{\hat{\pi}_c(1-\hat{\pi}_c)}{n_c}}} \right)$$

$$N = \frac{(Z_{\alpha} + Z_{\beta})^2 [(1 - TF) * \hat{\pi}_t(1 - \hat{\pi}_t) + TF * \hat{\pi}_c(1 - \hat{\pi}_c)]}{\delta^2(1 - TF)}$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2 [(1 - TF) * \hat{\pi}_t(1 - \hat{\pi}_t) + TF * \hat{\pi}_c(1 - \hat{\pi}_c)]}{\delta^2(1 - TF)}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{|\delta|}{\sqrt{\frac{\hat{\pi}_t(1-\hat{\pi}_t)}{n_t} + \frac{\hat{\pi}_c(1-\hat{\pi}_c)}{n_c}}} \right)$$

$$+ \Phi \left(-Z_{\alpha/2} - \frac{|\delta|}{\sqrt{\frac{\hat{\pi}_t(1-\hat{\pi}_t)}{n_t} + \frac{\hat{\pi}_c(1-\hat{\pi}_c)}{n_c}}} \right)$$

- Two sided asymmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2 [(1 - TF) * \hat{\pi}_t(1 - \hat{\pi}_t) + TF * \hat{\pi}_c(1 - \hat{\pi}_c)]}{\delta^2(1 - TF)}$$

Q Technical Reference and Formulas: Single Look Designs

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha_u} - \frac{|\delta|}{\sqrt{\frac{\hat{\pi}_t(1-\hat{\pi}_t)}{n_t} + \frac{\hat{\pi}_c(1-\hat{\pi}_c)}{n_c}}} \right) + \Phi \left(-Z_{\alpha_l} - \frac{|\delta|}{\sqrt{\frac{\hat{\pi}_t(1-\hat{\pi}_t)}{n_t} + \frac{\hat{\pi}_c(1-\hat{\pi}_c)}{n_c}}} \right)$$

Q.8.2 Two Independent Samples : Difference of Proportions: Superiority: Test Statistic Distribution: Normal: Variance : Pooled estimate

$$\delta = \mu_t - \mu_c$$

$$\hat{\pi} = \frac{n_t \hat{\pi}_t + n_c \hat{\pi}_c}{N}$$

$$Z = \frac{\hat{\pi}_t - \hat{\pi}_c}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_t} + \frac{1}{n_c}\right)}}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$Power = 1 - \Phi \left(Z_{\alpha} - \frac{|\delta|}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_t} + \frac{1}{n_c}\right)}} \right)$$

$$N = \frac{(Z_{\alpha} + Z_{\beta})^2 \hat{\pi}(1-\hat{\pi})}{\delta^2 TF(1-TF)}$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2 \hat{\pi}(1-\hat{\pi})}{\delta^2 TF(1-TF)}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{|\delta|}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_t} + \frac{1}{n_c}\right)}} \right)$$

$$+ \Phi \left(-Z_{\alpha/2} - \frac{|\delta|}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_t} + \frac{1}{n_c}\right)}} \right)$$

- Two sided asymmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2 \hat{\pi}(1-\hat{\pi})}{\delta^2 TF(1-TF)}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$\begin{aligned} Power = 1 - \Phi \left(Z_{\alpha_u} - \frac{|\delta|}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_t} + \frac{1}{n_c}\right)}} \right) \\ + \Phi \left(-Z_{\alpha_l} - \frac{|\delta|}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_t} + \frac{1}{n_c}\right)}} \right) \end{aligned}$$

Casagrande-Pike-Smith Correction

The Casagrande-Pike-Smith correction is applicable to Difference of Proportions - Superiority and Noninferiority. The correction is applicable in the case of equal allocation ratio only.

For the Alternative hypothesis $H_1 : \pi_t > \pi_c$ the corrected formula for sample size is

$$n_t = n_c = \frac{A \left[1 + \sqrt{1 + \frac{4(\pi_t - \pi_c)}{A}} \right]^2}{4(\pi_t - \pi_c)^2}$$

where

$$A = \left[Z_{1-\alpha} \sqrt{2\bar{\pi}(1-\bar{\pi})} + Z_{\beta} \sqrt{\pi_t(1-\pi_t) + \pi_c(1-\pi_c)} \right]^2$$

where

$$\bar{\pi} = \frac{\pi_t + \pi_c}{2}$$

Q.8.3 Two Independent Samples : Difference of Proportions: Noninferiority: Test Statistic Distribution: Normal

$$\delta = \mu_t - \mu_c$$

Q Technical Reference and Formulas: Single Look Designs

$$Z = \frac{\hat{\pi}_t - \hat{\pi}_c - \delta_0}{\sqrt{\frac{\hat{\pi}_t(1-\hat{\pi}_t)}{n_t} + \frac{\hat{\pi}_c(1-\hat{\pi}_c)}{n_c}}}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$Power = 1 - \Phi \left(Z_\alpha - \frac{|\delta - \delta_0|}{\sqrt{\frac{\hat{\pi}_t(1-\hat{\pi}_t)}{n_t} + \frac{\hat{\pi}_c(1-\hat{\pi}_c)}{n_c}}} \right)$$

$$N = \frac{(Z_\alpha + Z_\beta)^2 [(1 - TF) * \hat{\pi}_t(1 - \hat{\pi}_t) + TF * \hat{\pi}_c(1 - \hat{\pi}_c)]}{(\delta - \delta_0)^2 (1 - TF)}$$

Q.8.4 Two Independent Samples : Difference of Proportions: Equivalence: Test Statistic Distribution: Z

Effect Size: $\delta = \pi_t - \pi_c$,

δ_1 = Expected effect size,

δ_0 = Equivalence Margin,

$r = \frac{n_t}{N}$

$H_0 : |\pi_t - \pi_c| = \delta_0$ against $H_1 : |\pi_t - \pi_c| < \delta_0 > 0$

- Compute Sample Size

$$N = \frac{(Z_\alpha + Z_\beta)^2}{(\delta_0 - \delta_1)^2} \left(\frac{\pi_c(1 + \pi_c)}{1 - r} + \frac{(\pi_c - \delta_1)(1 - (\pi_c + \delta_1))}{r} \right)$$

Q.8.5 Two Independent Samples : Ratio of Proportions: Superiority: Test Statistic Distribution: Normal :Variance: Unpooled

$$\delta = \mu_t - \mu_c, TF = \frac{N_t}{N}$$

$$Z = \frac{\ln(\hat{\pi}_t) - \ln(\hat{\pi}_c)}{\sqrt{\frac{(1-\hat{\pi}_t)}{n_t \hat{\pi}_t} + \frac{(1-\hat{\pi}_c)}{n_c \hat{\pi}_c}}}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$Power = 1 - \Phi \left(Z_{\alpha} - \frac{|\delta|}{\sqrt{\frac{(1-\hat{\pi}_t)}{n_t \hat{\pi}_t} + \frac{(1-\hat{\pi}_c)}{n_c \hat{\pi}_c}}} \right)$$

$$N = \frac{(Z_{\alpha} + Z_{\beta})^2}{\delta^2} \left[\frac{1 - \hat{\pi}_t}{TF \hat{\pi}_t} + \frac{1 - \hat{\pi}_c}{(1 - TF) \hat{\pi}_c} \right]$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{\delta^2} \left[\frac{1 - \hat{\pi}_t}{TF \hat{\pi}_t} + \frac{1 - \hat{\pi}_c}{(1 - TF) \hat{\pi}_c} \right]$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\frac{\alpha}{2}} - \frac{|\delta|}{\sqrt{\frac{(1-\hat{\pi}_t)}{n_t \hat{\pi}_t} + \frac{(1-\hat{\pi}_c)}{n_c \hat{\pi}_c}}} \right) +$$

$$\Phi \left(-Z_{\frac{\alpha}{2}} - \frac{|\delta|}{\sqrt{\frac{(1-\hat{\pi}_t)}{n_t \hat{\pi}_t} + \frac{(1-\hat{\pi}_c)}{n_c \hat{\pi}_c}}} \right)$$

- Two sided asymmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{\delta^2} \left[\frac{1 - \hat{\pi}_t}{TF \hat{\pi}_t} + \frac{1 - \hat{\pi}_c}{(1 - TF) \hat{\pi}_c} \right]$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha_u} - \frac{|\delta|}{\sqrt{\frac{(1-\hat{\pi}_t)}{n_t \hat{\pi}_t} + \frac{(1-\hat{\pi}_c)}{n_c \hat{\pi}_c}}} \right) + \Phi \left(-Z_{\alpha_l} - \frac{|\delta|}{\sqrt{\frac{(1-\hat{\pi}_t)}{n_t \hat{\pi}_t} + \frac{(1-\hat{\pi}_c)}{n_c \hat{\pi}_c}}} \right)$$

**Q.8.6 Two Independent Samples : Ratio of Proportions: Superiority: Test
Statistic Distribution: Normal: Variance: Pooled**

$$\delta = \mu_t - \mu_c, TF = \frac{N_t}{N}$$

Q Technical Reference and Formulas: Single Look Designs

$$\hat{\pi} = \frac{n_t \hat{\pi}_t + n_c \hat{\pi}_c}{N}$$

$$Z = \frac{\ln(\hat{\pi}_t) - \ln(\hat{\pi}_c)}{\sqrt{\frac{(1-\hat{\pi})}{\hat{\pi}} \left(\frac{1}{n_t} + \frac{1}{n_c} \right)}}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$Power = 1 - \Phi \left(Z_{\alpha} - \frac{|\delta|}{\sqrt{\frac{(1-\hat{\pi})}{\hat{\pi}} \left(\frac{1}{n_t} + \frac{1}{n_c} \right)}} \right)$$

$$N = \frac{(Z_{\alpha} + Z_{\beta})^2 (1 - \hat{\pi})}{\delta^2 TF(1 - TF)\hat{\pi}}$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2 (1 - \hat{\pi})}{\delta^2 TF(1 - TF)\hat{\pi}}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{|\delta|}{\sqrt{\frac{(1-\hat{\pi})}{\hat{\pi}} \left(\frac{1}{n_t} + \frac{1}{n_c} \right)}} \right)$$

$$+ \Phi \left(-Z_{\alpha/2} - \frac{|\delta|}{\sqrt{\frac{(1-\hat{\pi})}{\hat{\pi}} \left(\frac{1}{n_t} + \frac{1}{n_c} \right)}} \right)$$

- Two sided asymmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2 \hat{\pi} (1 - \hat{\pi})}{\delta^2 TF(1 - TF)}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha_u} - \frac{|\delta|}{\sqrt{\frac{(1-\hat{\pi})}{\hat{\pi}} \left(\frac{1}{n_t} + \frac{1}{n_c} \right)}} \right) + \Phi \left(-Z_{\alpha_l} - \frac{|\delta|}{\sqrt{\frac{(1-\hat{\pi})}{\hat{\pi}} \left(\frac{1}{n_t} + \frac{1}{n_c} \right)}} \right)$$

Q.8.7 Two Independent Samples : Ratio of Proportions: Noninferiority: Farrington and Manning: Test Statistic Distribution: Normal

$\delta = \pi_t - \rho_0 \pi_c, TF = \frac{N_t}{N}, \rho_0 =$ Noninferiority margin

$$Z = \frac{\hat{\pi}_t - \rho_0 \hat{\pi}_t}{\sqrt{\frac{\hat{\pi}_t(1-\hat{\pi}_t)}{n_t} + \frac{\rho_0^2 \hat{\pi}_c(1-\hat{\pi}_c)}{n_c}}$$

$$\lambda = \frac{n_t}{n_c}$$

$$\theta = \frac{1}{\lambda}$$

$$a = 1 + \theta$$

$$b = -[\rho_0(1 - \theta \pi_c) + \theta + \pi_t]$$

$$c = \rho_0(\theta \pi_c + \pi_t).$$

$$\bar{\pi}_t = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \text{ and } \bar{\pi}_c = \frac{\bar{\pi}_t}{\rho_0}$$

$$N_t \geq \frac{\left[Z_{\alpha} \sqrt{[(\rho_0^2/\theta)\bar{\pi}_c(1 - \bar{\pi}_c) + \bar{\pi}_t(1 - \bar{\pi}_t)]} + Z_{\beta} \sqrt{[(\rho_0^2/\lambda)\pi_c(1 - \pi_c) + \pi_t(1 - \pi_t)]} \right]^2}{\delta^2}$$

$$Power = \Phi \left[\frac{|\delta| * \sqrt{N_t} - Z_{\alpha} \sqrt{[(\rho_0^2/\theta)\bar{\pi}_c(1 - \bar{\pi}_c) + \bar{\pi}_t(1 - \bar{\pi}_t)]}}{\sqrt{[(\rho_0^2/\theta)\pi_c(1 - \pi_c) + \pi_t(1 - \pi_t)]}} \right]$$

Q.8.8 Two Independent Samples : Ratio of Proportions: Noninferiority: Wald's Test: Test Statistic Distribution: Normal

$\delta = \ln(\pi_t/\pi_c), TF = \frac{N_t}{N}, \rho_0 =$ Noninferiority margin

$$Z = \frac{\ln(\hat{\pi}_t) - \ln(\hat{\pi}_c) - \ln(\rho_0)}{\sqrt{\frac{(1-\hat{\pi}_t)}{n_t \hat{\pi}_t} + \frac{(1-\hat{\pi}_c)}{n_c \hat{\pi}_c}}}$$

Q Technical Reference and Formulas: Single Look Designs

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$Power = 1 - \Phi \left(Z_{\alpha} - \frac{|\delta - \ln(\rho_0)|}{\sqrt{\frac{(1-\hat{\pi}_t)}{n_t \hat{\pi}_t} + \frac{(1-\hat{\pi}_c)}{n_c \hat{\pi}_c}}} \right)$$

Q.8.9 Two Independent Samples : Odds Ratio of Proportions: Superiority: Test Statistic Distribution: Normal

$$\delta = \ln \left(\frac{\hat{\pi}_t(1 - \hat{\pi}_c)}{\hat{\pi}_c(1 - \hat{\pi}_t)} \right)$$

$$Z = \frac{\ln \left(\frac{\hat{\pi}_t(1 - \hat{\pi}_c)}{\hat{\pi}_c(1 - \hat{\pi}_t)} \right)}{\sqrt{\frac{1}{n_t \hat{\pi}_t(1 - \hat{\pi}_t)} + \frac{1}{n_c \hat{\pi}_c(1 - \hat{\pi}_c)}}}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$Power = 1 - \Phi \left(Z_{\alpha} - \frac{|\delta|}{\sqrt{\frac{1}{n_t \hat{\pi}_t(1 - \hat{\pi}_t)} + \frac{1}{n_c \hat{\pi}_c(1 - \hat{\pi}_c)}}} \right)$$

$$N = \frac{(Z_{\alpha} + Z_{\beta})^2}{\delta^2 \left(\frac{1}{TF \hat{\pi}_t(1 - \hat{\pi}_t)} + \frac{1}{(1-TF) \hat{\pi}_c(1 - \hat{\pi}_c)} \right)}$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{\delta^2 \left(\frac{1}{TF \hat{\pi}_t(1 - \hat{\pi}_t)} + \frac{1}{(1-TF) \hat{\pi}_c(1 - \hat{\pi}_c)} \right)}$$

$$Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{|\delta|}{\sqrt{\frac{1}{n_t \hat{\pi}_t(1 - \hat{\pi}_t)} + \frac{1}{n_c \hat{\pi}_c(1 - \hat{\pi}_c)}}} \right) +$$

$$\Phi \left(-Z_{\alpha/2} - \frac{|\delta|}{\sqrt{\frac{1}{n_t \hat{\pi}_t(1 - \hat{\pi}_t)} + \frac{1}{n_c \hat{\pi}_c(1 - \hat{\pi}_c)}}} \right)$$

- Two sided asymmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{\delta^2 \left(\frac{1}{TF \hat{\pi}_t (1 - \hat{\pi}_t)} + \frac{1}{(1 - TF) \hat{\pi}_c (1 - \hat{\pi}_c)} \right)}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$\begin{aligned} Power = 1 - \Phi \left(Z_{\alpha_u} - \frac{|\delta|}{\sqrt{\frac{1}{n_t \hat{\pi}_t (1 - \hat{\pi}_t)} + \frac{1}{n_c \hat{\pi}_c (1 - \hat{\pi}_c)}}} \right) + \\ \Phi \left(-Z_{\alpha_l} - \frac{|\delta|}{\sqrt{\frac{1}{n_t \hat{\pi}_t (1 - \hat{\pi}_t)} + \frac{1}{n_c \hat{\pi}_c (1 - \hat{\pi}_c)}}} \right) \end{aligned}$$

Q.8.10 Two Independent Samples : Odds Ratio of Proportions: Noninferiority: Test Statistic Distribution: Normal

$$\delta = \ln \left(\frac{\pi_t (1 - \pi_c)}{\pi_c (1 - \pi_t)} \right) = \ln(\psi); \psi_0 = \text{noninferiority margin for odds ratio}$$

$$Z = \frac{\ln \left(\frac{\hat{\pi}_t (1 - \hat{\pi}_c)}{\hat{\pi}_c (1 - \hat{\pi}_t)} \right) - \ln(\psi_0)}{\sqrt{\frac{1}{n_t \hat{\pi}_t (1 - \hat{\pi}_t)} + \frac{1}{n_c \hat{\pi}_c (1 - \hat{\pi}_c)}}}$$

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$\begin{aligned} Power = 1 - \Phi \left(Z_{\alpha} - \frac{|\delta - \ln(\psi_0)|}{\sqrt{\frac{1}{n_t \hat{\pi}_t (1 - \hat{\pi}_t)} + \frac{1}{n_c \hat{\pi}_c (1 - \hat{\pi}_c)}}} \right) \\ N = \frac{(Z_{\alpha} + Z_{\beta})^2}{(\delta - \ln(\psi_0))^2 \left(\frac{1}{TF \hat{\pi}_t (1 - \hat{\pi}_t)} + \frac{1}{(1 - TF) \hat{\pi}_c (1 - \hat{\pi}_c)} \right)} \end{aligned}$$

Q.8.11 Two Independent Samples : Common Odds Ratio for Stratified 2 × 2 tables: Superiority: Test Statistic Distribution: Normal

G = Total number of strata

$$\delta = G^{-1} \sum_{g=1}^G \left\{ \ln \left(\frac{\hat{\pi}_{tg}}{1 - \hat{\pi}_{tg}} \right) \right\} - \ln \left(\frac{\hat{\pi}_{tg}}{1 - \hat{\pi}_{tg}} \right)$$

Q Technical Reference and Formulas: Single Look Designs

$$Z = \frac{G^{-1} \sum_{g=1}^G \left\{ \ln\left(\frac{\hat{\pi}_{tg}}{1-\hat{\pi}_{tg}}\right) - \ln\left(\frac{\hat{\pi}_{cg}}{1-\hat{\pi}_{cg}}\right) \right\}}{\sqrt{G^{-1} \sum_{g=1}^G \left\{ \frac{1}{n_{tg} \hat{\pi}_{tg} (1-\hat{\pi}_{tg})} + \frac{1}{n_{cg} \hat{\pi}_{cg} (1-\hat{\pi}_{cg})} \right\}}}$$

where $\hat{\pi}_{tg}$ and $\hat{\pi}_{cg}$ are the sample proportions based on n_{tg} and n_{cg} observations seen in the treatment and control arms respectively of the g^{th} stratum.

- One sided (for both $\delta > 0$ and $\delta < 0$)

$$Power = 1 - \Phi \left(Z_{\alpha} - \frac{|\delta|}{\sqrt{G^{-1} \sum_{g=1}^G \left\{ \frac{1}{n_{tg} \hat{\pi}_{tg} (1-\hat{\pi}_{tg})} + \frac{1}{n_{cg} \hat{\pi}_{cg} (1-\hat{\pi}_{cg})} \right\}}} \right)$$

$$N = \frac{(Z_{\alpha} + Z_{\beta})^2}{\delta^2 \left(G^{-1} \sum_{g=1}^G \left\{ \frac{1}{n_{tg} \hat{\pi}_{tg} (1-\hat{\pi}_{tg})} + \frac{1}{n_{cg} \hat{\pi}_{cg} (1-\hat{\pi}_{cg})} \right\} \right)}$$

- Two sided symmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{\delta^2 \left(G^{-1} \sum_{g=1}^G \left\{ \frac{1}{n_{tg} \hat{\pi}_{tg} (1-\hat{\pi}_{tg})} + \frac{1}{n_{cg} \hat{\pi}_{cg} (1-\hat{\pi}_{cg})} \right\} \right)}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$Power = 1 - \Phi \left(Z_{\alpha/2} - \frac{|\delta|}{\sqrt{G^{-1} \sum_{g=1}^G \left\{ \frac{1}{n_{tg} \hat{\pi}_{tg} (1-\hat{\pi}_{tg})} + \frac{1}{n_{cg} \hat{\pi}_{cg} (1-\hat{\pi}_{cg})} \right\}}} \right)$$

$$+ \Phi \left(-Z_{\alpha/2} - \frac{|\delta|}{\sqrt{G^{-1} \sum_{g=1}^G \left\{ \frac{1}{n_{tg} \hat{\pi}_{tg} (1-\hat{\pi}_{tg})} + \frac{1}{n_{cg} \hat{\pi}_{cg} (1-\hat{\pi}_{cg})} \right\}}} \right)$$

- Two sided asymmetric (for both $\delta > 0$ and $\delta < 0$)
Start with the initial solution as

$$N = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{\delta^2 \left(G^{-1} \sum_{g=1}^G \left\{ \frac{1}{n_{tg} \hat{\pi}_{tg} (1 - \hat{\pi}_{tg})} + \frac{1}{n_{cg} \hat{\pi}_{cg} (1 - \hat{\pi}_{cg})} \right\} \right)}$$

and solve using an iterative procedure the following equation so that the computed power matches with the desired power with 1.e-6 precision.

$$\text{Power} = 1 - \Phi \left(Z_{\alpha_u} - \frac{|\delta|}{\sqrt{G^{-1} \sum_{g=1}^G \left\{ \frac{1}{n_{tg} \hat{\pi}_{tg} (1 - \hat{\pi}_{tg})} + \frac{1}{n_{cg} \hat{\pi}_{cg} (1 - \hat{\pi}_{cg})} \right\}}} \right) + \Phi \left(-Z_{\alpha_l} - \frac{|\delta|}{\sqrt{G^{-1} \sum_{g=1}^G \left\{ \frac{1}{n_{tg} \hat{\pi}_{tg} (1 - \hat{\pi}_{tg})} + \frac{1}{n_{cg} \hat{\pi}_{cg} (1 - \hat{\pi}_{cg})} \right\}}} \right)$$

Q.9 Sample Size
:Discrete : Many Samples

- Q.9.1 Single Arm: Chi-square
- Q.9.2 Two Group Chi-square
- Q.9.3 Wilcoxon Rank Sum
- Q.9.4 Multi-arm: Trend Test
- Q.9.5 Multi-arm: Chi-square for Rx2
- Q.9.6 Multi-arm: Chi-square: Rx C

Q.9.1 Many Samples: Single Arm: Chi-square for specified proportions in C categories

C= number of categories
Proportions under Ho : $\{p_{0i}; i = 1, 2, 3, \dots, c\}$
Proportions under H1 : $\{p_{1i}; i = 1, 2, 3, \dots, c\}$
Effect size

$$\Delta^2 = \sum_{i=1}^c \frac{(p_{0i} - p_{1i})^2}{p_{0i}}$$

Test statistic

$$\chi^2_{c-1} = N \Delta^2$$

- Compute power
Find $\chi^2_{c-1, \alpha}$ such that $P(\chi^2_{c-1} > \chi^2_{c-1, \alpha}) = \alpha$ from central Chisquare with c-1

Q Technical Reference and Formulas: Single Look Designs

degrees of freedom.

$Power = P_{\lambda}(\chi^2 > \chi_{c-1,\alpha}^2)$ where $Power = P_{\lambda}(\chi^2 > \chi_{c-1,\alpha}^2)$ is a non-central chi square variable with c-1 degree of freedom and non-centrality parameter λ .

$$\lambda = N \Delta^2$$

- Compute sample size
 N is determined using iterative method so that the power is maintained. If user has given allocation $\{r_i; i = 1, 2, 3, \dots, c\}$ N is divided into $\{N_i; i = 1, 2, 3, \dots, c\}$. These N_i s are rounded up to nearest integers and added up to get actual N.

Q.9.2 Many Samples: Parallel Design: Two group Chi-square for proportions in C categories

n_t = sample size on treatment arm

n_c = sample size on control arm

Proportions for treatment : $\{\pi_{tj}; j = 1, 2, 3, \dots, c\}$

Proportions for control : $\{\pi_{cj}; j = 1, 2, 3, \dots, c\}$

Effect size:

$$\Delta^2 = Q_1(1 - Q_1) \sum_{i=1}^c \frac{(\pi_{tj} - \pi_{cj})^2}{(\pi_{cj}(1 - Q_1) + \pi_{tj} Q_1)}$$

Where $Q_1 = \frac{n_t}{N} = \frac{n_t}{n_t + n_c}$

Noncentrality parameter λ

$$\lambda = N \Delta^2$$

- Compute Power
 Find $\chi_{c-1,\alpha}^2$ such that $P(\chi_{c-1}^2 > \chi_{c-1,\alpha}^2) = \alpha$ from central Chisquare with c-1 degrees of freedom.

$$Power = P_{\lambda}(\chi^2 > \chi_{c-1,\alpha}^2)$$

where χ^2 is a non-central chi square variable with c-1 degree of freedom and non-centrality parameter .

- Compute Sample Size
 For given power, N is determined using iterative method.

Q.9.3 Many Samples: Parallel Design: Wilcoxon Rank Sum for ordered categorical data

$\{\pi_{tj}; j = 1, 2, , \dots, c\}$ proportions for category j for treatment, $j=1,2,\dots,J$

$\{\pi_{cj}; j = 1, 2, , \dots, c\}$ proportions for category j for control, $j=1,2,\dots,J$

$$\gamma_{ci} = \sum_{j=1}^i \pi_{cj}$$

$$\gamma_{ti} = \sum_{j=1}^i \pi_{tj}$$

Effect Size

$$\psi = \ln(\gamma_{ci}) - \ln(1 - \gamma_{ci}) - (\ln(\gamma_{ti}) - \ln(1 - \gamma_{ti}))$$

$H_0 : \psi = 0$ Vs $H_1 : \psi \neq 0$ or $H_1 : \psi > 0$

m_i = multinomial samples $i=c, t$

x_{ij} = number of these m_i = observations that fall into the j th ordered category.

$$x_{cj} + x_{tj} = n_j; m_c + m_t = N$$

$$X_t = (x_{t1}, x_{t2}, \dots, x_{tC})$$

$$X_c = (x_{c1}, x_{c2}, \dots, x_{cC})n = (n_1, n_2, \dots, n_C);$$

Test Statistic:

- Wilcoxon Rank Sum

$$T = \sum_1^C w_j x_j$$

Asymptotic approximation for the exact conditional power is given by:

$$\beta(n) = 1 - \Phi \left(\frac{t_\alpha(n) - E(T|n, H_1)}{\sqrt{\text{var}(T|n, H_1)}} \right)$$

Where

$$t_\alpha(n) = E(T|n, H_0) - Z_\alpha \sqrt{\text{var}(T|n, H_0)}$$

For more details, the user is referred to Rabee et al (2003)

Q.9.4 Many Samples: Multi-arm: Trend in R ordered proportions

Case 1: User based Probabilities $r_i = i^{th}$ Population Fraction

$w_i = i^{th}$ Population Score

$\pi_i = i^{th}$ Proportion response

$$\bar{w} = \sum r_i w_i$$

Q Technical Reference and Formulas: Single Look Designs

$$\delta = \sum \pi_i r_i (w_i - \bar{w})$$

$$N_i = N * r_i$$

N_i = Population size for the i^{th} group

$$\pi = \sum r_i \pi_i$$

$$N = \sum N_i$$

$$\text{Var(Pooled)} = N\pi(1 - \pi) \sum r_i (w_i - \bar{w})^2$$

$$\text{Var(Unpooled)} = N \sum \pi_i(1 - \pi_i)r_i (w_i - \bar{w})^2$$

- One sided

$$\text{Power} = 1 - \Phi \left[\left(Z_\alpha - \frac{N * \delta}{\sqrt{\text{var(Pooled)}}} \right) \sqrt{\frac{\text{var(Pooled)}}{\text{var(UnPooled)}}} \right]$$

- Two sided

$$\text{Power} = 1 - \Phi \left[\left(Z_{\alpha/2} - \frac{N * \delta}{\sqrt{\text{var(Pooled)}}} \right) \sqrt{\frac{\text{var(Pooled)}}{\text{var(UnPooled)}}} \right] + \Phi \left[\left(-Z_{\alpha/2} - \frac{N * \delta}{\sqrt{\text{var(Pooled)}}} \right) \sqrt{\frac{\text{var(Pooled)}}{\text{var(UnPooled)}}} \right]$$

Case 2: Model based probabilities

In this case, our first aim is to compute the vector of proportion responses i.e., π_i and then apply the methods described above.

We have

$$\text{log of common odds ratio (K)} = \frac{\pi_i(1 - \pi_{i-1}) / (\pi_{i-1}(1 - \pi_i))}{W_i - W_{i-1}}$$

$$\frac{\pi_i}{(1 - \pi_i)} = \frac{\pi_{i-1}}{(1 - \pi_{i-1})} e^{K(W_i - W_{i-1})}$$

$$\pi_i = \frac{\frac{\pi_{i-1}}{(1 - \pi_{i-1})} e^{K(W_i - W_{i-1})}}{1 + \frac{\pi_{i-1}}{(1 - \pi_{i-1})} e^{K(W_i - W_{i-1})}}$$

Determine all π_i 's and then apply the steps mentioned in Case 1 to compute Power. Sample size computation is by iterating on the power function.

Q.9.5 Many Samples: Multi-arm: Chi-square for Rx2 proportions

R= number of groups
 n_t = sample size for the i^{th} arm

$$r_i = \frac{n_i}{n_1}$$

$$\pi_0 = \frac{\sum r_i \pi_i}{\sum r_i}$$

$$V = \frac{\sum r_i (\pi_i - \pi_0)^2}{\sum r_i}$$

Effect size:

$$\Delta^2 = \frac{V}{\pi_0(1 - \pi_0)}$$

Where $Q_1 = \frac{n_t}{N} = \frac{n_t}{n_t + n_c}$ Noncentrality parameter λ

$$\lambda = N \Delta^2$$

- Compute Power Find $x_{R-1,\alpha}^2$ such that $P(\chi_{R-1}^2 > \chi_{R-1,\alpha}^2) = \alpha$ from central Chisquare with c-1 degrees of freedom.

$$Power = P_{\lambda}(\chi^2 > \chi_{R-1,\alpha}^2)$$

where χ^2 is a non-central chi square variable with R-1 d.f. and non-centrality parameter λ .

- Compute Sample Size For given power, N is determined using iterative method.

Q.9.6 Many Samples: Multi-arm: Chi-square for proportions in RxC tables

R= number of groups (arms)
 C= number of categories
 n_i = sample size for the i^{th} arm
 $r_i = \frac{n_i}{n_1}$

π_{ij} = proportion of subjects belonging to the i^{th} group and j^{th} category. $i=1,2,\dots, R$, $j=1,2,\dots,C$

π_j = proportion in the j^{th} category.

Effect size:

$$\Delta^2 = \frac{\sum_{i=1}^R r_i \sum_{j=1}^C \frac{(\pi_{ij} - \pi_j)^2}{\pi_j}}{\sum r_i}$$

Q Technical Reference and Formulas: Single Look Designs

Noncentrality parameter: $\lambda = N\Delta^2$

- Compute Power
Find $\chi^2_{R-1,\alpha}$ such that $P(\chi^2_{(R-1)(C-1)} > \chi^2_{(R-1)(C-1),\alpha}) = \alpha$ from central Chisquare with $(R - 1)(C - 1)$ degrees of freedom.

$$Power = P_\lambda(\chi^2 > \chi^2_{(R-1)(C-1),\alpha})$$

where χ^2 is a non-central chi square variable with $(R - 1)(C - 1)$ degrees of freedom and non-centrality parameter λ .

- Compute Sample Size
For given power, N is determined using iterative method.

Q.10 Sample Size :Discrete : Regression

Q.10.1 Logistic Regression: Odds Ratio

Q.10.1 Logistic Regression: Odds Ratio

One Covariate

P_0 = Proportion successes of events at the mean value of the covariate, μ

P_1 = Proportion successes of events at the mean value of the covariate, $\mu + \sigma$

$$\theta = \text{Odds ratio} = \frac{P_1(1-P_0)}{P_0(1-P_1)}$$

- One sided
 - Compute Power

$$Power = \Phi \left[e^{\left(\frac{\eta^2}{4}\right)} \left(\sqrt{\frac{NP_0\eta^2}{[1 + 2P_0\delta]}} - Z_\alpha \right) \right]$$

where

$$\delta = \frac{1 + (1 + \eta^2)e^{\left(\frac{5\eta^2}{4}\right)}}{1 + e^{\left(-\frac{\eta^2}{4}\right)}}$$

$$\eta = \ln(\theta) \text{ and } Z_\alpha = \Phi^{-1}(1 - \alpha)$$

- Compute Sample Size

$$N = \frac{[Z_\alpha + Z_\beta e^{\left(-\frac{\eta^2}{4}\right)}]^2}{P_0\eta^2} [1 + 2P_0\delta]$$

where

$$\delta = \frac{1 + (1 + \eta^2) e^{\left(\frac{5\eta^2}{4}\right)}}{1 + e^{\left(\frac{-\eta^2}{4}\right)}}$$

$$\eta = \ln(\theta)$$

- Two Sided
 - Compute Power

$$Power = \Phi \left[e^{\left(\frac{\eta^2}{4}\right)} \left(\sqrt{\frac{NP_0\eta^2}{[1 + 2P_0\delta]}} - Z_{\alpha/2} \right) \right]$$

where

$$\delta = \frac{1 + (1 + \eta^2) e^{\left(\frac{5\eta^2}{4}\right)}}{1 + e^{\left(\frac{-\eta^2}{4}\right)}}$$

$$\eta = \ln(\theta) \text{ and } Z_{\alpha} = \Phi^{-1}(1 - \alpha).$$

- Compute Sample Size

$$N = \frac{\left[Z_{\alpha/2} + Z_{\beta} e^{\left(\frac{-\eta^2}{4}\right)} \right]^2}{P_0\eta^2} [1 + 2P_0\delta]$$

where

$$\delta = \frac{1 + (1 + \eta^2) e^{\left(\frac{5\eta^2}{4}\right)}}{1 + e^{\left(\frac{-\eta^2}{4}\right)}}$$

$$\eta = \ln(OR) = \ln(\theta)$$

More Than One Covariate

P_0 = Proportion successes of events at the mean value of the covariate, μ

P_1 = Proportion successes of events at the mean value of the covariate, $\mu + \sigma$

θ = odds ratio = $\frac{P_1(1-P_0)}{P_0(1-P_1)}$, ρ^2 = the square of multiple correlation coefficient (ρ) (between X_1 and other remaining covariate.)

- One sided
 - Compute Power

$$Power = \Phi \left[e^{\left(\frac{\eta^2}{4}\right)} \left(\sqrt{\frac{NP_0\eta^2(1 - \rho^2)}{[1 + 2P_0\delta]}} - Z_{\alpha} \right) \right]$$

Q Technical Reference and Formulas: Single Look Designs

where

$$\delta = \frac{1 + (1 + \eta^2)e^{\left(\frac{5\eta^2}{4}\right)}}{1 + e^{\left(-\frac{\eta^2}{4}\right)}}$$

$$\eta = \ln(\theta) \text{ and } Z_\alpha = \Phi^{-1}(1 - \alpha)$$

– Compute Sample Size

$$N = \frac{N_1}{(1 - \rho^2)}$$

where

$$N_1 = \frac{\left[Z_\alpha + Z_\beta e^{\left(-\frac{\eta^2}{4}\right)} \right]^2}{P_0 \eta^2} [1 + 2P_0 \delta]$$

and

$$\delta = \frac{1 + (1 + \eta^2)e^{\left(\frac{5\eta^2}{4}\right)}}{1 + e^{\left(-\frac{\eta^2}{4}\right)}}$$

$$\eta = \ln(\theta)$$

■ Two Sided

– Compute Power

$$Power = \Phi \left[e^{\left(\frac{\eta^2}{4}\right)} \left(\sqrt{\frac{NP_0 \eta^2 (1 - \rho^2)}{[1 + 2P_0 \delta]}} - Z_{\alpha/2} \right) \right]$$

where

$$\delta = \frac{1 + (1 + \eta^2)e^{\left(\frac{5\eta^2}{4}\right)}}{1 + e^{\left(-\frac{\eta^2}{4}\right)}}$$

$$\eta = \ln(\theta)$$

– Compute Sample Size

$$N = \frac{N_1}{(1 - \rho^2)}$$

where

$$N_1 = \frac{\left[Z_{\alpha/2} + Z_\beta e^{\left(-\frac{\eta^2}{4}\right)} \right]^2}{P_0 \eta^2} [1 + 2P_0 \delta]$$

and

$$\delta = \frac{1 + (1 + \eta^2)e^{\left(\frac{5\eta^2}{4}\right)}}{1 + e^{\left(-\frac{\eta^2}{4}\right)}}$$

Q.11 Sample Size : Agreement

Q.11.1 Cohen's Kappa: Two Binary Ratings

Q.11.2 Cohen's Kappa: Two Categorical Ratings

Q.11.1 Cohen's Kappa: Two Binary Ratings

π_{ij} = Proportion of population given rating i by Rater 1 and j by Rater 2.

K_0 = Kappa under Null,

K_1 = Kappa under Alternative

- One Sided

$$Power = \Phi \left(\frac{\sqrt{N_1}(|K_1 - K_0|) - Z_\alpha \sqrt{Q_0}}{\sqrt{Q_1}} \right)$$

where

$$Q_1 = \sum_i \pi_{ii} [(1 - \pi_c) - (\pi_{i.} + \pi_{.i})(1 + \pi_0)]^2$$

- Compute Sample Size

$$N = \left[\frac{Z_\alpha \sqrt{Q_0} + Z_\beta \sqrt{Q_1}}{K_1 - K_0} \right]^2$$

- Two Sided

$$Power = \Phi \left(\frac{\sqrt{N_1}(|K_1 - K_0|) - Z_{\alpha/2} \sqrt{Q_0}}{\sqrt{Q_1}} \right)$$

Where

$$Q_1 = \sum_i \pi_{ii} [(1 - \pi_c) - (\pi_{i.} + \pi_{.i})(1 + \pi_0)]^2$$

- Compute Sample Size

$$N = \left[\frac{Z_\alpha \sqrt{Q_0} + Z_\beta \sqrt{Q_1}}{K_1 - K_0} \right]^2$$

Q.11.2 Agreement: Cohen's Kappa: Two Categorical Ratings

C = Number of ratings

π_0 = Proportion of agreement

Q Technical Reference and Formulas: Single Look Designs

- π_e = Expected proportion of agreement
- π_{ij} = Proportion of population given rating i by Rater 1 and j by Rater 2.
- K_0 = Kappa under Null
- K_1 = Kappa under Alternative

- Compute Power

$$Power = \Phi \left(\frac{\sqrt{N}(K_1 - K_0) - Z_{1-\alpha} \max \tau(\hat{k}|k = 0.4)}{\max \tau(\hat{k}|k = 0.6)} \right)$$

where

$$\tau(\hat{k}) = \frac{\sqrt{Q_1 + Q_2 - 2Q_3 - Q_4}}{(1 - \pi_e)^2}$$

$$Q_1 = \pi_0 (1 - \pi_e)^2$$

$$Q_2 = (1 - \pi_0)^2 \sum_i \sum_j \pi_{ij} (\pi_{i.} + \pi_{.j})^2$$

$$Q_3 = 2(1 - \pi_0)(1 - \pi_e) \sum_i \pi_{ij} (\pi_{i.} + \pi_{.j})$$

$$Q_4 = (\pi_0 \pi_e - 2\pi_e + \pi_0)^2$$

- Compute Sample Size

$$N \geq \left(\frac{Z_{1-\alpha} \max \tau(\hat{k}|k = k_0) + Z_{1-\beta} \max \tau(\hat{k}|k = k_1)}{k_1 - k_0} \right)^2$$

Ref : Flack, V.F., et. Al. (1988).

**Q.12 Sample Size :
Count Data**

Q.12.1 One Sample: Single Poisson rate

Q.12.2 Two Samples: Ratio of Poisson Rates

Q.12.3 Ratio of Negative Binomial Rates

Q.12.1 One Sample: Single Poisson rate

X : No. of events (outcomes) observed during an interval of specified length.

D = Exposure Duration (This could be time, length, volume, area etc)

$X \sim \text{Poisson}(\lambda D)$

λ = Poisson rate (mean number of occurrences of X during a unit length interval)

λ_0 = Hypothesized value of λ

λ_1 = Value of λ at which Power is to be computed.

n = sample size = Number of times observations on X taken over the Exposure duration D

$G(.,k)$ denote the CDF of chi square distribution with k d.f.

- One sided test (right tailed)

$$H_0 : \lambda = \lambda_0 \text{ Vs } H_1 : \lambda = \lambda_0$$

- Compute Power

1. Find 'k' such that

$$G(2nD \lambda_0; 2k) \leq \alpha \tag{Q.1}$$

2. Compute

$$\text{Power} = 1 - F(k - 1, nD \lambda_1) = G(2nD \lambda_1; 2k) \tag{Q.2}$$

where k is obtained from equation Q.1.

- Compute sample size

Solve equation (Q.1) and equation (Q.2) simultaneously for n and k .

- One sided test (left tailed)

$$H_0 : \lambda = \lambda_0 \text{ VS } H_1 : \lambda < \lambda_0$$

- Compute Power

1. Find 'k' such that

$$G(2nD \lambda_0; 2(k + 1)) \geq 1 - \alpha \tag{Q.3}$$

2. Compute

$$\text{Power} = 1 - G(2nD \lambda_1; 2(k + 1)) \tag{Q.4}$$

where k is obtained from equation

Q Technical Reference and Formulas: Single Look Designs

- Compute sample size Solve equation (Q.3) and equation (Q.4) simultaneously for n and k.

- Two sided test

$$H_0 : \lambda = \lambda_0 \text{ Vs } H_1 : \lambda \neq \lambda_0$$

For carrying out a two sided design,(compute power and sample size and duration) compute

$$\alpha' = \frac{\alpha}{2}$$

Execute the algorithm for one sided (right or left depending upon the sign of the difference $\lambda_1 - \lambda_0$) with α' as the value of level of significance, α .

Q.12.2 Two Samples: Ratio of Poisson Rates

- λ_c : Poisson rate for control arm
 λ_t : Poisson rate for treatment arm
 D_t : Duration of study for the treatment arm
 D_c : Duration of study for the control arm
 X_t : No of events (outcomes) observed on Treatment arm in time D_t
 $X_t \sim \text{Poisson}(\lambda_t D_t)$
 X_c : No of events (outcomes) observed on Control arm in time D_c
 $X_c \sim \text{Poisson}(\lambda_c D_c)$
 n_t : Number of observations on Treatment arm
 n_c : Number of observations on Control arm
 $r = \frac{n_t}{n_c}$ allocation ratio
 $d = \frac{D_c n_c}{D_t n_t}$
 $\rho_0 =$ Hypothecated value of the ratio, $\frac{\lambda_t}{\lambda_c}$
 $\rho_1 =$ value of the ratio at which the power is to be computed

- One sided test (right tailed)
 $H_0 : \frac{\lambda_t}{\lambda_c} > \rho_0 \geq 1 \text{ Vs } H_1 : \frac{\lambda_t}{\lambda_c} = \rho_0$
 $H_1 : \frac{\lambda_t}{\lambda_c} = \rho_1$
 where $\rho_1 > \rho_0$
 Test Statistic

$$W_3 = \frac{\ln\left(\frac{X_t}{X_c}\right) - \ln\left(\frac{\rho_0}{d}\right)}{\sqrt{\frac{1}{X_t} + \frac{1}{X_c}}}$$

In case X_t or $X_c = 0$, the value is set to 0.5 for that variable.

- Compute Power

$$Power = 1 - \Phi\left(Z_{1-\alpha} - \frac{\mu}{\sigma}\right) \tag{Q.5}$$

where $\mu = \ln\left(\frac{\rho_1}{\rho_0}\right)$ and $\sigma^2 = \frac{d + \rho_1}{D_c n_c \lambda_c \rho_1}$

- Compute Sample Size
Solve equation (Q.5) for n_c by using the following algorithm.

1. Compute

$$\sigma^2 = \left[\frac{\ln\left(\frac{\rho_1}{\rho_0}\right)}{Z_{1-\alpha} - \Phi^{-1}(1 - power)} \right]^2$$

2. Compute

$$n_c = \frac{d + \rho_1}{D_c \sigma^2 \lambda_c \rho_1}$$

3. Compute

$$n_t = r * n_c$$

$$n = n_t + n_c$$

- One sided test (left tailed)

$$H_0 : \frac{\lambda_t}{\lambda_c} = \rho_0 \geq 1 \text{ Vs } H_1 : \frac{\lambda_t}{\lambda_c} < \rho_0$$

$$H_1 : \frac{\lambda_t}{\lambda_c} = \rho_1$$

Where $\rho_1 < \rho_0$

$$Power = \Phi\left(Z_\alpha - \frac{\mu}{\sigma}\right) \tag{Q.6}$$

Where $\mu = \ln\left(\frac{\rho_1}{\rho_0}\right)$ and $\sigma^2 = \frac{d + \rho_1}{D_c n_c \lambda_c \rho_1}$ In case X_t or $X_c = 0$, the value is set to 0.5 for that variable.

- Compute Sample Size
Solve equation (Q.6) for n_c by using the following algorithm.

1. Compute

$$\sigma^2 = \left[\frac{\ln\left(\frac{\rho_1}{\rho_0}\right)}{Z_\alpha - \Phi^{-1}(1 - power)} \right]^2$$

2. Compute

$$n_c = \frac{d + \rho_1}{D_c \sigma^2 \lambda_c \rho_1}$$

Q Technical Reference and Formulas: Single Look Designs

3. Compute

$$n_t = r * n_c$$

$$n = n_t + n_c$$

■ Two Sided Test

$$H_0 : \frac{\lambda_t}{\lambda_c} = \rho_0 \geq 1 \text{ Vs } H_1 : \frac{\lambda_t}{\lambda_c} \neq \rho_0$$

Depending upon the ratio of rates > 1 or < 1 , use the power computation formula for $\rho_1 > \rho_0$ or $\rho_1 < \rho_0$ as the case may be with α replaced by $\frac{\alpha}{2}$.

Q.12.3 Two Samples: Ratio of Negative Binomial Rates

$$X_c \sim NB(\lambda_c, \Upsilon_c)$$

$$X_t \sim NB(\lambda_t = \theta\lambda_c, \Upsilon_t) \quad \theta = \frac{\lambda_t}{\lambda_c} \quad u = \text{Fixed follow up } k = \text{Allocation Ratio} \\ = n_t/n_c$$

■ One sided test (Left tailed)

$$H_0 : \theta = 1 \text{ Vs } H_1 : \theta < 1$$

– Compute power

$$\text{Power} = \Phi(E_\theta - z_\alpha)$$

$$\text{Where } E_\theta = \text{Test statistic} = -\sqrt{n_c} \frac{\ln(\hat{\theta})}{\sqrt{\frac{1+\gamma_c \lambda_c u}{\lambda_c \mu} + \frac{1+\gamma_t \lambda_c \theta u}{k \lambda_c \theta u}}}$$

$$A = \frac{1}{[\ln(\hat{\theta})]^2} \left[\sqrt{\frac{1 + \gamma_c \lambda_c u}{\lambda_c \mu} + \frac{1 + \gamma_t \lambda_c \theta u}{k \lambda_c \theta u}} \right]$$

– Compute Sample Size

$$n = A(z_\alpha + z_\beta)^2(1 + k)$$

■ One sided test (Right tailed)

$$H_0 : \theta = 1 \text{ Vs } H_1 : \theta > 1$$

– Compute Power

$$\text{Power} = \Phi(E_\theta - (-1 * z_\alpha))$$

- Compute Sample Size

$$n = A((-1 * z_{\alpha}) + z_{\beta})^2(1 + k)$$

- Two sided test
 $H_0 : \theta = 1$ Vs. $H_1 : \theta \neq 1$

- Compute Power

$$Power = 1 - \Phi(E_{\theta} - (-1 * z_{\alpha/2})) + \Phi(E_{\theta} - z_{\alpha/2})$$

- Compute Sample Size

$$n = A(z_{\alpha/2} + z_{\beta})^2(1 + k)$$

Q.13 Sample Size :Time to Event Data

Q.13.1 Two Samples: Superiority: Logrank

Effect Size: $\delta = \ln\left(\frac{\lambda_t}{\lambda_c}\right)$ where λ_t and λ_c are hazard rates for treatment and control arms respectively. In Time to event studies, maximum number of events are determined for given power.

$H_0: \delta = 0$ Vs $H_1: \delta = \delta_1$

Test Statistic (Log Rank) Suppose at the end of study, in all q failures are observed with failure times $\tau_1, \tau_2, \dots, \tau_i, \dots, \tau_q$. Accordingly, there will be q 2x2 tables of the following type.

The i^{th} table is shown below:

Where the subscripts t and c indicate values observed under treatment and control.

$$S = \sum_{i=1}^q \left\{ d_t(\tau_i) - \frac{n_t(\tau_i) d_t(\tau_i)}{n(\tau_i)} \right\}$$

Q Technical Reference and Formulas: Single Look Designs

| Status | Treatment T | Treatment C | Total |
|------------|-----------------------------|-----------------------------|-------------------------|
| Failed | $d_t(\tau_i)$ | $d_c(\tau_i)$ | $d(\tau_i)$ |
| Not Failed | $n_t(\tau_i) - d_t(\tau_i)$ | $n_c(\tau_i) - d_c(\tau_i)$ | $n(\tau_i) - d(\tau_i)$ |
| Total | $n_t(\tau_i)$ | $n_c(\tau_i)$ | $n(\tau_i)$ |

$$S \sim AN(\text{Mean} = \delta D_{\max} r(1 - r), \text{Variance} = r(1 - r) D_{\max})$$

Where r= proportion randomized to treatment T .

- One sided test (Variance under Null)

$$D_{\max} = \frac{(Z_{\alpha} + Z_{\beta})^2}{\delta_1^2 r(1 - r)}$$

- One sided test (Variance under Alternative)

$$D_{\max} = \frac{(Z_{\alpha} + Z_{\beta})^2}{\delta_1^2 p(1 - p)}$$

where p=proportion of D_{\max} estimated to be on the experimental arm under the alternative hypothesis. East uses an iterative procedure to estimate p.

- Two sided test (Variance under Null)

$$D_{\max} = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{\delta_1^2 r(1 - r)}$$

- Two sided test (Variance under Alternative)

$$D_{\max} = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{\delta_1^2 p(1 - p)}$$

where p=proportion of D_{\max} estimated to be on the experimental arm under the alternative hypothesis. East uses an iterative procedure to estimate p.

Q.13.2 Two Samples: Noninferiority: Logrank

Effect Size: $\delta = \ln\left(\frac{\lambda_t}{\lambda_c}\right)$ where λ_t and λ_c are hazard rates for treatment and control arms respectively. In Time to event studies, maximum number of events are determined for given power.

H0: $\delta > \delta_0$ Vs H1: $\delta < \delta_0$

Test Statistic (Log Rank)

Suppose at the end of study, in all q failures are observed with failure times $\tau_1, \tau_2, \dots, \tau_i, \dots, \tau_q$. Accordingly, there will be q 2x2 tables of the following type. The i^{th} table is shown below:

| Status | Treatment T | Treatment C | Total |
|------------|-----------------------------|-----------------------------|-------------------------|
| Failed | $d_t(\tau_i)$ | $d_c(\tau_i)$ | $d(\tau_i)$ |
| Not Failed | $n_t(\tau_i) - d_t(\tau_i)$ | $n_c(\tau_i) - d_c(\tau_i)$ | $n(\tau_i) - d(\tau_i)$ |
| Total | $n_t(\tau_i)$ | $n_c(\tau_i)$ | $n(\tau_i)$ |

Where the subscripts t and c indicate values observed under treatment and control.

$$S = \sum_{i=1}^q \left\{ d_t(\tau_i) - \frac{n_t(\tau_i) d(\tau_i)}{n(\tau_i)} \right\} - \delta_0$$

$$S \sim AN(\text{Mean} = \delta D_{\max} r(1-r) - \delta_0, \text{Variance} = r(1-r) D_{\max})$$

Where r= proportion randomized to treatment T

δ_0 = Noninferiority margin.

- One sided test (Variance under Null and Alternative both)

$$D_{\max} = \frac{(Z_\alpha + Z_\beta)^2}{(\delta_1 - \delta_0)^2 r(1-r)}$$

R Technical Reference and Formulas: Analysis

In this Appendix, we provide the theory used in **East 6.4** for analyzing data under the **Analysis** menu.

Note: The test statistics formulas provided in this Appendix can be used in interim analysis of data while monitoring a group sequential study or for analyzing data arising out of a single sample study. For common notations and references the user is referred to the Technical Reference and Formulas:Single Look Designs or the respective chapters of the tests.

R.1 Basic Statistics- Descriptive Statistics

R.1.1 Central Tendency

R.1.2 Dispersion

R.1.3 Distribution

R.1.4 Summary

R.1.1 Central Tendency

Mean If $X_i, i = 1, 2, \dots, n$ are n observations, then the mean \bar{X} is defined as

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \tag{R.1}$$

Further if f_i is the frequency or weight for $X_i, i = 1, 2, \dots, n$, then the mean \bar{X} is defined as

$$\bar{X} = \frac{1}{\sum_{i=1}^n f_i} \sum_{i=1}^n X_i f_i \tag{R.2}$$

Median: Median is the value of the middle most observation, when the observations are arranged in ascending or descending order. If the number of observations is even, then the median is defined as the mean of the middle most two observations.

Mode: Mode is the value of X_i with the maximum frequency f_i . If there are more than one X_i with maximum frequency, then the smallest of all such X_i 's will be used as the value of mode.

Geometric Mean: If $X_i, i = 1, 2, \dots, n$ are n observations, then the geometric mean GM is defined as

$$GM = \left[\prod_{i=1}^n X_i \right]^{\frac{1}{n}} \tag{R.3}$$

Further if f_i is the frequency or weight for $X_i, i = 1, 2, \dots, n$, then the geometric mean GM is defined as

$$GM = \left[\prod_{i=1}^n X_i^{f_i} \right]^{\frac{1}{\sum f_i}} \quad (R.4)$$

Harmonic Mean If $X_i, i = 1, 2, \dots, n$ are n observations, then the harmonic mean HM is defined as

$$HM = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}} \quad (R.5)$$

Further if f_i is the frequency or weight for $X_i, i = 1, 2, \dots, n$, then the harmonic mean HM is defined as

$$HM = \frac{\sum_{i=1}^n f_i}{\sum_{i=1}^n \frac{f_i}{X_i}} \quad (R.6)$$

R.1.2 Dispersion

Standard Deviation If $X_i, i = 1, 2, \dots, n$ are n observations, then the standard deviation is defined as

$$s = \left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^{0.5} \quad (R.7)$$

Further if f_i is the frequency or weight for $X_i, i = 1, 2, \dots, n$, then the standard deviation is defined as

$$s = \left[\frac{1}{\sum_{i=1}^n f_i - 1} \sum_{i=1}^n (X_i - \bar{X})^2 f_i \right]^{0.5} \quad (R.8)$$

Standard Error of Mean If $X_i, i = 1, 2, \dots, n$ are n observations and s is the standard deviation, then the standard error of mean is defined as

$$SE = \frac{s}{\sqrt{n}} \quad (R.9)$$

R Technical Reference and Formulas: Analysis

Further if f_i is the frequency or weight for $X_i, i = 1, 2, \dots, n$, then the standard error of mean is defined as

$$SE = \frac{s}{\sqrt{\sum_{i=1}^n f_i}} \quad (R.10)$$

Variance Variance is defined as the square of the standard deviation and is denoted as s^2 .

Coefficient of variation If \bar{x} and s are the mean and standard deviation respectively, then Coefficient of Variation is defined as follows:

$$CV = \frac{\bar{X}}{s} \quad (R.11)$$

Minimum is the minimum value of $X_i, i = 1, 2, \dots, n$.

Maximum is the maximum value of $X_i, i = 1, 2, \dots, n$.

Range is calculated as the difference: Maximum-Minimum.

R.1.3 Distribution

Skewness If $X_i, i = 1, 2, \dots, n$ are n observations, then a measure of skewness is defined as

$$skewness = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3}{\left(\frac{(n-1)}{n} s^2\right)^{(3/2)}} \quad (R.12)$$

Further if f_i is the frequency or weight for $X_i, i = 1, 2, \dots, n$, then a measure of skewness is defined as

$$skewness = \frac{\frac{1}{\sum_{i=1}^n f_i} \sum_{i=1}^n (X_i - \bar{X})^3 f_i}{\left(\frac{(n-1)}{n} s^2\right)^{(3/2)}} \quad (R.13)$$

For normal distribution, skewness is zero and for any symmetric data, the value of skewness should be zero or close to zero. A negative value of skewness indicates that the data are skewed to the left or the left tail is heavier than the right tail. A positive value of skewness can be interpreted in a similar way.

Kurtosis If $X_i, i = 1, 2, \dots, n$ are n observations, then a measure of kurtosis is defined as

$$Kurtosis = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{\left(\frac{(n-1)}{n} s^2\right)^2} - 3 \quad (R.14)$$

Further if f_i is the frequency or weight for $X_i, i = 1, 2, \dots, n$, then a measure of kurtosis is defined as

$$Kurtosis = \frac{\frac{1}{\sum_{i=1}^n f_i} \sum_{i=1}^n (X_i - \bar{X})^4 f_i}{\left(\frac{(n-1)}{n} s^2\right)^2} - 3 \quad (\text{R.15})$$

The standard normal distribution has a kurtosis of 3. A kurtosis value > 3 indicates a relatively **peaked** distribution and a value < 3 indicates relatively **flat** distribution of the data.

R.1.4 Summary

Sum If $X_i, i = 1, 2, \dots, n$ are n observations, then sum is defined as

$$Sum = \sum_{i=1}^n X_i \quad (\text{R.16})$$

Further if f_i is the frequency or weight for $X_i, i = 1, 2, \dots, n$, then sum is defined as

$$Sum = \sum_{i=1}^n X_i f_i \quad (\text{R.17})$$

Count If $X_i, i = 1, 2, \dots, n$ are n observations, then Count is defined as

$$Count = n \quad (\text{R.18})$$

Further if f_i is the frequency or weight for $X_i, i = 1, 2, \dots, n$, then Count is defined as

$$Count = \sum_{i=1}^n f_i \quad (\text{R.19})$$

R Technical Reference and Formulas: Analysis

R.2 Basic Statistics-Analytics

R.2.1 Independent t-test

R.2.2 Paired t-test

R.2.3 Analysis of Variance

R.2.4 Spearman's Rank-Order Correlation

R.2.5 Multiple Linear Regression

R.2.6 Collinearity Diagnostics

R.2.7 Multivariate Analysis of Variance

R.2.1 Independent t-test

Equal variance

If x_1, x_2, \dots, x_{n_x} is a random sample from a normal population with mean μ_x and standard deviation σ_x and y_1, y_2, \dots, y_{n_y} is a random sample from a normal population with mean μ_y and standard deviation σ_y , we want to test null hypothesis:

$$H_0: \mu_x = \mu_y \text{ under the assumption } \sigma_x = \sigma_y$$

The test statistic is:

$$t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \tag{R.20}$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \text{and} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \tag{R.21}$$

and s is the pooled standard deviation.

$$s = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}} \tag{R.22}$$

The above statistic is distributed as t with $(n_x + n_y - 2)$ degrees of freedom.

Unequal variance

If x_1, x_2, \dots, x_{n_x} is a random sample from a normal population with mean μ_x and standard deviation σ_x and y_1, y_2, \dots, y_{n_y} is a random sample from a normal population with mean μ_y and standard deviation σ_y , we want to test null hypothesis:

$$H_0: \mu_x = \mu_y \text{ under the assumption } \sigma_x \neq \sigma_y$$

The testing procedure uses the approximation described by Scheffe (1970) as follows:.

$$t = \frac{\hat{\delta}}{\sqrt{\frac{S_y^2}{n_y} + \frac{S_x^2}{n_x}}} \sim t_\nu \tag{R.23}$$

where ν is the degrees of freedom given by

$$\nu = \frac{\left(\frac{S_y^2}{n_y} + \frac{S_x^2}{n_x}\right)^2}{\frac{\left(\frac{S_y^2}{n_y}\right)^2}{n_y - 1} + \frac{\left(\frac{S_x^2}{n_x}\right)^2}{n_x - 1}} \tag{R.24}$$

R.2.2 Paired t-test

If $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ are n paired observations, we would like to test the hypothesis that the differences $d_1 = x_1 - y_1, d_2 = x_2 - y_2, \dots, d_n = x_n - y_n$ come from a normal distribution with mean 0. If μ is the population mean of the differences, then we want to test null hypothesis:

$$H_0: \mu = 0.$$

The test statistic is

$$t = \frac{\bar{d}}{s/\sqrt{n}} \tag{R.25}$$

where

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i \tag{R.26}$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2} \tag{R.27}$$

This statistic is distributed as t with degrees of freedom (n-1).

R.2.3 Analysis of Variance

One-way Analysis of Variance: Suppose n subjects have been allocated randomly to r treatments and measurements have been made on a variate x for all the subjects, with the resulting data being denoted as follows:

Treatment 1: $x_{11}, x_{12}, \dots, x_{1n_1}$

Treatment 2: $x_{21}, x_{22}, \dots, x_{2n_2}$

⋮

Treatment r: $x_{r1}, x_{r2}, \dots, x_{rn_r}$

We assume that the data of the r treatment groups come from r normally distributed populations with the same variance σ^2 and with means $\mu_1, \mu_2, \dots, \mu_r$. We want to test the hypothesis that these means are equal:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_r \tag{R.28}$$

The sum of squares is

$$S = \sum_{i=1}^r \sum_{k=1}^{n_i} (x_{ik} - \bar{x})^2, \tag{R.29}$$

R Technical Reference and Formulas: Analysis

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^r \sum_{k=1}^{n_i} x_{ik} = \frac{1}{n} \sum_{i=1}^r n_i \bar{x}_i, \tag{R.30}$$

$$\bar{x}_i = \frac{1}{n_i} (x_{i1} + x_{i2} + \dots + x_{in_i}) \tag{R.31}$$

and

$$n = \sum_{i=1}^r n_i. \tag{R.32}$$

We decompose the “sum of squares” S into two parts S_1 and S_2 ,

$$S = S_1 + S_2 \tag{R.33}$$

where

$$S_1 = \sum_{i=1}^r n_i (\bar{x}_i - \bar{x})^2 \tag{R.34}$$

$$S_2 = \sum_{i=1}^r \sum_{k=1}^{n_i} (x_{ik} - \bar{x}_i)^2. \tag{R.35}$$

S_1 refers to the variation between the treatments and S_2 the variation within treatments. The ratio,

$$F = \frac{S_1/(r-1)}{S_2/(n-r)} \tag{R.36}$$

follows F distribution with (r-1, n-r) degrees of freedom. All these computations can be displayed in the usual ANOVA table as shown below:

ANOVA Table Two-way Analysis of Variance: In a two-way experimental design,

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | F | P-Value |
|---------------------|----------------|--------------------|-------------------|-----------|---------|
| Between groups | S_1 | r-1 | $M_1 = S_1/(r-1)$ | M_1/M_2 | |
| Residuals | S_2 | n-r | $M_2 = S_2/(n-r)$ | | |
| Total | S | n-1 | | | |

there are two factors: A and B, with A having levels A_1, A_2, \dots, A_a and B having levels B_1, B_2, \dots, B_b . Suppose there are c observations for each combination of the factor levels A_i and B_j , then the data from such a study can be represented as:

$(x_{ijk}, i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, c)$ where the subscript i refers to level A_i of factor A, j refers to level B_j of factor B and k refers to k^{th} observation for the

combination of A_i and B_j . We assume that the number of replications for each combination of i and j is equal to c .

We assume that $n = abc$ observations x_{ijk} correspond to n random variables which are independent and are distributed normally with the same variance σ^2 . We want to test the hypotheses that:

- the means of A at all the a levels are same
- the means of B at all the b levels are same

For carrying out these tests, we proceed as follows: We decompose the total "sum of squares"

$$S = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (x_{ijk} - \bar{x})^2 \tag{R.37}$$

into three parts S_1 , S_2 , and S_3 ,

$$S = S_1 + S_2 + S_3 \tag{R.38}$$

where

$$S_1 = bc \sum_{i=1}^a (x_{i..} - \bar{x})^2 \tag{R.39}$$

refers to the sum of squares due to the variation between the levels of A,

$$S_2 = ac \sum_{j=1}^b (x_{.j.} - \bar{x})^2 \tag{R.40}$$

refers to the sum of squares due to the variation between the levels of B, and

$$S_3 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (x_{ijk} - \bar{x}_{i..} - \bar{x}_{.j.} - \bar{x})^2 \tag{R.41}$$

refers to the sum of squares due to the residual variation. Under the null hypothesis, the quantities $\frac{1}{(a-1)}S_1$, $\frac{1}{(b-1)}S_2$ and $\frac{1}{(n-a-b+1)}S_3$ have χ^2 distribution with $(a - 1)$, $(b - 1)$ and $(abc - a - b + 1)$ degrees of freedom respectively. From this it follows that the quantity

$$f_1 = \frac{S_1/(a-1)}{S_3/(abc-a-b+1)} \tag{R.42}$$

R Technical Reference and Formulas: Analysis

follows F-distribution with $(a - 1, n - a - b + 1)$ degrees of freedom and the quantity

$$f_2 = \frac{S_2/(b - 1)}{S_3/(abc - a - b + 1)} \tag{R.43}$$

follows F-distribution with $(b - 1, n - a - b + 1)$ degrees of freedom. All these computations are displayed in the usual ANOVA table as shown below:

ANOVA Table

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | F | P-Value |
|---------------------|----------------|--------------------|-------------------------|-----------|---------|
| Factor A | S_1 | a-1 | $M_1 = S_1/(a-1)$ | M_1/M_3 | |
| Factor B | S_2 | b-1 | $M_2 = S_1/(b-1)$ | M_2/M_3 | |
| Residuals | S_3 | abc-a-b+1 | $M_3 = S_3/(abc-a-b+1)$ | | |
| Total | S | n-1 | | | |

R.2.4 Correlations

Pearson's Product-Moment Correlation Coefficient

Let $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ be the n paired observations of two continuous random variables x and y . The Pearson product-moment correlation coefficient is a measure of association for these two variables. The formula for the Pearson product-moment correlation coefficient is

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \tag{R.44}$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \text{and} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \tag{R.45}$$

If f_i is the frequency or weight for the i^{th} paired observation (x_i, y_i) , then the formula for Pearson product-moment correlation coefficient can be written as:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) f_i}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 f_i} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 f_i}} \tag{R.46}$$

where

$$\bar{x} = \frac{1}{\sum_{i=1}^n f_i} \sum_{i=1}^n x_i f_i, \quad \text{and} \quad \bar{y} = \frac{1}{\sum_{i=1}^n f_i} \sum_{i=1}^n y_i f_i. \quad (\text{R.47})$$

Spearman’s Rank-Order Correlation Coefficient

If we are reluctant to make the assumption of bivariate normality, we may use Spearman’s rank-order correlation coefficient instead of Pearson’s product-moment correlation coefficient. The only difference between the two measures of association is that Pearson’s measure uses the raw data whereas Spearman’s uses ranks derived from the raw data. Spearman’s rank-order correlation coefficient can be computed by substituting the ranks of x_i and ranks of y_i in the formulas for Pearson product-moment correlation coefficient. If ties are present in the raw data, the average ranks are used.

Kendall’s Tau

Kendall’s Tau is an alternative to Pearson’s product-moment correlation coefficient and Spearman’s rank-order correlation coefficient for ordinal data. The main distinction between this measure and Pearson’s or Spearman’s measures is that we can compute Kendall’s Tau without specifying numerical values. The actual values are needed only to order the variables, hence, different values that preserve the order will output same values of Kendall’s taus. All that is needed is an implicit ordering of the data. Kendall’s Tau is a nonparametric measure of association. It is based on the number of concordances and discordances in paired observations. When paired observations vary together, it denotes concordance and when they vary differently, it indicates discordance. The formula for Kendall’s Tau can be written as,

$$\tau = \frac{\sum_{i < j} \text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j)}{\sqrt{(t_1 - t_2)(t_1 - t_3)}} \quad (\text{R.48})$$

with

$$\begin{aligned} t_1 &= n(n - 1)/2 \\ t_2 &= \sum_k u_k(u_k - 1)/2 \\ t_3 &= \sum_l v_l(v_l - 1)/2 \end{aligned}$$

where u_k is the number of tied x values in the k^{th} group of tied x values, v_l is the number of tied y values in the l^{th} group of tied y values.

R Technical Reference and Formulas: Analysis

R.2.5 Multiple Linear Regression

The regression procedures are performed using a variance-covariance updating procedure described in Maindonald, J. H. (1984). The least squared solution is facilitated by using Cholesky decomposition.

Model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1 + \dots + \beta_k X_k + \varepsilon$$

where Y is the dependent variable (response) and X_1, \dots, X_k are the independent variables (predictors) and ε is a random error with a normal distribution having mean=0 and variance= σ^2 . The multiple linear regression algorithm computes the estimates $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$, of the regression coefficients $\beta_0, \beta_1, \dots, \beta_k$, so as to minimize the sum of squares of residuals.

R.2.6 Collinearity Diagnostics

You can obtain **Collinearity Diagnostics** along the lines of Belsey, Kuh, and Welsh (1980), as a part of regression output. Under **Collinearity Diagnostics** the columns represent the variance components (related to principal components in multivariate analysis) and the rows represent the **variance proportion decomposition** explained by each variable in the model. The eigenvalues are those associated with the **singular value decomposition** of the covariance matrix of the coefficients (in fact the eigenvalues are the squares of the singular values) and the **condition numbers** are the ratios of the square root of the largest eigenvalue to all the rest. Since two or more variables are required to establish a dependency, it follows that two or more regression coefficient variances will be adversely affected by high variance decomposition proportions associated with a particular eigenvalue. It can be shown that only one high variance proportion in a given column cannot be indicative of a multicollinearity problem since the variance decomposition matrix of an orthogonal matrix (the ideal case indicating total independence) consists of only 0's and 1's. Thus, the broad rule for assessing collinearity is that there is an eigenvalue associated with a high condition index (> 30 , say) and with very high variance decomposition proportion (> 0.5 , say) for two or more regression coefficient variances. Interpretations are less obvious when there are competing dependencies (two or more near dependencies with the same condition index values) or two or more near dependencies with one condition index greatly dominating the others.

The general principle suggested by Belsley, Kuh and Welsh is that near dependencies or collinearity, problems exist if the condition index exceeds some threshold, variously quoted as 10, 15 or 30. It is suggested that a condition index greater than 30 indicates moderate to severe collinearity.

Parameters for Collinearity Diagnostics

The parameters you need to provide for these diagnostics are described below:

Number of collinearity components: Enter the number of collinearity components. This number can be between 2 and the number of degrees of freedom for the model. When the model is fitted without an intercept, the model degrees of freedom is equal to the number of predictors in the model. When the model is fitted with an intercept, the model degrees of freedom is equal to the number of predictors in the model plus one.

Multicollinearity Criterion: The default value is 0.05. It controls how small the determinant of the matrix inverted to compute the coefficient estimates, is allowed to be. If a finer tolerance is required, decrease this value, this achieves a coarser tolerance can be achieved. This value must be between 0 and 1.

Residuals

You can obtain the results of various types of residuals which are described in this section.

Unstandardized Residuals: These are computed by the formula *Unstandardized residual = Actual response - Predicted response*.

Standardized Residuals: These consist of residuals divided by their standard deviation. They have the drawback that they do not have a common standard deviation.

Studentized Residuals: These are computed by dividing the unstandardized residuals by quantities related to the diagonal elements of the hat matrix, using a common scale estimate computed without the i^{th} case in the model. (Cook and Weisberg refer to this as **external studentization**). These residuals have t - distributions with (n-k-1) degrees of freedom, so any residual with absolute value exceeding 3, usually requires attention. (n is the number of cases).

Deleted (predicted) Residuals: The deleted residual for the i^{th} observation is obtained by fitting the model with the i^{th} observation omitted, using the model to predict the i^{th} observation and then computing the difference from the actual i^{th} observation. The sum of squares of these deleted residuals is referred to as the

Predicted Residual Error Sum of Squares (PRESS) statistic and is often used to select from competing regression models. The expression for PRESS is based on the studentized residuals (see Cook and Weisberg (1982)).

Influence Statistics

Cook's Distance: Cook's Distance is an overall measure of the impact of the i^{th} datapoint on the estimated regression coefficient. In linear regression, Cook's distance has, approximately, an F distribution with k and (n-k) degrees of freedom. A guide to the influence of the i^{th} observation is given as follows: (see Bowerman, O'Connell, and Dickey (1986)).

- If D_i is less than $F(.8, k, n-k)$ (the upper 20th percentile of the F-distribution having k and n-k degrees of freedom), then the i^{th} observation should not be considered influential.

R Technical Reference and Formulas: Analysis

- If D_i is greater than $F(.5, k, n - k)$ (the 50th percentile of the F-distribution having k and $n - k$ degrees of freedom), then the i^{th} observation should be considered influential.
- If $F(.8, k, n - k) \leq D_i \leq F(.5, k, n - k)$ then the nearer D_i is to $F(.5, k, n - k)$ the greater the extent of the influence of the i^{th} observation.

DFFIT's (change in the regression fit): These reflect coefficient changes as well as forecasting effects when an observation is deleted and are similar to Cook's distance.

Covariance Ratios: This measure reflects the change in the covariance matrix of the estimated coefficients when the i^{th} observation is omitted. The suggestion is that $|covarianceratio - 1| \geq 3p/n$ warrants further investigation.

Diagonal of the hat matrix: This measure is also known as the leverage of the i^{th} observation. The diagonal elements sum to the number of parameters being fitted. Any value greater than $2 * p/n$ suggests further investigation.

R.2.7 Multivariate Analysis of Variance

One-way MANOVA

Suppose, n individuals have been subjected randomly to r treatments and measurements have been made on p variates with resulting data represented as follows:

Treatment 1: $X_{11}, X_{12}, \dots, X_{1n_1}$

Treatment 2: $X_{21}, X_{22}, \dots, X_{2n_2}$

⋮

Treatment r : $X_{r1}, X_{r2}, \dots, X_{rn_r}$.

where $n_1 + n_2 + \dots + n_r = n$ (we assume that there are at least two observations in each group). We note that each X_{ij} is a p -dimensional column vector. Assume that each vector observation $\sim N(\mu, \Sigma)$. We want to test the hypothesis that these mean vectors are equal i.e.:

$$H_0 : \mu_1 = \mu_2 \dots = \mu_r$$

We draw an analogy with univariate one way ANOVA. There we calculated various sums of squares, namely 'between groups sum of squares', 'residuals sum of squares', and 'total sum of squares. Here too, we will compute similar entities. In the multivariate situation, instead of one value, we will have a $p \times p$ matrix of values. The values along the diagonal and the off diagonal elements will be sums of cross products (SSP). The formulas are given in the table below.

Manova Table For Comparing Mean Vectors Of Populations

| Source of Variation | Matrix Sum of Squares and Cross Product | Degrees of Freedom |
|---------------------|---|--------------------|
| Treatment | $B = \sum_{i=1}^r n_i (X_i - \bar{X})(X_i - \bar{X})'$ | $r - 1$ |
| Residual | $W = \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'$ | $\sum n_i - r$ |
| Total | $B + W = \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - \bar{X})(X_{ij} - \bar{X})'$ | $\sum n_i - 1$ |

where $\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$ and $\bar{X} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} X_{ij}$. The test is based on Wilk's Λ which is given by

$$\text{Wilk's } \Lambda = \frac{|W|}{|B + W|}$$

Small values of Wilk's Λ suggest rejection of the null hypothesis. A general result regarding approximate distribution of Λ is that $-\ln \Lambda$ follows a chi-square distribution with $p(r - 1)$ d.f. For a detailed table regarding distribution of Wilk's λ , (see Johnson and Wichern, 1998).

Test for Parallel Profiles

Here the hypothesis to be tested is weaker than the earlier hypothesis that asserted equality of mean vectors. Instead, now we ask if differences in successive co-ordinate wise means are the same in all populations. In other words our hypothesis is: H_0 : The difference $\mu_{ij} - \mu_{i,j-1}$ is the same for all groups $i = 1, 2, \dots, r$, and for all components $j = 2, 3, \dots, p$. This hypothesis in matrix form can be expressed as

$$H_0 : C\mu_1 = C\mu_2 = \dots = C\mu_r$$

Where

$$C_{(p-1) \times p} = \begin{pmatrix} -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}$$

Clearly a test for this hypothesis is the same test as above after transforming the variables from $X_{p \times 1}$ to $Y_{(p-1) \times 1}$ where $Y = CX$.

R Technical Reference and Formulas: Analysis

R.3 Continuous

R.3.1 Single Arm: Single Mean

R.3.2 Paired Design: Mean of Paired Differences

R.3.3 Parallel Design: Difference of Means

R.3.4 Wilcoxon Signed Rank Test

R.3.5 Linear Regression

R.3.1 Single Arm: Single Mean

Normal Superiority Trials: One-Sample Test - Single Mean

- Hypothesis: $H_0 : \mu = \mu_0$
- Test statistic

$$Z = \frac{\hat{\mu} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}},$$

where $\hat{\mu}$ is the sample mean and $\hat{\sigma}^2$ is the sample variance based on the n observations.

- References:
 1. Jennison, C and Turnbull, BW (2000).
 2. Sheskin, DJ (2004).

R.3.2 Paired Design: Mean of Paired Differences

Normal Superiority Trials: One-Sample Test - Mean of Paired Differences

- Hypothesis: $H_0 : \delta = \mu_t - \mu_c = 0$
- Test statistic

$$Z = \frac{\hat{\mu}_t - \hat{\mu}_c}{\sqrt{\frac{\hat{\sigma}_d^2}{n}}},$$

where $\hat{\mu}_t$ and $\hat{\mu}_c$ are the sample means based on the n pairs of observations in the treatment and control arm, respectively, and $\hat{\sigma}_d^2$ is the sample variance of the paired differences. Denote the observed differences by d_l , for $l = 1, \dots, n$ pairs of observations, then the sample variance is given by:

$$\hat{\sigma}_d^2 = \frac{\sum_{l=1}^n d_l^2 - \frac{(\sum_{l=1}^n d_l)^2}{n}}{n - 1}.$$

- References:
 1. Jennison, C and Turnbull, BW (2000).
 2. Sheskin, DJ (2004).

Normal Superiority Trials: One-Sample Test - T-test for Single Mean

- Hypothesis: $H_0 : \mu = \mu_0$.

- Test statistic:

$$T = \frac{\hat{\mu} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}},$$

where $\hat{\mu}$ is the sample mean and $\hat{\sigma}^2$ is the sample variance based on n observations.

- References:

- Sheskin, DJ (2004).

Normal Superiority Trials: One-Sample Test - T-test for Mean of Paired Differences

- Hypothesis: $H_0 : \delta = \mu_t - \mu_c = 0$

- Test statistic:

$$T = \frac{\hat{\mu}_t - \hat{\mu}_c}{\sqrt{\frac{\hat{\sigma}_d^2}{n}}},$$

where $\hat{\mu}_t$ and $\hat{\mu}_c$ are the sample mean based on n pairs of observations in the treatment and control arm, respectively, and $\hat{\sigma}_d^2$ is the sample variance of the paired differences. Denote the observed differences by d_l , for $l = 1, \dots, n$ pairs of observations, then the sample variance is given by:

$$\hat{\sigma}_d^2 = \frac{\sum_{l=1}^n d_l^2 - \frac{(\sum_{l=1}^n d_l)^2}{n}}{n - 1}.$$

- References:

- Sheskin, DJ (2004).

R.3.3 Parallel Design: Difference of Means

Normal Superiority Trials: Two-Sample Test - Difference in Means

- Hypothesis : $H_0 : \delta = \mu_t - \mu_c = 0$

Variance : Equal

- Test statistic

$$Z = \frac{\hat{\mu}_t - \hat{\mu}_c}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_c} + \frac{1}{n_t} \right)}},$$

where $\hat{\mu}_t$ and $\hat{\mu}_c$ are the sample mean based on n_t and n_c observations, and $\hat{\sigma}^2$ is the pooled estimate of variance.

R Technical Reference and Formulas: Analysis

- References:
 1. Cytel East 3 User Manual (2004).
 2. Jennison, C and Turnbull, BW (2000).
 3. Sheskin, DJ (2004).

Normal Superiority Trials: Two-Sample Test - T-test for Difference of Independent Means

- Hypothesis : $H_0 : \delta = \mu_t - \mu_c = 0$
Variance : Equal
- Test statistic:

$$T = \frac{\hat{\mu}_t - \hat{\mu}_c}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_c} + \frac{1}{n_t} \right)}}$$

where $\hat{\mu}_t$ and $\hat{\mu}_c$ are the sample mean based on n_t and n_c observations in the treatment and control arm, respectively, and $\hat{\sigma}^2$ is the pooled estimate of variance.

- References:
 1. Sheskin, DJ (2004).

Normal Non-Inferiority Trials: Two-Sample Test - Difference in Means

- Hypothesis : $H_0 : \delta = \mu_t - \mu_c >= \delta_0$
- Test statistic :

$$Z = \frac{\hat{\mu}_c - \hat{\mu}_t - \delta_0}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_c} + \frac{1}{n_t} \right)}}$$

where $\hat{\mu}_t$ and $\hat{\mu}_c$ are the sample mean based on n_t and n_c observations in the treatment and control arm, respectively, and $\hat{\sigma}^2$ is the pooled estimate of variance. δ_0 is the non-inferiority margin.

- References:
 1. Cytel East 3 User Manual (2004).
 2. Jennison, C and Turnbull, BW (2000).
 3. Sheskin, DJ (2004).

Normal Non-Inferiority Trials: Two-Sample Test - T-test for Difference of Independent Means

- Hypothesis : $H_0 : \delta = \mu_t - \mu_c >= \delta_0$

- Test statistic:

$$T = \frac{\hat{\mu}_c - \hat{\mu}_t - \delta_0}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_c} + \frac{1}{n_t} \right)}}$$

where $\hat{\mu}_t$ and $\hat{\mu}_c$ are the sample mean based on n_t and n_c observations in the treatment and control arm, respectively, and $\hat{\sigma}^2$ is the pooled estimate of variance. δ_0 is the non-inferiority margin.

- References:

- Sheskin, DJ (2004).

Normal Equivalence Trials: Two-Sample Test - Difference of Means

- Hypothesis : $H_0 : \delta = \mu_t - \mu_c \leq \delta_L$ Or $\delta = \mu_t - \mu_c \geq \delta_U$

- Test statistics:

(This test is performed as two separate α -level one-sided hypothesis t-tests)

$$T_L = \frac{\hat{\mu}_c - \hat{\mu}_t - \delta_L}{\sqrt{\frac{\hat{\sigma}^2}{nr(1-r)}}},$$

and

$$T_U = \frac{\hat{\mu}_c - \hat{\mu}_t - \delta_U}{\sqrt{\frac{\hat{\sigma}^2}{nr(1-r)}}},$$

where $\hat{\mu}_t$ and $\hat{\mu}_c$ are the sample mean in the treatment and control arm, respectively, and $\hat{\sigma}^2$ is the pooled estimate of common variance, all based on n observations. The assigned fraction r is the probability of being randomized to the treatment arm, and δ_L and δ_U are the lower and upper equivalence limits, respectively. Denote the sample variance in the treatment and control arm by $\hat{\sigma}_t^2$ and $\hat{\sigma}_c^2$, respectively. The pooled estimate of common variance is given by:

$$\hat{\sigma}^2 = \frac{(n_t - 1)\hat{\sigma}_t^2 + (n_c - 1)\hat{\sigma}_c^2}{n - 2}.$$

- References:

- Schuirman, DJ (1987).
- Diletti, E, Hauschke, D. and Steinijs, VW (1991).
- Owen, DB (1965).

Normal Equivalence Trials: Two-Sample Test - Log Ratio of Means

- Hypothesis : $H_0 : \delta = \ln(\mu_t/\mu_c) \leq \delta_L$ Or $\delta = \ln(\mu_t/\mu_c) \geq \delta_U$

R Technical Reference and Formulas: Analysis

- Test statistics:
(This test is performed as two separate α -level one-sided hypothesis t-tests)

$$T_L = \frac{\ln(\hat{\mu}_c) - \ln(\hat{\mu}_t) - \delta_L}{\sqrt{\frac{\ln(1+\hat{C}V^2)}{nr(1-r)}}},$$

and

$$T_U = \frac{\ln(\hat{\mu}_c) - \ln(\hat{\mu}_t) - \delta_U}{\sqrt{\frac{\ln(1+\hat{C}V^2)}{n_j r(1-r)}}},$$

where $\hat{\mu}_t$ and $\hat{\mu}_c$ are the sample means in the treatment and control arm, respectively, and $\hat{C}V$ is the pooled estimate of the coefficient of variation, all based on n observations. The assigned fraction r is the probability of being randomized to the treatment arm, and δ_L and δ_U are the lower and upper equivalence limits, respectively.

- References:
 1. Schuirmann, D.J. (1987).
 2. Hauschke, D, Kieser, M, Diletti, E and Burke, M (1998).
 3. Diletti, E, Hauschke, D and Steinijans, VW (1991).
 4. Owen, D.B. (1965).

Normal Equivalence Trials: Two-Sample Test - Difference of Means in Crossover Designs

- Hypothesis : $H_0 : \delta = \mu_t - \mu_c \leq \delta_L$ Or $\delta = \mu_t - \mu_c \geq \delta_U$
To determine by a difference metric whether the unknown mean μ_t under treatment is equal to the unknown mean μ_c under control for n subjects enrolled in a 2×2 crossover trial.
- Test statistics:
(This test is performed as two separate α -level one-sided hypothesis t-tests)

$$T_L = \frac{\hat{\mu}_c - \hat{\mu}_t - \delta_L}{\sqrt{\frac{MSE}{nr(1-r)}}},$$

and

$$T_U = \frac{\hat{\mu}_c - \hat{\mu}_t - \delta_U}{\sqrt{\frac{MSE}{nr(1-r)}}},$$

where $\hat{\mu}_t$ and $\hat{\mu}_c$ are the sample mean in the treatment and control arm, respectively, and MSE is the mean squared error obtained by fitting a linear

model to the crossover data, all based on n observations. The assigned fraction r is the probability of being randomized to the treatment arm, and δ_L and δ_U are the lower and upper equivalence limits, respectively.

■ References:

1. Schuirmann, DJ (1987).
2. Diletti, E, Hauschke, D. and Steinijans, VW (1991).
3. Owen, DB (1965).

Normal Equivalence Trials: Two-Sample Test - Log Ratio of Means in Crossover Designs

■ $H_0 : \delta = \ln(\mu_t/\mu_c) \leq \delta_L$ Or $\delta = \ln(\mu_t/\mu_c) \geq \delta_U$

To determine by a log ratio metric whether the unknown mean μ_t under treatment is equal to the unknown mean μ_c under control for n subjects enrolled in a 2×2 crossover trial.

■ Test statistics:

(This test is performed as two separate α -level one-sided hypothesis t-tests)

$$T_L = \frac{\hat{\mu}_c - \hat{\mu}_t - \delta_L}{\sqrt{\frac{MSE}{nr(1-r)}}},$$

and

$$T_U = \frac{\hat{\mu}_c - \hat{\mu}_t - \delta_U}{\sqrt{\frac{MSE}{nr(1-r)}}},$$

where $\hat{\mu}_t$ and $\hat{\mu}_c$ are the sample means in the treatment and control arm, respectively, and MSE is the mean squared error obtained by fitting a linear model to the crossover log data, all based on n observations. The assigned fraction r is the probability of being randomized to the treatment arm, and δ_L and δ_U are the lower and upper equivalence limits, respectively.

■ References:

1. Schuirmann, D.J. (1987).
2. Hauschke, D, Kieser, M, Diletti, E and Burke, M (1998).
3. Diletti, E, Hauschke, D and Steinijans, VW (1991).
4. Owen, D.B. (1965).

R.3.4 Wilcoxon Signed Rank Test

Notation

R_i : Rank of $|D_i|$ when absolute values are arranged in ascending order.

I : is the indicator function.

R Technical Reference and Formulas: Analysis

- **Hypothesis:** $H_0 : \lambda = 0$
- **Test Statistic:**

$$W^+ = \sum_{i=1}^n R_i \mathbf{I}(D_i > 0) \sim AN(\mu_W, \sigma_W^2) \quad (\text{R.49})$$

where,

$$\mu_W = \frac{n(n+1)}{4}. \quad (\text{R.50})$$

and

$$\sigma_W^2 = \frac{n(n+1)(2n+1)}{24}. \quad (\text{R.51})$$

R.3.5 Linear Regression

Normal Superiority Trials: Linear Regression - Comparing Slope to Predefined Value

- Hypothesis : $H_0 : \theta_t = \theta_c$
- Model:
Given a response Y_l and a covariate $X_l \sim N(\mu_x, \sigma_x^2)$ for subject $l = 1, \dots, n_j$, consider the linear model:

$$Y_l = \gamma + \theta X_l + \varepsilon_l,$$

where all ε_l 's are independent and identically distributed (i.i.d.) as $N(0, \sigma_\varepsilon^2)$.

- Test statistic

$$Z = \frac{\hat{\theta} - \hat{\theta}_0}{\sqrt{\frac{\hat{\sigma}_\varepsilon^2}{n\hat{\sigma}_x^2}}},$$

where $\hat{\theta}$ is the estimated regression slope parameter, $\hat{\sigma}_x^2$ is the sample variance of the covariate X in the sample, and $\hat{\sigma}_\varepsilon^2$ is the sample error variance, all based on the n observations.

- References:
 1. Dupont, WD and Plummer, WD, Jr. (1998).
 2. Jennison, C and Turnbull, BW (2000).

Normal Superiority Trials: Linear Regression - Comparing Two Slopes

- Hypothesis : $H_0 : \theta_t = \theta_c$

- Model:
Given a response Y_{il} and a covariate $X_{il} \sim N(\mu_{xi}, \sigma_{xi}^2)$ for subject $l = 1, \dots, n_j$ submitted to treatment $i = c, t$, consider the linear model:

$$Y_{il} = \gamma + \theta_i X_{il} + \varepsilon_{il}.$$

where all ε'_{il} s are i.i.d. $N(0, \sigma_\varepsilon^2)$

- Test statistic:

$$Z = \frac{\hat{\theta}_t - \hat{\theta}_c}{\sqrt{\hat{\sigma}_\varepsilon^2 \left(\frac{1}{n_t \hat{\sigma}_{xt}^2} + \frac{1}{n_c \hat{\sigma}_{xc}^2} \right)}}$$

where $\hat{\theta}_t$ and $\hat{\theta}_c$ are the estimated regression slope parameters and $\hat{\sigma}_{xt}^2$ and $\hat{\sigma}_{xc}^2$ are the sample variances of the covariate X in the treatment and control arm, respectively, based on the n_t and the n_c observations, while $\hat{\sigma}_\varepsilon^2$ is the sample error variance.

- References:
 1. Dupont, WD and Plummer, WD, Jr (1998).
 2. Jennison, C and Turnbull, BW (2000).

Normal Superiority Trials: Repeated Measures Regression - Comparing Two Slopes

- Hypothesis : $H_0 : \theta_t = \theta_c$
where θ_t and θ_c are regression fixed slope parameters for two distinct population regressions using independent random samples of subject-specific repeated measures.
- Model:
Given a final response Y_{iml} and a prior series of repeated measurements on the response variable at times $v_m, m = 1, \dots, M$ for subject $l = 1, \dots, n$ submitted to treatment $i = c, t$, consider the linear mixed effects model:

$$Y_{iml} = \gamma_i + \theta_i v_m + a_l + b_l v_m + \varepsilon_{ml},$$

where the random effect $(a_l, b_l)'$ is multivariate normal with mean $(0, 0)'$ and variance-covariance matrix:

$$G = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix},$$

and all ε_{ml} are i.i.d. $N(0, \sigma_w^2)$.

R Technical Reference and Formulas: Analysis

- Test statistic:

$$Z = \frac{\hat{\theta}_t - \hat{\theta}_c}{\sqrt{\left(\hat{\sigma}_b^2 + \frac{12(M-1)\hat{\sigma}_w^2}{M(M+1)S^2}\right) \left(\frac{1}{n_t} + \frac{1}{n_c}\right)}}$$

where $\hat{\theta}_t$ and $\hat{\theta}_c$ are the estimated regression fixed slope parameters based on n_t and n_c observations in the treatment and control arm, respectively, $\hat{\sigma}_b^2$ and $\hat{\sigma}_w^2$ are the between and within sample variances, respectively, M is the total number of measurements on each subject, and S is the follow-up time for each subject.

- References:

- Fitzmaurice, GM, Laird, NM and Ware, JH (2004).
- Jennison, C and Turnbull, BW (2000).

R.4 Discrete

R.4.1 Test for Proportion in One Sample Binomial

R.4.2 McNemar's Test for Paired Binomial

R.4.1 Test for Proportion in One Sample Binomial

- Hypothesis : $H_0 : \pi = \pi_0$
to be tested against a two-sided alternative hypothesis $H_1 : \pi \neq \pi_0$ or a one-sided alternative hypothesis $H_1 : \pi < \pi_0$ or $H_1' : \pi > \pi_0$. In this analysis, the hypothesis is tested asymptotically as well as using Exact Inference.

- Asymptotic Inference**

Test Statistic: Using the variance estimated under the null hypothesis:

$$Z = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

where $\hat{\pi}$ is the sample proportion based on the n observations. **East** computes 1-sided and 2-sided asymptotic p-values using standard normal distribution of the test statistic Z . Also, confidence interval for the population proportion is derived for the specified value of confidence level.

- Exact Inference**

Suppose the data consist of t successes, and $n - t$ failures, in n independent Bernoulli trials. Let π be the true underlying success rate. Then the outcome $T = t$ has the Binomial probability

$$\Pr(T = t|\pi) = \binom{n}{t} \pi^t (1 - \pi)^{n-t} . \tag{R.52}$$

East computes the maximum likelihood estimate of π as

$$\hat{\pi} = t/n .$$

Next, **East** computes a $100 \times (1 - \gamma)\%$ exact confidence interval for π using the method of Clopper and Pearson (1934). This method computes the interval in the form $(\pi_*(t), \pi^*(t))$, where $\pi_*(t)$ is such that:

$$\pi_*(t) = 0, \text{ if } t = 0 \quad (\text{R.53})$$

$$\Pr(T \geq t | \pi_*(t)) = \frac{\gamma}{2}, \text{ if } 0 < t \leq n \quad (\text{R.54})$$

and $\pi^*(t)$ is such that:

$$\Pr(T \leq t | \pi^*(t)) = \frac{\gamma}{2}, \text{ if } 0 \leq t < n \quad (\text{R.55})$$

$$\pi^*(t) = 1, \text{ if } t = n . \quad (\text{R.56})$$

A unique and very useful option available in **East** is Casella's procedure for computing confidence intervals (Casella, 1986). This procedure guarantees *uniformly shorter* exact confidence intervals than the commonly used Clopper-Pearson confidence intervals described above. In other words, for any value of n and any observed value of t , we will obtain shorter confidence intervals for π . The Casella procedure generalizes the technique of Blyth and Still (1983); in **East** we refer to these intervals as Blyth-Still-Casella intervals. To test the null hypothesis:

$$H_0: \pi = \pi_0 , \quad (\text{R.57})$$

East computes the following 1 and 2-sided p-values:

$$p_1 = \min\{\Pr(T \leq t | \pi_0), \Pr(T \geq t | \pi_0)\} , \quad (\text{R.58})$$

$$p_2 = 2 * p_1 . \quad (\text{R.59})$$

East also computes the power against the alternative hypothesis:

$$H_1: \pi = \pi_1 (\pi_1 > \pi_0) . \quad (\text{R.60})$$

Let α be the probability of a Type I error and t_0 be the smallest integer such that:

$$\Pr(T \geq t_0 | \pi_0) \leq \alpha . \quad (\text{R.61})$$

R Technical Reference and Formulas: Analysis

Then, the exact (one-sided) power is given by:

$$1 - \beta = \Pr(T \geq t_0 | \pi_1) . \tag{R.62}$$

If $\pi_1 < \pi_0$,

$$1 - \beta = \Pr(T \leq t_0 | \pi_1) , \tag{R.63}$$

where t_0 is the largest integer for which

$$\Pr(T \leq t_0 | \pi_0) \leq \alpha . \tag{R.64}$$

R.4.2 McNemar's Test for Paired Binomial

Suppose that two binomial responses are observed on each of N pairs. Let y_{11} be the count of the number of individuals whose first and second responses are both positive. Let y_{22} be the count of the number of individuals whose first and second responses are both negative. Let y_{12} be the count of the number of individuals whose first response is positive and whose second response is negative. Finally let y_{21} be the count of the number of individuals whose first response is negative and whose second response is positive. Then McNemar's test is defined on a single 2×2 table of the form

$$\mathbf{y} = \begin{matrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{matrix} .$$

Let $(\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$, denote the four cell probabilities for this table. The null hypothesis of interest is:

$$H_0: \pi_{12} = \pi_{21} .$$

verus

$$H_1: \pi_{12} \neq \pi_{21} .$$

McNemar's statistic only depends on the values of the off-diagonal elements of the 2×2 table. The Test Statistic is:

$$MC(\mathbf{y}) = y_{12} - y_{21} . \tag{R.65}$$

Let y represent any generic 2×2 contingency table and suppose that \mathbf{x} is the 2×2 table actually observed. The exact permutation distribution of the test statistic (R.65) is obtained by conditioning on the observed sum of off-diagonal terms, or "discordant pairs",

$$N_d = y_{12} + y_{21} \tag{R.66}$$

We define the reference set by

$$\Gamma = \{ \mathbf{y}: \mathbf{y} \text{ is } 2 \times 2; y_{12} + y_{21} = N_d \} . \tag{R.67}$$

Given

$$\mu = \frac{\pi_1}{\pi_1 + \pi_2}, \tag{R.68}$$

We see that evaluation of H_0 versus H_1 is equivalent to testing

$$H'_0 : \mu = 0.5 \tag{R.69}$$

versus

$$H'_1 : \mu \neq 0.5. \tag{R.70}$$

The conditional probability $P(\mathbf{y})$ of observing any $\mathbf{y} \in \Gamma$ is binomial with parameters (μ, N_d) . Thus

$$P(\mathbf{y}) = \binom{N_d}{y_{12}} \mu^{y_{12}} (1 - \mu)^{N_d - y_{12}}, \tag{R.71}$$

which reduces under (R.69) to

$$P(\mathbf{y}) = \frac{(0.5)^{N_d} N_d!}{y_{12}! y_{21}!}. \tag{R.72}$$

Hence, under the null hypothesis the probability that McNemar's statistic equals or exceeds its observed value $MC(\mathbf{x})$ is readily evaluated as

$$\Pr(MC(\mathbf{Y}) \geq MC(\mathbf{x})) = \sum_{MC(\mathbf{Y}) \geq MC(\mathbf{x})} \mathbf{P}(\mathbf{Y}), \tag{R.73}$$

the sum being taken over all $\mathbf{y} \in \Gamma$. The probability that McNemar's statistic is less than or equal to $MC(\mathbf{x})$ is similarly obtained. The exact one-sided p-value is then defined as

$$p_1 = \min\{\Pr(MC(\mathbf{Y}) \leq MC(\mathbf{x})), \Pr(MC(\mathbf{Y}) \geq MC(\mathbf{x}))\} \tag{R.74}$$

We can show that the exact distribution of the test statistic $MC(\mathbf{Y})$ is symmetric about 0. Therefore the exact two-sided p-value is defined as double the exact one-sided p-value:

$$p_2 = 2p_1. \tag{R.75}$$

In large samples, the standardized test statistic (which we report in the output for both exact and asymptotic options)

$$MC^*(\mathbf{y}) = \frac{y_{12} - y_{21}}{\sqrt{N_d}} \tag{R.76}$$

R Technical Reference and Formulas: Analysis

is asymptotically normally distributed with zero mean and unit variance. The 1-sided asymptotic p-value is defined as:

$$\tilde{p}_1 = \min\{\Phi(MC^*(\mathbf{x})), (1 - \Phi(MC^*(\mathbf{x})))\} \tag{R.77}$$

where $\Phi(z)$ is the left tail of the standard normal distribution at z , and x is the observed 2×2 contingency table. The 2-sided asymptotic p-value is double the 1-sided asymptotic p-value. The confidence interval is obtained for the difference of proportions based on the asymptotic distribution.

R.5 Two Independent Binomials

R.5.1 Exact Superiority

Test:Diff

R.5.2 Exact Noninferiority

Test :Diff

R.5.3 Exact Equivalence

Test:Diff

R.5.4 Exact CI for Diff of

Prop

R.5.5 Exact CI for Ratio of

Prop

R.5.6 Exact Noninferiority

Test: Ratio

R.5.7 CI for Binomial

Ratio

R.5.8 Restricted Nuisance

Parameter Range

R.5.9 Noninferiority:Odds

Ratio of Proportions

R.5.10 Common Odds

Ratio for Stratified

2x2 Tables

R.5.11 Fisher's Exact Test

R.5.1 Exact Unconditional Test of Superiority : Difference of Proportions

This section presents the statistical theory underlying Exact unconditional inference for data sampled from two independent binomial populations. Although the problems we will discuss are commonly encountered, the underlying theory is not easily accessible elsewhere. Consider a randomized clinical trial comparing an experimental treatment T, to a control treatment C, on the basis of a binomially distributed outcome variable, X , with probability of success π_t and π_c respectively. Consider the data presented in the 2×2 contingency table coming from control and treatment arm, \mathbf{x} , displayed in Table R.1:

Table R.1: The Observed 2x2 Contingency Table, \mathbf{x} .

| Response | Population C | Population T | Row_Total |
|-----------|--------------|--------------|-----------|
| Success | x_{1c} | x_{1t} | m_1 |
| Failure | x_{2c} | x_{2t} | m_2 |
| Col.Total | n_c | n_t | N |

The two columns of Table R.1 arise from two independent binomial populations. In the first column for control arm, there are x_{1c} successes and x_{2c} failures in n_c independent Bernoulli trials, each with probability π_c of success. Second column corresponds to data on the treatment arm. The sum of successes from the two arms is $m_1 = x_{1c} + x_{1t}$. The sample sizes n_c and n_t are number of observations on control and treatment arm. Define the difference in proportions between treatment group and control group to be $\delta = \pi_t - \pi_c$. The null hypothesis of interest is $H_0 : \delta = 0$ which is tested against a 2-sided alternative hypothesis $H_1 : \delta \neq 0$ or a 1-sided alternative hypothesis $H_1 : \delta > 0$ or $H'_1 : \delta < 0$ as the case maybe. Let $\hat{\pi}_t$ and $\hat{\pi}_c$ be the sample

proportions based on n_t and n_c observations in the treatment and control arm respectively. Then the estimate of δ is $\hat{\delta} = \hat{\pi}_t - \hat{\pi}_c$.

Asymptotic Inference

Test statistic is defined as:

$$Z = \frac{\hat{\pi}_t - \hat{\pi}_c}{\sqrt{\left(\frac{x_{1c} + x_{1t}}{N}\right) \left(\frac{x_{2c} + x_{2t}}{N}\right) \left(\frac{1}{n_t} + \frac{1}{n_c}\right)}} \tag{R.78}$$

Z is distributed as variable that follows $N(0, 1)$ distribution under the null hypothesis.

Exact Unconditional Inference

Suppose that H_0 is true and let the common probability of success for the two binomial populations be $\pi_c = \pi_t = \pi$. Then the probability of observing the data in Table R.1 is a product of two binomial probabilities, denoted by

$$f_0(\mathbf{x}) = \binom{n_c}{x_{1c}} \binom{n_t}{x_{1t}} \pi^{x_{1c} + x_{1t}} (1 - \pi)^{x_{2c} + x_{2t}} \tag{R.79}$$

The p-value is defined to be the probability, under H_0 , of obtaining a 2×2 table at least as extreme as the observed table, \mathbf{x} . Before we can compute this p-value, however, we need to answer two questions:

1. What criterion should we use to establish that a 2×2 contingency table is at least as extreme as \mathbf{x} ?
2. What is the exact null probability of each of these extreme 2×2 contingency tables?

To answer these questions we must introduce some more notation. Let \mathbf{Y} denote any generic 2×2 table that can arise if we take two independent samples, one of size n_c from binomial population C and the other of size n_t from binomial population T. Such a generic 2×2 table is displayed below in Table R.3

Table R.2: Any Generic 2x2 Contingency Table, \mathbf{Y}

| Response | Control | Treatment | Row_Total |
|-----------|----------|-----------|-------------------|
| Success | y_{1c} | y_{1t} | $y_{1c} + y_{1t}$ |
| Failure | y_{2c} | y_{2t} | $y_{2c} + y_{2t}$ |
| Col_Total | n_c | n_t | N |

R Technical Reference and Formulas: Analysis

The probability of observing this table is $f_0(\mathbf{Y})$ which, as shown by equation (R.79), contains an unknown (nuisance) parameter, π . So long as the probability of observing any generic 2×2 table depends on π , exact inference is not possible, since the p-value is based on summing up the probabilities of many such tables, each depending on an unknown parameter. The key to exact inference is getting rid of π , the nuisance parameter. The unconditional approach is to eliminate π by taking a supremum over its entire range so as to provide for the worst-case. (Barnard, 1945, was the first to propose this idea.) The unconditional probability of observing \mathbf{x} under H_0 is $f_0(\mathbf{x})$, specified by equation (R.79). In order to compute an exact p-value we need to specify a reference set of 2×2 contingency tables and sum the probabilities of tables that are at least as extreme as \mathbf{x} in it. Unconditional inference uses reference set of 2×2 contingency tables in which only the column sums, or the binomial sample sizes, are fixed. The row sums are treated as random variables. Denote this reference set by

$$\Omega = \{ \mathbf{Y}: y_{1j} + y_{2j} = n_j, j = c, t \}, \tag{R.80}$$

and order each table $\mathbf{Y} \in \Omega$ according to the test statistic

$$D(\mathbf{Y}) = \frac{\hat{\pi}_t - \hat{\pi}_c}{\sqrt{\left(\frac{y_{1c} + y_{1t}}{N}\right)\left(\frac{y_{2c} + y_{2t}}{N}\right)\left(\frac{1}{n_c} + \frac{1}{n_t}\right)}}, \tag{R.81}$$

where $\hat{\pi}_j = y_{1j}/n_j, j = c, t$. If $y_{11} = y_{12} = 0$, or $y_{21} = y_{22} = 0$, set $D(\mathbf{Y}) = 0$. The denominator of (R.81) is the standard error of the observed difference of binomial proportions under the null hypothesis. Therefore the statistic $D(\mathbf{Y})$ has a mean of 0 and variance of 1 under H_0 . A large positive value for the observed statistic $D(\mathbf{x})$ furnishes evidence against H_1 while a large negative value furnishes evidence against H_1' . The exact p-value is the sum of probabilities of all tables $\mathbf{Y} \in \Omega$ that are more extreme than the observed table \mathbf{x} with respect to the test statistic (R.81). The trouble is that each such extreme table has a probability $f_0(\mathbf{Y})$ which, by equation (R.79) depends on the unknown nuisance parameter, π . We compute the p-value in two stages. At the first stage we express the p-value as a function of π . Then, at the second stage, we obtain the supremum of this function over all values of $\pi \in (0, 1)$. We use this supremum as the p-value. Since the p-value based on the actual value of π can never exceed the supremum over all possible values of π , this procedure guarantees that the type-1 error will always be preserved. In effect we compute a conservative p-value that will preserve the desired type-1 error rate no matter what the true value of π might be, since it is designed to cater for the worst case. Specifically, the exact one-sided p-value given π is computed as

$$p_1(\pi) = \min \left\{ \sum_{D(\mathbf{Y}) \leq D(\mathbf{x})} f_0(\mathbf{Y}), \sum_{D(\mathbf{Y}) \geq D(\mathbf{x})} f_0(\mathbf{Y}) \right\}. \tag{R.82}$$

The exact two-sided p-value given π is computed as

$$p_2(\pi) = \sum_{|D(\mathbf{Y})| \geq |D(\mathbf{x})|} f_0(\mathbf{Y}) . \tag{R.83}$$

Finally we obtain one and two-sided p-values that are independent of π by taking a supremum over all possible values of π and arguing that even in the worst possible case, the true p-value could never exceed the supremum. Thus

$$p_1 = \sup\{p_1(\pi): 0 \leq \pi \leq 1\} \tag{R.84}$$

and

$$p_2 = \sup\{p_2(\pi): 0 \leq \pi \leq 1\} . \tag{R.85}$$

R.5.2 Exact Test of Noninferiority: Difference of Proportions

An important biomedical application arises in so-called “active control” clinical trials. In these studies the goal is to demonstrate the noninferiority rather than the superiority of the new treatment relative to the active control. Define the difference in proportions

$$\delta = \pi_t - \pi_c . \tag{R.86}$$

In a noninferiority clinical trial the objective is not to demonstrate that the experimental treatment is superior to the control but rather to demonstrate that the experimental treatment is not significantly inferior. Accordingly a noninferiority margin, $\delta_0 > 0$, is specified a priori and we test the null hypothesis of inferiority.

$$H_0: \delta \geq \delta_0 \tag{R.87}$$

versus the one sided alternative hypothesis of noninferiority

$$H_1: \delta < \delta_0 . \tag{R.88}$$

The test is carried out under the assumption that δ is at its threshold null value $\delta = \delta_0$.

When $\delta_0 < 0$, **East** tests the null hypothesis $H_0: \delta \leq \delta_0$ against the alternative hypothesis $H_1: \delta > \delta_0$. When $\delta_0 > 0$, the null hypothesis $H_0: \delta \geq \delta_0$ is tested against the alternative hypothesis $H_1: \delta < \delta_0$. Let $\hat{\pi}_t$ and $\hat{\pi}_c$ be the sample proportions based on n_t and n_c observations in the treatment and control arm. Then the estimate of δ is $\hat{\delta} = \hat{\pi}_t - \hat{\pi}_c$. Test statistics for Wald test and Score Test are defined as follows:

R Technical Reference and Formulas: Analysis

Noninferiority (WALD)

$$Z = \frac{\hat{\pi}_t - \hat{\pi}_c - \delta_0}{\sqrt{\frac{\hat{\pi}_c(1 - \hat{\pi}_c)}{n_c} + \frac{\hat{\pi}_t(1 - \hat{\pi}_t)}{n_t}}} \sim N(0, 1) \quad (\text{R.89})$$

Noninferiority (Score)

$$Z = \frac{\tilde{\pi}_t - \tilde{\pi}_c - \delta_0}{\sqrt{\frac{\tilde{\pi}_c(1 - \tilde{\pi}_c)}{n_c} + \frac{\tilde{\pi}_t(1 - \tilde{\pi}_t)}{n_t}}} \sim N(0, 1) \quad (\text{R.90})$$

where $\tilde{\pi}_t$ and $\tilde{\pi}_c$ are the restricted mle's of π_t and π_c as suggested by Mittinen and Nurminen(1985) whereas the test statistic has been recommended by Farrington and Manning(1990).

Z is distributed as variable that follows $N(0, 1)$ distribution under the null hypothesis.

Exact Inference

Let $\mathbf{Y} \in \Omega$ denote any generic 2×2 table of the form of Table R.3 that might be observed if we generated n_c independent bernoulli trials each with probability π_c and n_t independent bernoulli trials each with probability π_t . The probability of observing any $\mathbf{Y} \in \Omega$ under H_0 is

$$f_{\pi_c, \delta_0}(\mathbf{y}) = \binom{n_c}{y_{1c}} \binom{n_t}{y_{1t}} \pi_c^{y_{1c}} (1 - \pi_c)^{y_{2c}} (\pi_c + \delta_0)^{y_{1t}} (1 - \pi_c - \delta_0)^{y_{2t}} \quad (\text{R.91})$$

The test statistic (see Chan, 1998) is defined as

$$D(\mathbf{Y}) = \frac{\hat{\pi}_t - \hat{\pi}_c - \delta_0}{\sqrt{\frac{(\hat{\pi}_c)(1 - \hat{\pi}_c)}{n_c} + \frac{(\hat{\pi}_t)(1 - \hat{\pi}_t)}{n_t}}} \quad (\text{R.92})$$

where

$$\hat{\pi}_j = \frac{y_{1j}}{n_j}, \quad (\text{R.93})$$

for $j = c, t$, and $\tilde{\pi}_c$ and $\tilde{\pi}_t$ are the maximum likelihood estimates of π_c and π_t , respectively, restricted under the null hypothesis so as to satisfy the requirement $\tilde{\pi}_t - \tilde{\pi}_c = \delta_0$. Mittinen and Nurminen (1985) have shown that one may obtain these restricted maximum likelihood estimates by solving the third degree likelihood equation

$$\sum_{k=1}^3 L_k \tilde{\pi}_c^k = 0 \quad (\text{R.94})$$

for $\tilde{\pi}_c$ and setting $\tilde{\pi}_t = \tilde{\pi}_c + \delta_0$, where

$$L_3 = N ,$$

$$L_2 = (n_t + 2n_c)\delta_0 - N - y_{1c} - y_{1t} ,$$

$$L_1 = (n_c\delta - N - 2y_{1c})\delta_0 + y_{1c} + y_{1t} ,$$

$$L_0 = y_{1c}\delta_0(1 - \delta_0) .$$

The test statistic (R.92) is known as the score statistic. Under H_0 this test statistic has mean 0 and variance 1. Let the data in Table R.1, denoted by \mathbf{x} , be the 2×2 table actually observed. Then the observed value of the test statistic is $D(\mathbf{x})$, and the left tail of the distribution of $D(\mathbf{Y})$ at its observed value under H_0 is

$$P_{\pi_c, \delta_0}(D(\mathbf{x})) = \sum_{\mathbf{D}(\mathbf{Y}) \leq \mathbf{D}(\mathbf{x})} \mathbf{f}_{\pi_c, \delta_0}(\mathbf{Y}) \quad (\text{R.95})$$

If we knew the value of π_c , then $P_{\pi_c, \delta_0}(D(\mathbf{x}))$ would be the exact p-value for testing H_0 versus H_1 . Since π_c is unknown, however, we take the supremum of (R.95) over all values of π_c in its range, just as we did for Barnard's test in Section R.5.1. This produces a conservative p-value that is guaranteed to ensure that the true type-1 error of the test will never exceed its nominal significance level. Since $\delta_0 > 0$ the range of possible values for π_c is

$$I(\delta_0) = \{\pi_c : 0 < \pi_c < 1 - \delta_0\} . \quad (\text{R.96})$$

Thereupon the unconditional exact one-sided p-value is

$$p_1 \equiv P_{\delta_0}(D(\mathbf{x})) = \sup\{\mathbf{P}_{\pi_c, \delta_0}(\mathbf{D}(\mathbf{x})) : \pi_c \in I(\delta_0)\} . \quad (\text{R.97})$$

Note that in practice the supremum in equation (R.97) is taken over a restricted range for π rather than over the entire range $I(\delta_0)$. This restriction, proposed by Berger and Boos (1994), adds stability and reduces the conservatism of the procedure. The p-values are suitably adjusted so that the restricted search for the supremum does not compromise the type-1 error. Finally, it is worth noting that when $\delta_0 = 0$ the above p-value specializes to the left tail p-value obtained by Barnard's test.

Additional Remarks

1. The score statistic $D(\mathbf{Y})$ specified by equation (R.92) is always defined except for the special case where $y_{1c} = y_{1t} = 0$ and $\delta_0 = 0$ or where $y_{2c} = y_{2t} = 0$ and $\delta_0 = 0$. For these special cases the one- and two-sided p-values are both set to 1. These special cases never arise when performing a noninferiority test with $\delta_0 > 0$.

R Technical Reference and Formulas: Analysis

2. The one-sided asymptotic p-value corresponding to p_1 is obtained by assuming that the test statistic $D(\mathbf{Y})$ converges in distribution to the standard normal.

Thus

$$\tilde{p}_1 = 1 - \Phi(D(\mathbf{x})) . \tag{R.98}$$

3. An alternative equivalent way to perform a level- α test of non-inferiority is to compute an exact $100 \times (1 - \alpha)$ lower confidence bound for δ , say δ_L , using the method described in Section R.5.4. If $\delta_L < \delta_0$ we reject the null hypothesis of inferiority.

R.5.3 Exact Test of Equivalence: Difference of Proportions

Suppose π_c is the response rate of Control and π_t is the response rate of Treatment. Define the absolute difference in proportions

$$\delta = |\pi_t - \pi_c| . \tag{R.99}$$

Suppose that for a pre-specified equivalence margin $\delta_0 > 0$ we wish to test the null hypothesis of inequivalence

$$H_0: \delta \geq \delta_0 \tag{R.100}$$

against the alternative hypothesis of equivalence

$$H_1: \delta < \delta_0 . \tag{R.101}$$

We test the above null hypothesis by performing two separate one-sided non-inferiority hypothesis tests of the form

$$H_{01}: \pi_c - \pi_t \geq \delta_0 \text{ versus } H_{11}: \pi_c - \pi_t < \delta_0 \tag{R.102}$$

and

$$H_{02}: \pi_t - \pi_c \geq \delta_0 \text{ versus } H_{12}: \pi_t - \pi_c < \delta_0 . \tag{R.103}$$

Each hypothesis test is carried out separately using the method described in Section R.5.2. Hypothesis test H_{01} is performed under the assumption that $\pi_c - \pi_t$ is at its threshold null value $\pi_c - \pi_t = \delta_0$. Similarly hypothesis test H_{02} is tested under the assumption that $\pi_t - \pi_c$ is at its threshold null value $\pi_t - \pi_c = \delta_0$. We reject the null hypothesis of inequivalence and accept the alternative hypothesis of equivalence only if **both** H_{01} and H_{02} are rejected. The probability of observing any $\mathbf{Y} \in \Omega$ under H_{01} is

$$f_{\pi_c, \delta_0}^{01}(\mathbf{Y}) = \binom{n_c}{y_{1c}} \binom{n_t}{y_{1t}} \pi_c^{y_{1c}} (1 - \pi_c)^{y_{2c}} (\pi_c - \delta_0)^{y_{1t}} (1 - \pi_c + \delta_0)^{y_{2t}} , \tag{R.104}$$

and the statistic used to test H_{01} is

$$D^{01}(\mathbf{Y}) = \frac{\hat{\pi}_c - \hat{\pi}_t - \delta_0}{\sqrt{\frac{(\hat{\pi}_c)(1-\hat{\pi}_c)}{n_c} + \frac{(\hat{\pi}_t)(1-\hat{\pi}_t)}{n_t}}}, \quad (\text{R.105})$$

where the $\tilde{\pi}_c$ and $\tilde{\pi}_t$ are restricted maximum likelihood estimates of π_c and π_t , respectively, under the restriction that $\tilde{\pi}_c - \tilde{\pi}_t = \delta_0$. We compute

$$P_{\tilde{\pi}_c, \delta_0}^{01}(D(\mathbf{x})) = \sum_{\mathbf{D}^{01}(\mathbf{Y}) \leq \mathbf{D}^{01}(\mathbf{x})} \mathbf{f}_{\tilde{\pi}_c, \delta_0}^{01}(\mathbf{Y}), \quad (\text{R.106})$$

and then take the supremum over all $\pi_c \in I^{01}(\delta_0)$ where

$$I^{(01)}(\delta_0) = \{\pi_c : \delta_0 < \pi_c < 1\}. \quad (\text{R.107})$$

The exact unconditional one-sided p-value for testing H_{01} is thus

$$p_{01} \equiv P_{\delta_0}^{01}(D(\mathbf{x})) = \sup\{\mathbf{P}_{\tilde{\pi}_c, \delta_0}^{01}(\mathbf{D}(\mathbf{x})) : \pi_c \in I^{01}(\delta_0)\}. \quad (\text{R.108})$$

The probability of observing any $\mathbf{Y} \in \Omega$ under H_{02} is

$$f_{\pi_c, \delta_0}^{02}(\mathbf{Y}) = \binom{n_c}{y_{1c}} \binom{n_t}{y_{1t}} \pi_c^{y_{1c}} (1 - \pi_c)^{y_{2c}} (\pi_c + \delta_0)^{y_{1t}} (1 - \pi_c - \delta_0)^{y_{2t}}, \quad (\text{R.109})$$

and the statistic used to test H_{02} is

$$D^{02}(\mathbf{Y}) = \frac{\hat{\pi}_t - \hat{\pi}_c - \delta_0}{\sqrt{\frac{(\hat{\pi}_c)(1-\hat{\pi}_c)}{n_c} + \frac{(\hat{\pi}_t)(1-\hat{\pi}_t)}{n_t}}}, \quad (\text{R.110})$$

where the $\tilde{\pi}_c$ and $\tilde{\pi}_t$ are maximum likelihood estimates of π_c and π_t , respectively, under the restriction that $\tilde{\pi}_t - \tilde{\pi}_c = \delta_0$. We compute

$$P_{\tilde{\pi}_c, \delta_0}^{02}(D(\mathbf{x})) = \sum_{\mathbf{D}^{02}(\mathbf{Y}) \leq \mathbf{D}^{02}(\mathbf{x})} \mathbf{f}_{\tilde{\pi}_c, \delta_0}^{02}(\mathbf{Y}), \quad (\text{R.111})$$

and then take the supremum over all $\pi_c \in I^{02}(\delta_0)$ where

$$I^{(02)}(\delta_0) = \{\pi_c : 0 < \pi_c < 1 - \delta_0\}. \quad (\text{R.112})$$

The exact unconditional one-sided p-value for testing H_{02} is thus

$$p_{02} \equiv P_{\delta_0}^{02}(D(\mathbf{x})) = \sup\{\mathbf{P}_{\tilde{\pi}_c, \delta_0}^{02}(\mathbf{D}(\mathbf{x})) : \pi_c \in I^{02}(\delta_0)\}. \quad (\text{R.113})$$

Additional Remarks

R Technical Reference and Formulas: Analysis

1. The test statistics $D^{01}(\mathbf{Y})$ and $D^{02}(\mathbf{Y})$, specified by equations (R.105) and (R.110), respectively, are always defined except for the special cases where $y_{11} = y_{12} = 0$ and $\delta_0 = 0$ or where $y_{21} = y_{22} = 0$ and $\delta_0 = 0$. For these special case the one- and two-sided p-values are both set to 1. These special cases never arise when performing an equivalence test with $\delta_0 > 0$.

2. The one-sided asymptotic p-value corresponding to p_{01} is obtained by assuming that the test statistic $D^{01}(\mathbf{Y})$ converges in distribution to the standard normal. Thus,

$$\tilde{p}_{01} = 1 - \Phi(D^{01}(\mathbf{x})) . \tag{R.114}$$

3. The one-sided asymptotic p-value corresponding to p_{02} is obtained by assuming that the test statistic $D^{02}(\mathbf{Y})$ converges in distribution to the standard normal. Thus

$$\tilde{p}_{02} = 1 - \Phi(D^{02}(\mathbf{x})) . \tag{R.115}$$

4. An alternative equivalent way to perform a level- α test of equivalence is to compute an exact $100 \times (1 - 2\alpha)$ confidence interval for δ , say (δ_L, δ_U) , using the method described in Section R.5.4. If δ_0 is excluded from this interval, we reject the null hypothesis of equivalence.

R.5.4 Unconditional Exact Confidence Intervals for the Difference of Proportions

Suppose π_c is the binomial response rate of Control and π_t is the binomial response rate of Treatment. We wish to compute an exact $100(1 - \alpha)\%$ confidence interval for

$$\delta = \pi_t - \pi_c .$$

We use a test based procedure. That is, we invert hypothesis tests of the form $\delta = \delta_0$, where, in general, $\delta_0 \neq 0$. If we are dealing with the superiority, δ_0 will be zero.

In case of noninferiority δ_0 is nonzero. Accordingly, this section is applicable to superiority, noninferiority and equivalence. There is one further complication, however, since the p-values which we compute under these alternative hypotheses depend on a nuisance parameter. We handle this problem the same way we handled it for Barnard's unconditional exact hypothesis test; i.e., by taking a supremum over all possible values of the nuisance parameter.

Interval Estimation

Suppose we take n_c independent Bernoulli samples from control and n_t independent Bernoulli samples from treatment. Let $\mathbf{Y} \in \Omega$ (see Table R.3 denote any generic 2×2 table that might be observed, and let \mathbf{x} (see Table R.1), be the 2×2 table that was actually observed. Define

$$\hat{\pi}_j = \frac{y_{1j}}{n_j}$$

for $j = c, t$. In **East**, we provide a test based exact confidence interval using the standardized statistic

$$D(\mathbf{Y}) = \frac{\hat{\pi}_t - \hat{\pi}_c - \delta_0}{\sqrt{\frac{(\tilde{\pi}_c)(1-\tilde{\pi}_c)}{n_c} + \frac{(\tilde{\pi}_t)(1-\tilde{\pi}_t)}{n_t}}} \quad (\text{R.116})$$

where $\tilde{\pi}_c$ and $\tilde{\pi}_t$ are the maximum likelihood estimates of π_c and π_t computed, under the restriction that $\tilde{\pi}_t - \tilde{\pi}_c = \delta_0$. This statistic is known as the score statistic. The use of (R.116) as the test statistic has been proposed by Farrington and Manning(1990) for asymptotic confidence intervals and by Chan and Zhang (1999) for exact confidence intervals. We note that the score statistic specified by equation (R.116) is always defined except for the special cases where $y_{1c} = y_{1t} = 0$ and $\delta_0 = 0$, or $y_{2c} = y_{2t} = 0$ and $\delta = 0$. These special cases never arise when computing a confidence interval.

Test Based Exact Confidence Intervals: Inverting Two One-Sided Tests

Let (δ_*, δ^*) be the desired $100(1 - \alpha)\%$ exact confidence interval, evaluated at $D(\mathbf{x})$, the observed value of the test statistic. This exact confidence interval may be constructed by inverting two one-sided hypothesis tests, each at the $\alpha/2$ significance level, under appropriate alternative hypotheses about δ . The probability of observing any $\mathbf{Y} \in \Omega$, for any given value of δ , is

$$f_{\pi_c, \delta}(\mathbf{Y}) = \binom{n_c}{y_{1c}} \binom{n_t}{y_{1t}} \pi_c^{y_{1c}} (1 - \pi_c)^{y_{2c}} (\pi_c + \delta)^{y_{1t}} (1 - \pi_c - \delta)^{y_{2t}} . \quad (\text{R.117})$$

Define

$$P_{\pi_c, \delta}(D(\mathbf{x})) = \sum_{\mathbf{D}(\mathbf{Y}) \leq \mathbf{D}(\mathbf{x})} \mathbf{f}_{\pi_c, \delta}(\mathbf{Y}) \quad (\text{R.118})$$

and

$$Q_{\pi_c, \delta}(D(\mathbf{x})) = \sum_{\mathbf{D}(\mathbf{Y}) \geq \mathbf{D}(\mathbf{x})} \mathbf{f}_{\pi_c, \delta}(\mathbf{Y}) . \quad (\text{R.119})$$

We must eliminate the nuisance parameter π_c from equations (R.118) and (R.119) by taking the supremum over its range. It is easy to see that the permissible range for π_c given δ is the interval

$$I(\delta) = \{\pi_c: \max(0, -\delta) \leq \pi_c \leq \min(1, 1 - \delta)\} . \quad (\text{R.120})$$

R Technical Reference and Formulas: Analysis

Thus we define

$$P_\delta(D(\mathbf{x})) = \sup\{\mathbf{P}_{\pi_c, \delta}(\mathbf{D}(\mathbf{x})): \pi_c \in \mathbf{I}(\delta)\} \quad (\text{R.121})$$

and

$$Q_\delta(D(\mathbf{x})) = \sup\{\mathbf{Q}_{\pi_c, \delta}(\mathbf{D}(\mathbf{x})): \pi_c \in \mathbf{I}(\delta)\} . \quad (\text{R.122})$$

Starting with $\delta = -1$, the desired lower confidence bound is obtained by increasing the value of δ until we find a value, denoted by δ_* , such that the equality

$$Q_{\delta_*}(D(\mathbf{x})) = \alpha/2 \quad (\text{R.123})$$

is satisfied but for any $\delta < \delta_*$, $Q_\delta(D(\mathbf{x})) < \alpha/2$. The upper confidence bound, δ^* is obtained in an analogous fashion. Starting with $\delta = 1$, the desired upper confidence bound is obtained by decreasing the value of δ until we find a value, denoted by δ^* , such that the equality

$$P_{\delta^*}(D(\mathbf{x})) = \alpha/2 \quad (\text{R.124})$$

is satisfied but for any $\delta > \delta^*$, $P_\delta(D(\mathbf{x})) < \alpha/2$.

East reports (δ_*, δ^*) as the $100 \times (1 - \alpha)\%$ confidence interval for the parameter δ . Suppose δ_0 is the true (unknown) value of δ . The long run relative frequency with which, in repeated trials, this interval excludes δ_0 is $\Pr(\delta_0 \leq \delta_*) + \Pr(\delta_0 \geq \delta^*)$. We shall show at the end of this Section that neither term in the above sum can exceed $\alpha/2$. Therefore the probability of the confidence interval excluding δ_0 cannot exceed α . However, due to the discreteness of the distribution of $D(\mathbf{Y})$, and the conservatism induced by taking a supremum over all $\pi_c \in \mathbf{I}(\delta)$, the above exclusion probability is usually less than α instead of equaling α . Thus, (δ_*, δ^*) may be regarded as a conservative confidence interval. In addition to the exact confidence interval, **East** also reports an exact one-sided p-value, defined as the smaller of the two tail areas,

$$p_1 = \min(P_0(D(\mathbf{x})), \mathbf{Q}_0(\mathbf{D}(\mathbf{x}))) , \quad (\text{R.125})$$

and the two sided exact p-value is twice the one-sided:

$$p_2 = 2p_1 . \quad (\text{R.126})$$

The two-sided p-value is weakly consistent with the corresponding exact confidence interval for δ . That is, if $0 \notin [\delta_*, \delta^*]$ then $p_t < \alpha$. The stronger consistency requirement, that $p_t < \alpha$ if and only if $0 \notin [\delta_*, \delta^*]$ cannot be established unless $P_\delta(D(\mathbf{x}))$ and $Q_\delta(D(\mathbf{x}))$ are monotone functions of δ for any given $D(\mathbf{x})$. This need not be the case, however.

The above procedure can be slightly modified in practice. The suprema in equations (R.121) and (R.122) can be taken over a restricted range for π rather than

over the entire range $I(\delta)$. This restriction, proposed by Berger and Boos (1994), adds stability and reduces the conservatism of the procedure. The right hand sides of equations (R.123) and (R.124) are suitably adjusted so that the restricted search for the supremum does not compromise the coverage properties of the resulting confidence interval.

Proof of Coverage: We shall now prove that the probability that the above confidence interval excludes the true parameter δ_0 cannot exceed α . For simplicity, denote the random variable $D(\mathbf{Y})$ by D , and its observed value $D(\mathbf{x})$ by d . In order to make explicit the dependence of the confidence interval on d , denote the lower confidence bound by $\delta_*(d)$ and the upper confidence bound by $\delta^*(d)$. Thus the lower confidence bound satisfies the relationship

$$Q_{\delta_*(d)}(D(\mathbf{x})) = \alpha/2, \tag{R.127}$$

and furthermore, by the way we conduct the search for $\delta_*(d)$,

$$Q_\delta(D(\mathbf{x})) < \alpha/2, \text{ if } \delta < \delta_*(d). \tag{R.128}$$

Define $H(\delta_0)$ to be the *smallest* value of D satisfying the inequality

$$Q_{\delta_0}(H(\delta_0)) \leq \alpha/2. \tag{R.129}$$

Observe, from the definition of $H(\delta_0)$ in (R.129), that if $d < H(\delta_0)$ we must have $Q_{\delta_0}(d) > \alpha/2$. But we know from (R.128) that there is no value of $\delta \leq \delta_*(d)$ for which $Q_\delta(d) > \alpha/2$. Therefore if $d < H(\delta_0)$, it must be the case that $\delta_0 > \delta_*(d)$. It follows that

$$\Pr\{\delta_0 > \delta_*(d)\} \geq \Pr\{D < H(\delta_0) | \delta_0, \pi_1\}. \tag{R.130}$$

We use a weak inequality instead of a strict equality in (R.130) because it is also possible in some situations to have $\delta_0 > \delta_*(d)$ when $d \geq H(\delta_0)$. Taking the complementary probability on both sides of (R.130) we have

$$\Pr\{\delta_0 \leq \delta_*(d)\} \leq \Pr\{D \geq H(\delta_0) | \delta_0, \pi_c\}. \tag{R.131}$$

Taking the supremum over all $\pi_c \in I(\delta_0)$ on the right hand side of (R.131) we have

$$\Pr\{\delta_* \geq \delta_0\} \leq Q_{\delta_0}(H(\delta_0)) \leq \alpha/2. \tag{R.132}$$

By an analogous argument we can establish that

$$\Pr\{\delta^* \leq \delta_0\} \leq \alpha/2. \tag{R.133}$$

Therefore the probability that the interval (δ_*, δ^*) excludes the parameter δ_0 cannot exceed α .

R Technical Reference and Formulas: Analysis

Asymptotic Confidence Interval

East computes asymptotic p-values and test based asymptotic confidence intervals for δ , under the assumption that the test statistic is asymptotically normally distributed. The asymptotic $100 \times (1 - \alpha)\%$ confidence interval $(\tilde{\delta}_*, \tilde{\delta}^*)$ is obtained by inverting the corresponding one-sided hypothesis tests. Thus $\tilde{\delta}_*$ satisfies the equality

$$1 - \Phi \left\{ \frac{x_{1t}/n_t - x_{1c}/n_c - \delta_*}{\sqrt{\frac{(\tilde{\pi}_c)(1-\tilde{\pi}_c)}{n_c} + \frac{(\tilde{\pi}_t)(1-\tilde{\pi}_t)}{n_t}}} \right\} = \alpha/2, \quad (\text{R.134})$$

where $\tilde{\pi}_c$ and $\tilde{\pi}_t$ are the maximum likelihood estimates of π_c and π_t , respectively, under the null hypothesis that $\pi_t - \pi_c = \delta_*$. Similarly $\tilde{\delta}^*$ satisfies the equality

$$\Phi \left\{ \frac{x_{1t}/n_t - x_{1c}/n_c - \delta^*}{\sqrt{\frac{(\tilde{\pi}_c)(1-\tilde{\pi}_c)}{n_c} + \frac{(\tilde{\pi}_t)(1-\tilde{\pi}_t)}{n_t}}} \right\} = \alpha/2, \quad (\text{R.135})$$

where $\tilde{\pi}_c$ and $\tilde{\pi}_t$ are the maximum likelihood estimates of π_c and π_t , respectively, under the null hypothesis that $\pi_t - \pi_c = \delta^*$.

R.5.5 Unconditional Exact Confidence Intervals for the Ratio of Proportions

In the **Ratio of Proportions** test, let π_t and π_c denote the proportions of the successes from the experimental treatment (T) and the control treatment (C), respectively. To test the null hypothesis $H_0: \pi_t/\pi_c = 1$ against the 2-sided alternative hypothesis

$H_1: \pi_t/\pi_c \neq 1$ or a 1-sided alternative hypothesis $H_1: \pi_t/\pi_c < 1$ or $H_1': \pi_t/\pi_c > 1$.

Test Statistic Using the pooled estimate of variance:

$$Z = \frac{\ln(\hat{\pi}_t) - \ln(\hat{\pi}_c)}{\sqrt{\frac{(1-\hat{\pi})}{\hat{\pi}} \left(\frac{1}{n_t} + \frac{1}{n_c} \right)}},$$

where

$$\hat{\pi} = \frac{n_t \hat{\pi}_t + n_c \hat{\pi}_c}{n_t + n_c},$$

where $\hat{\pi}_t$ and $\hat{\pi}_c$ are the sample proportions based on n_t and n_c observations in the treatment and control arm, respectively. Asymptotically, Z is distributed as variable that follows $N(0, 1)$ distribution under the null hypothesis.

We wish to compute an exact $100(1 - \alpha)\%$ confidence interval for

$$\rho = \frac{\pi_t}{\pi_c}. \quad (\text{R.136})$$

The procedure parallels that described in Section R.5.4 for the difference of proportions, δ . We invert hypothesis tests of the form $\rho = \rho_0$, where, in general, $\rho_0 \neq 1$. There is one further complication, however, since the p-values which we compute under these alternative hypotheses depend on a nuisance parameter. We handle this problem the same way we handled it for Barnard's unconditional exact hypothesis test; by taking a supremum over all possible values of the nuisance parameter. Suppose we take n_c independent Bernoulli samples from control and n_t independent Bernoulli samples from treatment. Let $\mathbf{Y} \in \Omega$ (see Table R.3, page 2596) denote any generic 2×2 table that might be observed, and let \mathbf{x} (see Table R.1, page 2568), be the 2×2 table that was actually observed. Define

$$\hat{\pi}_j = \frac{y_{1j}}{n_j} \tag{R.137}$$

for $j = c, t$. In **East** we provide a test based exact confidence interval using the standardized statistic.

$$D(\mathbf{Y}) = \frac{\hat{\pi}_t - \rho_0 \hat{\pi}_c}{\sqrt{\frac{(\hat{\pi}_t)(1-\hat{\pi}_t)}{n_t} + \frac{\rho_0^2(\hat{\pi}_c)(1-\hat{\pi}_c)}{n_c}}} \tag{R.138}$$

where $\hat{\pi}_c$ and $\hat{\pi}_t$ are the restricted maximum likelihood estimates of π_c and π_t computed, under the restriction that $\hat{\pi}_t/\hat{\pi}_c = \rho_0$. The restricted MLE's are suggested by Miettinen and Nurminen (1985). The use of (R.164) as the test statistic has been proposed by Farrington and Manning (1990) for asymptotic confidence intervals and by Chan and Zhang (1999) for exact confidence intervals.

Test Based Exact Confidence Intervals: Inverting Two One-Sided Tests

Let (ρ_*, ρ^*) be the desired $100(1 - \alpha)\%$ exact confidence interval, evaluated at $D(\mathbf{x})$, the observed value of the test statistic. This exact confidence interval may be constructed by inverting two one-sided hypothesis tests, each at the $\alpha/2$ significance level, under appropriate alternative hypotheses about ρ . The computations are very similar to the p-value computations performed in the next Section. The probability of observing any $\mathbf{Y} \in \Omega$ for any given value of ρ is

$$f_{\pi_c, \rho}(\mathbf{Y}) = \binom{n_c}{y_{1c}} \binom{n_t}{y_{1t}} \pi_c^{y_{1c}} (1 - \pi_c)^{y_{2c}} (\rho \pi_c)^{y_{1t}} (1 - \rho \pi_c)^{y_{2t}} . \tag{R.139}$$

Define

$$P_{\pi_c, \rho}(D(\mathbf{x})) = \sum_{\mathbf{D}(\mathbf{Y}) \leq \mathbf{D}(\mathbf{x})} f_{\rho \pi_c}(\mathbf{Y}) \tag{R.140}$$

and

$$Q_{\pi_c, \rho}(D(\mathbf{x})) = \sum_{\mathbf{D}(\mathbf{Y}) \geq \mathbf{D}(\mathbf{x})} f_{\rho \pi_c}(\mathbf{Y}) . \tag{R.141}$$

R Technical Reference and Formulas: Analysis

We must eliminate the nuisance parameter π_c from equations (R.166) and (R.167) by taking the supremum over its range. It is easy to see that the permissible range for π_c given ρ is the interval

$$I(\rho) = \{\pi_c: 0 \leq \pi_c \leq \min(1/\rho, 1)\} . \quad (\text{R.142})$$

Thus we define

$$P_\rho(D(\mathbf{x})) = \sup\{\mathbf{P}_{\pi_c, \rho}(\mathbf{D}(\mathbf{x})): \pi_c \in I(\rho)\} \quad (\text{R.143})$$

and

$$Q_{\rho}(D(\mathbf{x})) = \sup\{\mathbf{Q}_{\pi_c, \rho}(\mathbf{D}(\mathbf{x})): \pi_c \in I(\rho)\} . \quad (\text{R.144})$$

Starting with $\rho = 0$, the desired lower confidence bound is obtained by increasing the value of ρ until we find a value, denoted by ρ_* , such that the equality

$$Q_{\rho_*}(D(\mathbf{x})) = \alpha/2 \quad (\text{R.145})$$

is satisfied but for any $\rho < \rho_*$, $Q_\rho(D(\mathbf{x})) < \alpha/2$. The upper confidence bound, ρ^* is obtained in an analogous fashion. Starting with $\rho = \infty$ (i.e., a very large positive number), the desired upper confidence bound is obtained by decreasing the value of ρ until we find a value, denoted by ρ^* , such that the equality

$$P_{\rho^*}(D(\mathbf{x})) = \alpha/2 \quad (\text{R.146})$$

is satisfied but for any $\rho > \rho^*$, $P_\rho(D(\mathbf{x})) < \alpha/2$. **East** reports (ρ_*, ρ^*) as the $100 \times (1 - \alpha)\%$ confidence interval for the parameter ρ . Suppose ρ_0 is the true (unknown) value of ρ . The long run relative frequency with which, in repeated trials, this interval excludes ρ_0 is $\Pr(\rho_0 \leq \rho_*) + \Pr(\rho_0 \geq \rho^*)$. Using arguments similar to those given on page 2579 for the binomial difference, δ_0 , we can show that neither term in the above sum can exceed $\alpha/2$. Therefore the probability of the confidence interval excluding ρ_0 cannot exceed α . However, due to the discreteness of the distribution of $D(\mathbf{Y})$, and the conservatism induced by taking a supremum over all $\pi_c \in I(\rho)$, the above exclusion probability is usually less than α instead of equaling α . Thus, (ρ_*, ρ^*) may be regarded as a conservative confidence interval. In addition to the exact confidence interval, **East** also reports an exact one-sided p-value, defined as the smaller of the two tail areas,

$$p_c = \min(P_0(D(\mathbf{x})), \mathbf{Q}_0(\mathbf{D}(\mathbf{x}))) , \quad (\text{R.147})$$

and the two sided exact p-value is twice the one-sided:

$$p_t = 2p_c . \quad (\text{R.148})$$

The two-sided p-value is weakly consistent with the corresponding exact confidence interval for ρ . That is, if $1 \notin [\rho_*, \rho^*]$ then $p_t < \alpha$. The stronger consistency

requirement, that $p_t < \alpha$ if and only if $1 \notin [\rho_*, \rho^*]$ cannot be established unless $P_\rho(D(\mathbf{x}))$ and $Q_\rho(D(\mathbf{x}))$ are monotone functions of ρ for any given $D(\mathbf{x})$. This need not be the case, however.

R.5.6 Exact Test of Noninferiority: Ratio of Proportions

Suppose π_c is the response rate of an experimental treatment and π_t is the response rate of an active control treatment. Define the ratio of binomial proportions

$$\rho = \frac{\pi_t}{\pi_c} . \tag{R.149}$$

In a non-inferiority clinical trial the objective is not to demonstrate that the experimental treatment is superior to the control but rather to demonstrate that the experimental treatment is not significantly inferior. Accordingly a non-inferiority margin, ρ_0 , is specified a priori and we test the null hypothesis of inferiority $H_0 : \rho \leq \rho_0$ against $H_1 : \rho > \rho_0$ if $\rho_0 < 1$ Or $H_0 : \rho \geq \rho_0$ against $H_1 : \rho < \rho_0$ if $\rho_0 > 1$

- Test statistic: (Wald)

$$Z = \frac{\ln(\hat{\pi}_t) - \ln(\hat{\pi}_c) - \ln(\rho_0)}{\sqrt{\frac{(1-\hat{\pi}_t)}{n_t \hat{\pi}_t} + \frac{(1-\hat{\pi}_c)}{n_c \hat{\pi}_c}}},$$

where $\hat{\pi}_t$ and $\hat{\pi}_c$ are the sample proportions based on n_t and n_c observations in the treatment and control arm, respectively. $\delta_0 = \ln(\rho_0)$ is the noninferiority margin. Under Ho, Z follows asymptotic Normal distribution with mean 0 and variance 1. Only asymptotic inference is available with Wald test.

- Test statistic:(Farrington Manning)

$$Z = \frac{\hat{\pi}_t - \rho_0 \hat{\pi}_c}{\sqrt{\frac{\hat{\pi}_t(1-\hat{\pi}_t)}{n_t} + \frac{\rho_0^2 \hat{\pi}_c(1-\hat{\pi}_c)}{n_c}}},$$

where $\hat{\pi}_t$ and $\hat{\pi}_c$ are the sample proportions based on n_t and n_c observations in the treatment and control arm, respectively and $\tilde{\pi}_c$ and $\tilde{\pi}_t$ are the restricted maximum likelihood estimates of π_c and π_t , respectively.

The test is carried out under the assumption that ρ is at its threshold null value $\rho = \rho_0$.

Exact Inference

Let $\mathbf{Y} \in \Omega$ denote any generic 2×2 table of the form of Table R.3 that might be observed if we generated n_c independent bernoulli trials each with probability π_c and

R Technical Reference and Formulas: Analysis

n_t independent bernoulli trials each with probability π_t . The probability of observing any $\mathbf{Y} \in \Omega$ under H_0 is

$$f_{\pi_c, \rho_0}(\mathbf{Y}) = \binom{n_c}{y_{1c}} \binom{n_t}{y_{1t}} \pi_c^{y_{1c}} (1 - \pi_c)^{y_{2c}} (\rho_0 \pi_c)^{y_{1t}} (1 - \rho_0 \pi_c)^{y_{2t}} . \quad (\text{R.150})$$

The test statistic (see Farrington and Manning(1990) is defined as

$$D(\mathbf{Y}) = \frac{\hat{\pi}_t - \rho_0 \hat{\pi}_c}{\sqrt{\frac{(\hat{\pi}_t)(1-\hat{\pi}_t)}{n_t} + \frac{\rho_0^2(\hat{\pi}_c)(1-\hat{\pi}_c)}{n_c}}} \quad (\text{R.151})$$

where

$$\hat{\pi}_j = \frac{y_{1j}}{n_j} , \quad (\text{R.152})$$

for $j = c, t$, and $\tilde{\pi}_c$ and $\tilde{\pi}_t$ are the maximum likelihood estimates of π_c and π_t , respectively, restricted under the null hypothesis to satisfy the requirement that $\tilde{\pi}_t/\tilde{\pi}_c = \rho_0$. Miettinen and Nurminen (1985) have shown that one may obtain these restricted maximum likelihood estimates by solving a quadratic likelihood equation. Thus

$$\tilde{\pi}_1 = \frac{-B - \sqrt{B^2 - 4AC}}{2A} , \quad (\text{R.153})$$

and

$$\tilde{\pi}_t = \rho_0 \tilde{\pi}_c , \quad (\text{R.154})$$

where

$$A = \rho_0 N , \quad (\text{R.155})$$

$$B = -(\rho_0 n_t + y_{1t} + n_c + \rho_0 y_{1c}) , \quad (\text{R.156})$$

$$C = y_{1c} + y_{1t} . \quad (\text{R.157})$$

The test statistic (R.151) is known as the score statistic. Under H_0 this test statistic has mean 0 and variance 1. Let the data in Table R.1, denoted by \mathbf{x} , be the 2×2 table actually observed. Then the observed value of the test statistic is $D(\mathbf{x})$, and the left tail of the distribution of $D(\mathbf{Y})$ at its observed value under H_0 is

$$P_{\pi_c, \rho_0}(D(\mathbf{x})) = \sum_{\mathbf{D}(\mathbf{Y}) \leq \mathbf{D}(\mathbf{x})} \mathbf{f}_{\pi_c, \rho_0}(\mathbf{Y}) . \quad (\text{R.158})$$

If we knew the value of π_c , then $P_{\pi_c, \rho_0}(D(\mathbf{x}))$ would be the exact p-value for testing H_0 versus H_1 . Since π_c is unknown, however, we take the supremum of (R.158) over

all values of π_c in its range, just as we did for Barnard’s test in Section R.5.1. This produces a conservative p-value that is guaranteed to ensure that the true type-1 error of the test will never exceed its nominal significance level. Since $\rho_0 > 1$ the range of possible values for π_c is

$$I(\rho_0) = \{\pi_c : 0 < \pi_c < \min(1, 1/\rho_0)\}. \tag{R.159}$$

Thereupon the unconditional exact one-sided p-value is

$$p_1 \equiv P_{\rho_0}(D(\mathbf{x})) = \sup\{\mathbf{P}_{\pi_c, \rho_0}(D(\mathbf{x})) : \pi_c \in I(\rho_0)\}. \tag{R.160}$$

Note that in practice the supremum in equation (R.160) is taken over a restricted range for π rather than over the entire range $I(\rho_0)$. This restriction, proposed by Berger and Boos (1994), adds stability and reduces the conservatism of the procedure. The p-values are suitably adjusted so that the restricted search for the supremum does not compromise the type-1 error. Finally, it is worth noting that when $\rho_0 = 1$ the above p-value specializes to the left tail p-value obtained by Barnard’s test.

Additional Remarks

1. The score statistic $D(\mathbf{Y})$ specified by equation (R.151) is undefined when $y_{1c} = y_{1t} = 0$ and $\rho_0 = 1$, or when $y_{2c} = y_{2t} = 0$ and $\rho_0 = 1$. For these special cases the one- and two-sided p-values are both set to 1. These special cases never arise when performing a non-inferiority test with $\rho_0 \neq 1$.
2. The one-sided asymptotic p-value corresponding to p_1 is obtained by assuming that the test statistic $D(\mathbf{Y})$ converges in distribution to the standard normal.

Thus,

$$\tilde{p}_1 = 1 - \Phi(D(\mathbf{x})). \tag{R.161}$$

3. An alternative equivalent way to perform a level- α test of non-inferiority is to compute an exact $100 \times (1 - \alpha)$ lower confidence bound for ρ , say (ρ_L, ∞) , using the method described in Section R.5.4. If $\rho_L < \rho_0$ we reject the null hypothesis of inferiority.

R.5.7 Unconditional Exact Confidence Interval for the Ratio of Proportions

Suppose π_c is the binomial response rate of Control and π_t is the binomial response rate of treatment. We wish to compute an exact $100(1 - \alpha)\%$ confidence interval for

$$\rho = \frac{\pi_t}{\pi_c}. \tag{R.162}$$

The procedure parallels that described in Section R.5.4 for the difference of proportions, δ . We use a test based procedure. That is, we invert hypothesis tests of the

R Technical Reference and Formulas: Analysis

form $\rho = \rho_0$, where, in general, $\rho_0 \neq 1$. There is one further complication, however, since the p-values which we compute under these alternative hypotheses depend on a nuisance parameter. We handle this problem the same way we handled it for Barnard's unconditional exact hypothesis test and for the various exact tests of non-inferiority; by taking a supremum over all possible values of the nuisance parameter. **Choice of Test Statistic for Test Based Interval Estimation**

Suppose we take n_c independent Bernoulli samples from Control and n_t independent Bernoulli samples from Treatment. Let $\mathbf{Y} \in \Omega$ (see Table R.3 denote any generic 2×2 table that might be observed, and let \mathbf{x} (see Table R.1, page 2568), be the 2×2 table that was actually observed. Define

$$\hat{\pi}_j = \frac{y_{1j}}{n_j} \tag{R.163}$$

for $j = c, t$. In **East** we provide a test based exact confidence interval using the standardized statistic.

$$D(\mathbf{Y}) = \frac{\hat{\pi}_t - \rho_0 \hat{\pi}_c}{\sqrt{\frac{(\hat{\pi}_t)(1-\hat{\pi}_t)}{n_t} + \frac{\rho_0^2(\hat{\pi}_c)(1-\hat{\pi}_c)}{n_c}}} \tag{R.164}$$

where $\hat{\pi}_c$ and $\hat{\pi}_t$ are the maximum likelihood estimates of π_c and π_t computed, under the restriction that $\hat{\pi}_t/\hat{\pi}_c = \rho_0$. The use of (R.164) as the test statistic has been proposed by Farrington and Manning (1990) for asymptotic confidence intervals and by Chan and Zhang (1999) for exact confidence intervals.

Test Based Exact Confidence Intervals: Inverting Two One-Sided Tests

Let (ρ_*, ρ^*) be the desired $100(1 - \alpha)\%$ exact confidence interval, evaluated at $D(\mathbf{x})$, the observed value of the test statistic. This exact confidence interval may be constructed by inverting two one-sided hypothesis tests, each at the $\alpha/2$ significance level, under appropriate alternative hypotheses about ρ . The probability of observing any $\mathbf{Y} \in \Omega$ for any given value of ρ is

$$f_{\pi_c, \rho}(\mathbf{Y}) = \binom{n_c}{y_{1c}} \binom{n_t}{y_{1t}} \pi_c^{y_{1c}} (1 - \pi_c)^{y_{2c}} (\rho \pi_c)^{y_{1t}} (1 - \rho \pi_c)^{y_{2t}} . \tag{R.165}$$

Define

$$P_{\pi_c, \rho}(D(\mathbf{x})) = \sum_{\mathbf{D}(\mathbf{Y}) \leq \mathbf{D}(\mathbf{x})} f_{\rho \pi_c}(\mathbf{Y}) \tag{R.166}$$

and

$$Q_{\pi_c, \rho}(D(\mathbf{x})) = \sum_{\mathbf{D}(\mathbf{Y}) \geq \mathbf{D}(\mathbf{x})} f_{\rho \pi_c}(\mathbf{Y}) . \tag{R.167}$$

We must eliminate the nuisance parameter π_c from equations (R.166) and (R.167) by taking the supremum over its range. It is easy to see that the permissible range for π_c given ρ is the interval

$$I(\rho) = \{\pi_c: 0 \leq \pi_c \leq \min(1/\rho, 1)\}. \quad (\text{R.168})$$

Thus we define

$$P_\rho(D(\mathbf{x})) = \sup\{\mathbf{P}_{\pi_c, \rho}(\mathbf{D}(\mathbf{x})): \pi_c \in I(\rho)\} \quad (\text{R.169})$$

and

$$Q_\rho(D(\mathbf{x})) = \sup\{\mathbf{Q}_{\pi_c, \rho}(\mathbf{D}(\mathbf{x})): \pi_c \in I(\rho)\}. \quad (\text{R.170})$$

Starting with $\rho = 0$, the desired lower confidence bound is obtained by increasing the value of ρ until we find a value, denoted by ρ_* , such that the equality

$$Q_{\rho_*}(D(\mathbf{x})) = \alpha/2 \quad (\text{R.171})$$

is satisfied but for any $\rho < \rho_*$, $Q_\rho(D(\mathbf{x})) < \alpha/2$. The upper confidence bound, ρ^* is obtained in an analogous fashion. Starting with $\rho = \infty$ (i.e., a very large positive number), the desired upper confidence bound is obtained by decreasing the value of ρ until we find a value, denoted by ρ^* , such that the equality

$$P_{\rho^*}(D(\mathbf{x})) = \alpha/2 \quad (\text{R.172})$$

is satisfied but for any $\rho > \rho^*$, $P_\rho(D(\mathbf{x})) < \alpha/2$.

East reports (ρ_*, ρ^*) as the $100 \times (1 - \alpha)\%$ confidence interval for the parameter ρ . Suppose ρ_0 is the true (unknown) value of ρ . The long run relative frequency with which, in repeated trials, this interval excludes ρ_0 is $\Pr(\rho_0 \leq \rho_*) + \Pr(\rho_0 \geq \rho^*)$. Using arguments similar to those given on page 2579 for the binomial difference, δ_0 , we can show that neither term in the above sum can exceed $\alpha/2$. Therefore the probability of the confidence interval excluding ρ_0 cannot exceed α . However, due to the discreteness of the distribution of $D(\mathbf{Y})$, and the conservatism induced by taking a supremum over all $\pi_c \in I(\rho)$, the above exclusion probability is usually less than α instead of equaling α . Thus, (ρ_*, ρ^*) may be regarded as a conservative confidence interval.

In addition to the exact confidence interval, East also reports an exact one-sided p-value, defined as the smaller of the two tail areas,

$$p_1 = \min(P_0(D(\mathbf{x})), \mathbf{Q}_0(\mathbf{D}(\mathbf{x}))), \quad (\text{R.173})$$

R Technical Reference and Formulas: Analysis

and the two sided exact p-value is twice the one-sided:

$$p_2 = 2p_1 . \tag{R.174}$$

The two-sided p-value is weakly consistent with the corresponding exact confidence interval for ρ . That is, if $1 \notin [\rho_*, \rho^*]$ then $p_2 < \alpha$. The stronger consistency requirement, that $p_2 < \alpha$ if and only if $1 \notin [\rho_*, \rho^*]$ cannot be established unless $P_\rho(D(\mathbf{x}))$ and $Q_\rho(D(\mathbf{x}))$ are monotone functions of ρ for any given $D(\mathbf{x})$. This need not be the case, however.

We shall see in Section R.5.8 that the above procedure can be slightly modified in practice. The suprema in equations (R.169) and (R.170) can be taken over a restricted range for π rather than over the entire range $I(\rho)$. This restriction, proposed by Berger and Boos (1994), adds stability and reduces the conservatism of the procedure. The right hand sides of equations (R.171) and (R.172) are suitably adjusted so that the restricted search for the supremum does not compromise the coverage properties of the resulting confidence interval.

Asymptotic Results

East provides asymptotic confidence interval and p-values for ρ . They are due to Farrington and Manning (1990). The standardized test statistic (R.164) is adopted and assumed to have a standard normal distribution. Under the null hypothesis that $\rho = 1$ the standardized test statistic is identical to the statistic used for Barnard's test. Therefore the asymptotic one-sided p-value is the same as the asymptotic one-sided p-value for Barnard's test. The asymptotic two-sided p-value is double the asymptotic one-sided p-value.

The asymptotic $100 \times (1 - \alpha)\%$ confidence interval $(\tilde{\rho}_*, \tilde{\rho}^*)$ is obtained by inverting the corresponding one-sided hypothesis tests. Thus $\tilde{\rho}_*$ satisfies the equality

$$1 - \Phi \left\{ \frac{(x_{1t}/n_t) - \rho_*(x_{1c}/n_c)}{\sqrt{\frac{\rho_*^2(\tilde{\pi}_c)(1-\tilde{\pi}_c)}{n_c} + \frac{(\tilde{\pi}_t)(1-\tilde{\pi}_t)}{n_t}}} \right\} = \alpha/2 , \tag{R.175}$$

where $\tilde{\pi}_c$ and $\tilde{\pi}_t$ are the maximum likelihood estimates of π_c and π_t , respectively, under the restriction that $\tilde{\pi}_t/\tilde{\pi}_c = \rho_*$. Similarly $\tilde{\rho}^*$ satisfies the equality

$$\Phi \left\{ \frac{(x_{1t}/n_t) - \rho^*(x_{1c}/n_c)}{\sqrt{\frac{\rho_*^2(\tilde{\pi}_c)(1-\tilde{\pi}_c)}{n_c} + \frac{(\tilde{\pi}_t)(1-\tilde{\pi}_t)}{n_t}}} \right\} = \alpha/2 , \tag{R.176}$$

where $\tilde{\pi}_c$ and $\tilde{\pi}_t$ are the maximum likelihood estimates of π_c and π_t , respectively, under the restriction that $\tilde{\pi}_t/\tilde{\pi}_c = \delta^*$.

R.5.8 Searching for Nuisance Parameters in a Restricted Range: Berger-Boos Correction

A source of conservatism, present in all the unconditional procedures covered in this chapter, is the fact that we must cater for the worst possible value of the nuisance parameter, π_c , by taking a supremum over its range. If this source of conservatism could be reduced in some way, it would result in shorter confidence intervals. A modification based on a proposal by Berger and Boos (1994) achieves this end. It should be noted that Berger and Boos (1994) actually proposed their method only for hypothesis tests. To our knowledge the extension to confidence intervals is new. To avoid unnecessary repetition, we will discuss the Berger-Boos modification only as it applies to Section (R.5.4), for computing an unconditional exact confidence interval for the difference of two binomial parameters based on inverting two one-sided hypothesis tests. It will be clear from this discussion that the same type of Berger-Boos correction also applies to all the other settings in this chapter, such as exact unconditional tests of superiority, non-inferiority or equivalence, and exact confidence intervals for ratios of binomials.

Let x be the observed 2×2 table. The main idea is that the information available in x about π_c and π_t can be used to reduce the conservatism of the exact confidence interval for δ . As a first step, we compute an exact $100(1 - \gamma/2)\%$ confidence interval, $A_1(x) = [l_1(x), u_1(x)]$, for π_c , and, independently, an exact $100(1 - \gamma/2)\%$ confidence interval, $A_2(x) = [l_2(x), u_2(x)]$, for π_t . Let \mathcal{E} denote the event $(\pi_c, \pi_t) \in A_1(x) \times A_2(x)$. If \mathcal{E} is true that restricts the range of δ and the associated range of π_c . It is easy to show that if \mathcal{E} is true, the range of possible values for δ must be restricted to the interval $[\delta_{\min}, \delta_{\max}]$, where

$$\delta_{\min} = l_2(x) - u_1(x), \tag{R.177}$$

and

$$\delta_{\max} = u_2(x) - l_1(x). \tag{R.178}$$

For any δ in this interval, π_c must lie in the restricted interval

$$I_x(\delta) = \{\pi_c: \max(l_1(x), l_2(x) - \delta) \leq \pi_c \leq \min(u_1(x), u_2(x) - \delta)\}. \tag{R.179}$$

Clearly $I_x(\delta) \subseteq I(\delta)$.

Define the right tail probability

$$Q_{\pi_c, \delta}(D(\mathbf{x})) = \sum_{\mathbf{D}(\mathbf{Y}) \geq \mathbf{D}(\mathbf{x})} \mathbf{f}_{\pi_c, \delta}(\mathbf{Y}) \tag{R.180}$$

R Technical Reference and Formulas: Analysis

and its supremum

$$Q_{\delta|\mathcal{E}}(D(\mathbf{x})) = \sup\{Q_{\pi_c, \delta}(D(\mathbf{x})): \pi_c \in I_x(\delta)\} . \quad (\text{R.181})$$

Notice the difference between $Q_{\delta}(D(x))$ given by equation (R.122) and $Q_{\delta|\mathcal{E}}(D(x))$ given by above equation. The first expression eliminates π_c by searching over the unrestricted range $I(\delta)$ while the second expression eliminates π_c by searching over the restricted range $I_x(\delta)$. The restricted search reduces conservatism since we must have

$$Q_{\delta|\mathcal{E}}(D(x)) \leq Q_{\delta}(D(x)) . \quad (\text{R.182})$$

In a similar manner we define the left tail probability

$$P_{\pi_c, \delta}(D(x)) = \sum_{D(y) \leq D(x)} f_{\pi_c, \delta}(y) \quad (\text{R.183})$$

and its supremum

$$P_{\delta|\mathcal{E}}(D(x)) = \sup\{P_{\pi_c, \delta}(D(x)): \pi_c \in I_x(\delta)\} . \quad (\text{R.184})$$

We next compute upper and lower confidence bounds for δ as described in Section R.5.4. Equations (R.123) and (R.124) must be modified, however, to compensate for the fact that we are now searching over a subset of the original parameter space. This adjustment is made by decreasing the right hand side of each equation by $\gamma/2$. Thus the lower confidence bound is the value of δ_* that satisfies the condition

$$Q_{\delta_*|\mathcal{E}}(D(x)) = \alpha/2 - \gamma/2 , \quad (\text{R.185})$$

such that for any δ satisfying $\delta_{\min} \leq \delta < \delta_*$, $Q_{\delta|\mathcal{E}}(D(x)) < \alpha/2 - \gamma/2$. If no value of δ_* can be found in the interval $[\delta_{\min}, \delta_{\max}]$ such that equation (R.185) is satisfied, we set $\delta_* = \delta_{\min}$. The upper confidence bound is the value of δ^* that satisfies the condition

$$P_{\delta^*|\mathcal{E}}(D(x)) = \alpha/2 - \gamma/2 , \quad (\text{R.186})$$

such that for any δ satisfying $\delta_{\max} \geq \delta > \delta^*$, $P_{\delta|\mathcal{E}}(D(x)) < \alpha/2 - \gamma/2$. If no value of δ^* can be found in the interval $[\delta_{\min}, \delta_{\max}]$ such that equation (R.186) is satisfied, we set $\delta^* = \delta_{\max}$. Thus, no matter what the data, we will always have

$$(\delta_*, \delta^*) \subseteq (\delta_{\min}, \delta_{\max}) . \quad (\text{R.187})$$

Suppose that δ_0 is the true (unknown) value of δ . With the above adjustment to the right hand sides of equations (R.185) and (R.186) one can show that

$$\Pr\{\delta_0 \notin (\delta_*, \delta^*)\} \leq \alpha , \quad (\text{R.188})$$

the desired exclusion probability. To see this observe that

$$\begin{aligned} \Pr\{\delta_0 \notin (\delta_*, \delta^*)\} &= \Pr\{[\delta \notin (\delta_*, \delta^*)] \cap \mathcal{E}\} + \Pr\{[\delta_0 \notin (\delta_*, \delta^*)] \cap \mathcal{E}^c\} \\ &\leq \Pr\{\delta_0 \notin (\delta_*, \delta^*)\} + \Pr(\mathcal{E}^c) \end{aligned} \tag{R.189}$$

$$\begin{aligned} &= \Pr(\delta_* \geq \delta_0) + \Pr(\delta^* \leq \delta_0) + \Pr(\mathcal{E}^c) \\ &\leq \Pr(\delta_* \geq \delta_0) + \Pr(\delta^* \leq \delta_0) + \gamma. \end{aligned} \tag{R.190}$$

Inequality (R.189) uses the fact that a probability cannot exceed 1. Since, for $i = 1, 2$, the interval $A_i(x)$ excludes the parameter π_c with probability $\gamma/2$, it follows by the Bonferroni inequality that $\Pr(\mathcal{E}^c) \leq \gamma$. Inequality (R.190) follows. We show next that neither $\Pr(\delta_* \geq \delta_0)$ nor $\Pr(\delta^* \leq \delta_0)$ can exceed $\alpha/2 - \gamma/2$. Define $H_{\mathcal{E}}(\delta_0)$ to be the smallest value of the random variable $D(\mathbf{Y})$ satisfying the inequality

$$Q_{\delta_0|\mathcal{E}}(H_{\mathcal{E}}(\delta_0)) \leq \alpha/2 - \gamma/2. \tag{R.191}$$

This definition implies that if $D(x) < H_{\mathcal{E}}(\delta_0)$, then $Q_{\delta_0|\mathcal{E}}(D(x)) > \alpha/2 - \gamma/2$. But we know from (R.185) that there is no value of $\delta \leq \delta_*$ for which $Q_{\delta}(D(x)) > \alpha/2 - \gamma/2$. It follows that $\delta_0 > \delta_*$ whenever $D(x) < H_{\mathcal{E}}(\delta_0)$. Therefore the random event $\{D(y) < H_{\mathcal{E}}(\delta_0)\}$ is a proper subset of the random event $\{\delta_0 > \delta_*\}$ and hence

$$\Pr(\delta_0 > \delta_*) \geq \Pr\{D(y) < H_{\mathcal{E}}(\delta_0)\}. \tag{R.192}$$

Taking complimentary probabilities on both sides of equation (R.192) we have

$$\Pr(\delta_* \geq \delta_0) \leq \Pr\{D(y) \geq H_{\mathcal{E}}(\delta_0)\} = Q_{\pi_c, \delta_0}(H_{\mathcal{E}}(\delta_0)). \tag{R.193}$$

Next, taking the supremum over all possible values of $\pi_c \in I_x(\delta_0)$, we have

$$Q_{\pi_c, \delta_0}(H_{\mathcal{E}}(\delta_0)) \leq \sup\{Q_{\pi_c, \delta_0}(H_{\mathcal{E}}(\delta_0)): \pi_c \in I_x(\delta_0)\} = Q_{\delta_0|\mathcal{E}}(H_{\mathcal{E}}(\delta_0)) \leq \alpha/2 - \gamma/2. \tag{R.194}$$

This establishes that

$$\Pr(\delta_* \geq \delta_0) \leq \alpha/2 - \gamma/2. \tag{R.195}$$

By a similar argument, $\Pr(\delta^* \leq \delta_0) \leq \alpha/2 - \gamma/2$. Therefore

$$\Pr\{\delta_0 \notin (\delta_*, \delta^*)\} \leq 2(\alpha/2 - \gamma/2) + \gamma = \alpha \tag{R.196}$$

The above modifications generally give shorter confidence intervals than the unmodified approach wherein we search the entire sample space of the nuisance parameters. One ambiguity about the procedure, however, is the choice of γ . The

R Technical Reference and Formulas: Analysis

smaller we make γ the more the modified method resembles the original method. The choice $\gamma = 0$ corresponds to making no modification to the original approach. At the other extreme the larger we make γ , the more we restrict the region in which we search for the supremum, and the more we must compensate for this restriction on the right hand sides of equations (R.185), (R.186). These equations show that we cannot increase γ beyond $\alpha/2$.

The Implementation in East

In **East** we have set the default value to $\gamma = 0.99e - 7$. This value is very small relative to α , which is usually 0.05. Therefore it does not usually affect the right hand sides of equations (R.185) and (R.186) by much. On the other hand it can provide greater stability, narrower confidence intervals, and faster execution times, in unbalanced settings, by cutting off regions near the extremes, 0 and 1, of the parameter space. We have observed, empirically, that the functions P_{δ, π_c} and Q_{δ, π_c} can have multiple high peaks at values of π_c near 0 or 1. By cutting off these regions from the parameter space we are able to reduce conservatism and add stability to the computation of the supremum.

R.5.9 Noninferiority:Odds Ratio of Proportions

The odds ratio of proportions denoted by Ψ is defined as $\Psi = \frac{\pi_t(1 - \pi_c)}{\pi_c(1 - \pi_t)}$. In Noninferiority trial, we are interested in testing $H_0 : \Psi \geq \Psi_0$ against $H_0 : \Psi < \Psi_0$ if $\Psi_0 > 1$

Or, $H_0 : \Psi \leq \Psi_0$ against $H_0 : \Psi > \Psi_0$ if $\Psi_0 < 1$

The test statistics for the two tests are :

Noninferiority (Wald)

$$Z = \frac{\ln \hat{\Psi} - \ln \Psi_0}{\sqrt{\frac{\hat{\pi}_t}{n_t(1 - \hat{\pi}_t)} + \frac{\hat{\pi}_c}{n_c(1 - \hat{\pi}_c)}}} \sim N(0, 1) \tag{R.197}$$

Noninferiority (Score)

$$Z = \frac{n_c(\hat{\pi}_c - \tilde{\pi}_c)}{SE} \sim N(0, 1) \tag{R.198}$$

where

$$SE = \left[\frac{1}{n_t \pi_t (1 - \pi_t)} + \frac{1}{n_c \pi_c (1 - \pi_c)} \right]^{-1} \tag{R.199}$$

R.5.10 Common Odds Ratio for Stratified 2x2 Tables

Breslow-Day Test for Homogeneity of Odds-Ratios

$$H_0: \Psi_i = \Psi, i = 1, 2, \dots s .$$

Breslow and Day (1980) statistic:

$$\chi_{BD}^2 = \sum_{i=1}^s \frac{[X_i - A_i(\hat{\Psi})]^2}{var(X_i | \hat{\Psi})} , \tag{R.200}$$

where $A_i(\hat{\Psi})$ is the positive root of the quadratic equation

$$\frac{A_i(N_i - m_i - n_i + A_i)}{(m_i - A_i)(n_i - A_i)} = \hat{\Psi} , \tag{R.201}$$

formed by expressing the i^{th} table as

$$\begin{matrix} m_i - A_i & A_i \\ N_i - m_i - n_i + A_i & n_i - A_i , \end{matrix}$$

and equating its empirical Odds-Ratio to the Mantel-Haenszel common Odds-Ratio

$$\hat{\Psi} = \frac{\sum_{i=1}^s x_i(N_i - m_i - n_i + x_i)/N_i}{\sum_{i=1}^s (n_i - x_i)(m_i - x_i)/N_i} . \tag{R.202}$$

The variance of X_i is estimated by:

$$var(X_i | \hat{\Psi}) = \left[\frac{1}{A_i(\hat{\Psi})} + \frac{1}{m_i - A_i(\hat{\Psi})} + \frac{1}{n_i - A_i(\hat{\Psi})} + \frac{1}{N_i - m_i - n_i + A_i(\hat{\Psi})} \right]^{-1} \tag{R.203}$$

Tarone correction for Breslow-Day test

$$\chi_{BDT}^2 = \sum_{i=1}^s \frac{[X_i - A_i(\hat{\Psi})]^2}{var(X_i | \hat{\Psi})} - \frac{\left[\sum_{i=1}^s X_i - \sum_{i=1}^s A_i(\hat{\Psi}) \right]^2}{\sum_{i=1}^s var(X_i | \hat{\Psi})} , \tag{R.204}$$

R Technical Reference and Formulas: Analysis

where A_i , and $var(X_i | \hat{\Psi})$ are defined as above. In large samples, both χ_{BD}^2 and χ_{BDT}^2 are chi-squared distributed with $s - 1$ degrees of freedom, and the 2-sided p-values for testing H_0 is:

$$p_{BD} = \Pr(\chi_{BD}^2 \geq \chi_{0,BD}^2) \tag{R.205}$$

$$p_{BDT} = \Pr(\chi_{BDT}^2 \geq \chi_{0,BDT}^2), \tag{R.206}$$

where $\chi_{0,BD}^2$ and $\chi_{0,BDT}^2$ are the observed values of χ_{BD}^2 and χ_{BDT}^2 .

Mantel-Haenszel Inference for the Common Odds-Ratio

$$H_0: \Psi = 1.$$

Mantel-Haenszel (1959) test

$$\chi_{MH}^2 = \left[\sum_{i=1}^s \frac{x_i y'_i - x'_i y_i}{N_i} \right]^2 / \sum_{i=1}^s \frac{m_i m'_i n_i (N_i - n_i)}{(N_i - 1) N_i^2} \tag{R.207}$$

is chi-squared distributed with one degree of freedom.

$$p_{MH} = \Pr(\chi_{MH}^2 \geq \chi_0^2) \tag{R.208}$$

where χ_0^2 is the observed value of χ_{MH}^2 . The RBG variance is

$$var(\log \hat{\Psi}) = \sum_{i=1}^s \left(\frac{a_i c_i}{2c_+^2} + \frac{a_i d_i + b_i c_i}{2c_+ d_+} + \frac{b_i d_i}{2d_+^2} \right) \tag{R.209}$$

where

$$\begin{aligned} a_i &= \frac{x_i + y'_i}{N_i}, \\ b_i &= \frac{x'_i + y_i}{N_i}, \\ c_i &= \frac{x_i y'_i}{N_i}, \\ d_i &= \frac{x'_i y_i}{N_i}, \\ c_+ &= \sum_{i=1}^s c_i, \\ d_+ &= \sum_{i=1}^s d_i. \end{aligned}$$

A $100(1 - \alpha)\%$ confidence interval for $\log \Psi$

$$CI_{RBG} = \log \hat{\Psi} \pm z_{\alpha/2} [\text{var}(\log \hat{\Psi})]^{1/2}. \quad (\text{R.210})$$

The 2-sided p-value for testing

$$H_0: \Psi = 1,$$

based on the RBG variance is

$$p_{RBG} = 2[1 - \Phi(\frac{|\log \hat{\Psi}|}{\sqrt{\text{var}(\log \hat{\Psi})}})]. \quad (\text{R.211})$$

R.5.11 Fisher's Exact Test

As in the **Difference of Proportions** test, suppose π_t and π_c denote the proportions of the successes from the experimental treatment (T) and the control treatment (C). To test the null hypothesis:

$$H_0: \pi_t = \pi_c, \quad (\text{R.212})$$

against 1-sided alternatives of the form,

$$H_1: \pi_t > \pi_c, \quad (\text{R.213})$$

or

$$H_1': \pi_t < \pi_c, \quad (\text{R.214})$$

and against 2-sided alternatives of the form

$$H_2: \pi_t \neq \pi_c. \quad (\text{R.215})$$

Suppose that H_0 is true and let the common probability of success for the two binomial populations be $\pi_t = \pi_c = \pi$. Then the probability of observing the data in Table R.1 is a product of two binomial probabilities, denoted by

$$f_0(\mathbf{x}) = \binom{n_c}{x_{1c}} \binom{n_t}{x_{1t}} \pi^{x_{1c} + x_{1t}} (1 - \pi)^{x_{2c} + x_{2t}}. \quad (\text{R.216})$$

The p-value is defined to be the probability, under H_0 , of obtaining a 2×2 table at least as extreme as the observed table, \mathbf{x} . Let \mathbf{Y} denote any generic 2×2 table that can arise if we take two independent samples, one of size n_c from binomial population c and the other of size n_t from binomial population t. Such a generic 2×2 table is

R Technical Reference and Formulas: Analysis

Table R.3: Any Generic 2x2 Contingency Table, \mathbf{y}

| Response | Control | Treatment | Row_Total |
|-----------|----------|-----------|-------------------|
| Success | y_{1c} | y_{1t} | $y_{1c} + y_{1t}$ |
| Failure | y_{2c} | y_{2t} | $y_{2c} + y_{2t}$ |
| Col.Total | n_c | n_t | N |

displayed below in Table R.3: The probability of observing this table is $f_0(\mathbf{Y})$ which, as shown by equation (R.216), contains an unknown (nuisance) parameter, π . As long as the probability of observing any generic 2×2 table depends on π , exact inference is not possible. Since the p-value is based on summing up the probabilities of many such tables, each depending on an unknown parameter. The key to exact inference is getting rid of π , the nuisance parameter.

In Fisher's Exact Test, conditional approach is used. The sufficient statistic for π is $y_{11} + y_{12}$, the sum of successes from the two populations. The observed value of the sufficient statistic is m_1 . Thus, by the sufficiency principle, if the condition on $y_{11} + y_{12} = m_1$, the probability of any generic 2×2 table, \mathbf{Y} , no longer depends on the nuisance parameter π . To see this let

$$\Gamma = \left\{ \mathbf{Y} : \sum_{j=1}^2 y_{ij} = \mathbf{m}_i, \sum_{i=1}^2 y_{ij} = \mathbf{n}_j \right\} \tag{R.217}$$

denote a reference set of 2×2 contingency tables with fixed row and column margins. Since we are dealing with the case of two independent binomial samples, each of size $n_i, i = 1, 2$, (considering $n_1 = n_c$ and $n_2 = n_t$), conditioning on $y_{11} + y_{12} = m_1$ is equivalent to conditioning on $\mathbf{Y} \in \Gamma$. Let $h(\mathbf{Y})$ denote the probability of observing any $\mathbf{Y} \in \Gamma$ under the null hypothesis (R.212). Then

$$h(\mathbf{Y}) = \frac{f_0(\mathbf{Y})}{\sum_{\mathbf{Y} \in \Gamma} f_0(\mathbf{Y})} \tag{R.218}$$

which simplifies to

$$h(\mathbf{Y}) = \frac{\binom{n_1}{y_{11}} \binom{n_2}{y_{12}}}{\binom{N}{m_1}}, \tag{R.219}$$

a hypergeometric probability free of π . Exact inference is thus possible only if we confine our attention to 2×2 tables in Γ . Next turn to the question of how to

determine if a 2×2 contingency table, \mathbf{Y} , is at least as extreme as the observed table, \mathbf{x} . Let $D : \Gamma \rightarrow \mathcal{R}$ be a function assigning a real number, $D(\mathbf{Y})$, to each $\mathbf{Y} \in \Gamma$ in such a way that \mathbf{Y} is judged to be at least as extreme as \mathbf{x} provided $D(\mathbf{Y}) \geq D(\mathbf{x})$. We refer to $D(\mathbf{Y})$ as a discrepancy measure. Fisher's test statistic is given by:

$$D(\mathbf{Y}) = -2 \log(\gamma \mathbf{h}(\mathbf{Y})) . \tag{R.220}$$

where

$$\gamma = (2\pi N^{-3} m_1 m_2 n_1 n_2)^{-\frac{1}{2}} \tag{R.221}$$

The exact 2-sided p-value is defined as:

$$p_2 = \Pr(D(\mathbf{Y}) \geq D(\mathbf{x})) = \sum_{D(\mathbf{Y}) \geq D(\mathbf{x})} \mathbf{h}(\mathbf{Y}) , \tag{R.222}$$

the sum being taken over all $\mathbf{Y} \in \Gamma$ such that $D(\mathbf{Y}) \geq D(\mathbf{x})$. In large samples the distribution of $D(\mathbf{Y})$ conditional on $\mathbf{Y} \in \Gamma$ converges to the chi-square distribution with 1 degree of freedom. (Kendall and Stuart (1979)). The asymptotic 2-sided p-value is given by:

$$\tilde{p}_2 = \Pr(\chi_1^2 \geq D(\mathbf{x})) , \tag{R.223}$$

where χ_1^2 is a random variable distributed as chi-square with 1 df. You can also define the 1-sided exact p-value. It is based on the test statistic:

$$D(\mathbf{Y}) = y_{11} . \tag{R.224}$$

Since we have confined our attention only to 2×2 contingency tables in Γ , the value of y_{11} suffices to specify the entire 2×2 table \mathbf{Y} , and the exact probability of y_{11} is $h(\mathbf{Y})$. Moreover it is easy to see that y_{11} ranges from a minimum of

$$t_{\min} = \max(0, n_1 - m_2) , \tag{R.225}$$

to a maximum of

$$t_{\max} = \min(m_1, n_1) . \tag{R.226}$$

The exact 1-sided p-value for the Fisher tests is then defined as either the right or left tail area of the exact distribution of y_{11} at the observed value, x_{11} , based on the location of x_{11} relative to $n_1 m_1 / N$, the mean of y_{11} . That is,

$$p_1 = \begin{cases} \sum_{y_{11}=t_{\min}}^{x_{11}} h(\mathbf{Y}) & \text{if } x_{11} > n_1 m_1 / N \\ \sum_{y_{11}=x_{11}}^{t_{\max}} h(\mathbf{Y}) & \text{if } x_{11} \leq n_1 m_1 / N \end{cases} \tag{R.227}$$

A small 1-sided p-value furnishes evidence against the 1-sided alternative (R.213) if it is computed as the right tail of the exact distribution of y_{11} , and against the 1-sided alternative (R.214), if it is computed as the left tail of the exact distribution of y_{11} .

R Technical Reference and Formulas: Analysis

R.6 Many Proportions

- R.6.1 Contingency Coefficients
- R.6.2 Wilcoxon Rank Sum Test for Ordered Categories Data
- R.6.3 Trend in R ordered proportions
- R.6.4 Chi-square for R Unordered Binomial Properties
- R.6.5 Chi-square for R Unordered multinomial Properties

R.6.1 Contingency Coefficients

The **Contingency Coefficients** are derived from Pearson's chi-square statistic. The Phi contingency coefficient is given by,

$$\phi = \sqrt{\frac{\chi^2(\mathbf{x})}{N}} . \tag{R.228}$$

Pearson's contingency coefficient is given by:

$$CC = \sqrt{\frac{\chi^2(\mathbf{x})}{\chi^2(\mathbf{x}) + N}} . \tag{R.229}$$

The Sakoda contingency coefficient is given by

$$CC_1 = \sqrt{\frac{q\chi^2(\mathbf{x})}{(q-1)(\chi^2(\mathbf{x}) + N)}} . \tag{R.230}$$

The Tschuprov contingency coefficient ranges between 0 and 1, with 0 signifying no association and 1 signifying perfect association. It is given by,

$$CC_2 = \left(\frac{\chi^2(\mathbf{x})}{N\sqrt{(r-1)(c-1)}} \right)^{1/2} . \tag{R.231}$$

Finally, Cramer's V coefficient ranges between 0 and 1, with 0 signifying no association and 1 signifying perfect association. It is given by,

$$V = \sqrt{\frac{\chi^2(\mathbf{x})}{N(q-1)}} . \tag{R.232}$$

The $100 \times (1 - \alpha)\%$ confidence interval for any measure of association

$$CI = M(\mathbf{x}) \pm z_{\alpha/2} \times ASE.MLE , \tag{R.233}$$

where z_{β} is the value of the $(1 - \beta)$ percentile point of the standard normal distribution.

R.6.2 Wilcoxon Rank Sum Test for Ordered Categories Data

Each response must fall into one of c ordinal categories according to a multinomial distribution.

$$\gamma_{jk} = \sum_{i=1}^j \pi_{ik} ,$$

$$\gamma'_{jk} = \sum_{i=1}^j \pi'_{ik} ,$$

for $j = 1, 2, \dots, c$, and $k = 1, 2, \dots, s$. Then the Wilcoxon test is especially suited to detecting departures from the null of the form

$$H_1: \gamma_{jk} \geq \gamma'_{jk} ,$$

or

$$H'_1: \gamma'_{jk} \geq \gamma_{jk} ,$$

for all $j \in \{1, 2, \dots, c\}$, $k \in \{1, 2, \dots, s\}$, with strict inequality at at-least 6.3 one j, k . The 2-sided alternative hypothesis is that either H_1 or H'_1 is true; the alternative hypothesis does not specify which of the two possibilities is true, however. The Wilcoxon Rank Sum Test Statistic is of the form,

$$T = \sum_{k=1}^s \sum_{j=1}^c w_j X_{jk} , \tag{R.234}$$

where w_j are Wilcoxon-Mann-Whitney scores which are the ranks (mid-ranks in the case of tied observations) of the underlying responses.

$$w_j = n_1 + \dots + n_{j-1} + (n_j + 1)/2 \tag{R.235}$$

The mean of T , under the null hypothesis of no row and column interaction is given by

$$E(T) = \left(\frac{m}{N}\right) \sum_{j=1}^c w_j n_j \tag{R.236}$$

And the variance is

$$\sigma^2(T) = \left[\frac{mm'}{N(N-1)}\right] \sum_{j=1}^c \left[w_j - \frac{E(T)}{m}\right]^2 n_j \tag{R.237}$$

Under H_0 , T follows asymptotic Normal distribution with mean $E(T)$ and variance $\sigma^2(T)$

R Technical Reference and Formulas: Analysis

R.6.3 Trend in R ordered proportions

To determine whether a trend exists in the unknown proportions of response π_g for $g = 1, \dots, K$ ordered binomially distributed populations using independent random samples. Test statistic:

$$T = \sum_{j=1}^c w_j Y_j, \quad (\text{R.238})$$

where

$$w_j = j - 1. \quad (\text{R.239})$$

The mean of the test statistic is

$$E(T) = \left(\frac{m}{N}\right) \sum_{j=1}^c w_j n_j. \quad (\text{R.240})$$

and the variance of the test statistic is

$$\sigma^2(T) = \left[\frac{m(N-m)}{N(N-1)}\right] \sum_{j=1}^c \left[w_j - \frac{E(T)}{m}\right]^2 n_j. \quad (\text{R.241})$$

Under H_0 , $Z = \frac{T-E(T)}{\sqrt{Var(T)}}$ follows $N(0, 1)$ distribution.

R.6.4 Chi-square for R Unordered Binomial Properties

Hypothesis $H_0 : \pi_{1j} = \pi_{2j} \dots = \pi_{Rj}$ for all $j = 1, 2$

Vs H_1 : at least one π_{ij} differs for $i = 1, 2, \dots, R$ and $j = 1, 2$.

Let the $R \times 2$ contingency Table R.4: displayed in Table R.4 be the one actually observed.

Test Statistic

$$\chi_{R-1}^2 = \sum_{i=1}^R \sum_{j=1}^2 \frac{(x_{ij} - m_i n_j / N)^2}{m_i n_j / N}$$

R.6.5 Chi-square for R Unordered multinomial Properties

Hypothesis $H_0 : \pi_{12} = \pi_{2j} \dots = \pi_{Rj}$ for all $j = 1, 2, \dots, C$

Table R.4: The Observed Rx2 Contingency Table

| Rows | Failure | Success | Row Total |
|------------|----------|----------|-----------|
| Row 1 | x_{11} | x_{12} | m_1 |
| Row 2 | x_{21} | x_{22} | m_2 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| Row R | x_{R1} | x_{R2} | m_R |
| Col. Total | n_1 | n_2 | N |

Vs

H_1 : at least one π_{ij} differs.

Let the $R \times C$ contingency Table R.5 displayed in Table R.4 be the one actually observed.

Table R.5: The Observed RxC Contingency Table

| Rows | Col.1 | Col.2 | Col.3 | Col. C | Row Total |
|------------|----------|----------|-------|----------|-----------|
| Row 1 | x_{11} | x_{12} | ... | x_{1C} | m_1 |
| Row 2 | x_{21} | x_{22} | ... | x_{2C} | m_2 |
| ⋮ | ⋮ | ⋮ | | ⋮ | ⋮ |
| Row R | x_{R1} | x_{R2} | ... | x_{RC} | m_R |
| Col. Total | n_1 | n_2 | ... | n_C | N |

Test Statistic

$$\chi^2_{R-1} = \sum_{i=1}^R \sum_{j=1}^2 \frac{(x_{ij} - m_i n_j / N)^2}{m_i n_j} / N$$

R.7 Agreement

R.7.1 Cohen's Kappa

R.7.1 Cohen's Kappa

Hypothesis

R Technical Reference and Formulas: Analysis

H_0 : Agreement between two refers is purely from random variation

Vs

H_1 : Agreement between two refers is not purely from random variation (for two sided test)

Either

H_1 : Agreement between two refers is greater than that is expected from radom variation only.

Or

H_1 : Agreement between two refers is less than that is expected from random variation only.

(For 1-sided test)

Test Statistic

For Kappa

$$K = \frac{n \sum_{i=1}^r x_{ii} - \sum_{i=1}^r m_i n_i}{n^2 - \sum_{i=1}^r m_i n_i} \sim N(0, 1).$$

For Weighted Kappa

$$K_w = \frac{\sum_{i=1}^r \sum_{j=1}^r w_{ij} x_{ij} - \sum_{i=1}^r \sum_{j=1}^r w_{ij} m_i n_j}{n^2 - \sum_{i=1}^r \sum_{j=1}^r w_{ij} m_i n_j} \sim N(0, 1)$$

R.8 Survival : Two Samples

Let

$t_i, I = 1, 2, 3, \dots, M$ be the distinct time points of event on any Arm

$d_{i,t}$ = Number of events on treatment arm at time t_i

$n_{i,t}$ = Number of subjects at risk on treatment arm just before time t_i

$d_{i,c}$ = Number of events on control arm at time t_i

$n_{i,c}$ = Number of subjects at risk on control arm just before time t_i

$$d_i = d_{i,t} + d_{i,c}$$

$$n_i = n_{i,t} + n_{i,c}$$

Assumption : Censored observations are considered in the risk set if they are tied with time point at which event of treatment is observed.

For Superiority

$$Num_i = d_{i,t} - n_{i,t} \frac{d_i}{n_i}$$

$$Den_i = \frac{n_{i,t}n_{i,c}(n_i - d_i)d_i}{n_i^2(n_i - 1)}$$

For Non-inferiority

δ_0 = Non-Inferiority margin

$$n_i^* = n_{i,t} + n_{i,c}^* e^{-\delta_0}$$

$$Num_i = d_{i,t} - n_{i,t} \frac{d_i}{(n_i^*)}$$

$$Den_i = \frac{n_{i,t}n_{i,c}d_i e^{-\delta_0}}{n_i^* n_i^*}$$

Weighted Test Statistic is defined as (for both superiority and non-inferiority)

$$Num = \sum_{i=1}^M W_i Num_i \tag{R.242}$$

$$Den = \sum_{i=1}^M W_i^2 Den_i \tag{R.243}$$

$$TS = \frac{Num}{\sqrt{Den}}$$

where weights are defined as follows.

R.8.1 Logrank Test

$W_i = 1$ For all i .

R Technical Reference and Formulas: Analysis

R.8.2 Wilcoxon-Gehan

$$W_i = n_i$$

R.8.3 Harrington-Fleming

$$W_i = (\hat{S}_{i-1})^p (1 - \hat{S}_{i-1})^q$$

Where

$$W_i = \begin{cases} 1 & \text{if } q = 0 \\ 0 & \text{if } q > 0 \end{cases}$$

For $i > 1$

$$W_i = (\hat{S}_{i-1})^p (1 - \hat{S}_{i-1})^q$$

With

$$\hat{S}_t = \prod_{t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right)$$

Test Statistic for Stratified Simulations

Let

S = Number of Strata

Num_j = Numerator for j^{th} stratum using (R.242)

Den_j = Denominator for j^{th} stratum using (R.243)

Test Statistic is given by

$$TS = \frac{\sum_{j=1}^S Num_j}{\sqrt{\sum_{j=1}^S Den_j}} \quad (R.244)$$

S Theory - Design - Binomial One-Sample Exact Test

This appendix lays out the theory behind East’s power and sample size computations in the case of the exact fixed sample test and the exact group sequential test of a proportion π being equal to a constant π_0 .

Both Schultz et al. (1973) and Fleming (1982) have proposed multi-stage procedures for rejecting the null hypothesis under strict assumptions about the type 1 and type 2 errors. The methods used in East and described below are based on Jennison and Turnbull (2000).

Section (S.1) explains how to calculate the power and the sample size of the exact fixed sample test. Section (S.2) continues by considering the power and sample size of the group sequential test. It also explains how the boundaries of this test are computed.

S.1 Power and Sample Size for the Exact Fixed Sample Test

S.1.1 Power

S.1.2 Sample Size

Consider a clinical trial of fixed sample size N . The goal is to test - based on the observed number of successes $S = s$ - whether the binary probability π of response is equal to some a priori hypothesized value π_0 .

The null hypothesis of interest is

$$H_0 : \pi = \pi_0$$

East computes the power of the exact test of H_0 against one of the following one-sided alternatives

$$H_1 : \pi = \pi_1 \quad \pi_1 > \pi_0$$

or

$$H_1 : \pi = \pi_1 \quad \pi_1 < \pi_0$$

In what follows, we will assume that interest resides in detecting the former alternative where $\pi_1 > \pi_0$.

The exact test is based on the binomial probability distribution of the response variable S . Recall that for a binomial distribution $Bin(N, \pi)$ the probability that $S = s$ is given by

$$\Pr(S = s|\pi) = \binom{N}{s} \pi^s (1 - \pi)^{(N-s)}$$

and that the tail end probability is the cumulative sum of probabilities

$$\Pr(S \geq s|\pi) = \sum_{i=s}^N \binom{N}{i} \pi^i (1 - \pi)^{(N-i)}.$$

S Theory - Design - Binomial One-Sample Exact Test

Suppose data from the trial provide an observed number of responses $S = s$. Then a test of the null hypothesis H_0 consists in calculating the probability under the null distribution $Bin(N, \pi_0)$ of observing s or more responses among N subjects, and then comparing this probability to the type 1 error rate α . If

$$\Pr(S \geq s | \pi_0) \leq \alpha$$

then the null hypothesis that $\pi = \pi_0$ can be rejected in favor of the alternative hypothesis that $\pi = \pi_1$.

S.1.1 Power of the Exact Fixed Sample Test

Since the power and type 1 error of a design are intimately related and because in an exact test the desired false positive rate is often not attainable, we first consider calculation of the design's type 1 error before moving on to the design's power.

Suppose a type 1 error probability of α has been specified for the study design. Due to the discreteness of the binomial distribution, this false positive rate α will more likely than not be unattainable. Instead the design will attain a type 1 error of $\alpha^* \leq \alpha$.

Under the null hypothesis H_0 , the number of responses S follows a $Bin(N, \pi_0)$. Define s_0 to be the smallest integer, such that

$$\Pr(S \geq s_0 | \pi_0) \leq \alpha$$

Then the attained significance level α^* is given by

$$\alpha^* = \Pr(S \geq s_0 | \pi_0)$$

Upon knowing the critical value s_0 that gives us type 1 error α^* under the null hypothesis distribution $Bin(N, \pi_0)$, we can calculate the exact power of the design by considering the probability distribution under the alternative hypothesis $Bin(N, \pi_1)$.

The exact power of the procedure is given by

$$(1 - \beta) = \Pr(S \geq s_0 | \pi_1)$$

S.1.2 Sample Size Calculation for the Exact Fixed Sample Design

Calculation of a sample size N for a pre-specified type 1 error α and power $(1 - \beta)$ is

complicated by the fact that neither α nor β are attainable given the discreteness of the binomial distribution.

It is well known that a plot of the power versus sample size of a design displays a saw-tooth behavior, but this zig-zag is mostly due to the concurrently varying type 1 error of the design. Due to this behavior though, multiple sample sizes may be provided in answer to the design problem. Really, however, the optimal sample size will depend on the priorities given to type 1 and 2 errors by the investigator.

The sample size N is calculated such that both attained type 1 and 2 errors α^* and β^* are controlled. A search of the parameter space of N must be performed to find those values satisfying both of the following equations

$$\alpha^* = \Pr(S \geq s_0 | N, \pi_0) \leq \alpha$$

and

$$\beta^* = \Pr(S \leq s_0 | N, \pi_1) \leq \beta$$

while either (1) primarily maximizing the attained type 1 error while maintaining it below α , (2) primarily maximizing the attained type 2 error while maintaining it below β , or (3) optimizing to get α^* and β^* as close to α and β , respectively as possible.

The most practical choice of sample size, however, may be that sample size above which power is guaranteed to be at least $(1 - \beta)$.

S.2 Power and Sample Size for the Exact Group Sequential Test

Instead of a fixed sample test of the null hypothesis, let us now consider a group sequential test. This procedure tests the null hypothesis not just once at the end of the trial, but a total of K times after each group of n_k $k = 1, \dots, K$ subjects' data have been observed. In what follows, we still consider a 1-sided test of the null hypothesis with the alternative hypothesis specified in the direction of $\pi_1 > \pi_0$.

Suppose an error-spending function has been pre-specified to control the type 1 error of the group sequential test. Let $\{\alpha_1, \dots, \alpha_K\}$ be the fractions of the type 1 error at each stage, such that they sum up to α . The efficacy boundary corresponding to this error-spending function is given by the set of critical values $\{c_1, \dots, c_K\}$.

Before considering the construction of the boundary itself and the calculation of the test's power, the probability distribution of the number of responses at stage k must be established.

S Theory - Design - Binomial One-Sample Exact Test

Define $C_k(s; \pi, N_k)$ to be the probability of observing s responses at stage k where $1 \leq k \leq K$. Here N_k refers to the cumulative sample size up to and including stage k so that $n_k = N_k - N_{k-1}$ is the sample size for stage k only. Then for the first stage, the probability distribution of response is binomial with

$$C_1(s; \pi, N_1) = \binom{N_1}{s} \pi^s (1 - \pi)^{(N_1 - s)}.$$

Thereafter, the probability of s responses at stage $k > 1$ depends on how many of those responses have been observed up to but excluding stage k . This distribution is given by

$$C_k(s; \pi, N_k) = \sum_{i=a_k(s)}^{b_k(s)} C_{k-1}(i; \pi, N_{k-1}) * B_k((s - i); \pi, n_k)$$

where

$$B_k(s; \pi, n_k) = \binom{N_k}{s} \pi^s (1 - \pi)^{(N_k - s)}$$

and

$$a_k(s) = \max(0, (s - n_k))$$

$$b_k(s) = \min(s, (c_{(k-1)} - 1))$$

S.2.1 Computing Boundaries for the Exact Group Sequential Test

Given an a priori specified α -spending function, the type 1 error fractions used at each stage are provided as $\alpha_1, \dots, \alpha_K$. However, due to the discreteness of the binomial distribution, those values are not achievable at each step. It is reasonable however to carry-over the unspent type 1 error at any stage to the subsequent stage. This informs the following calculations and adjustments to the boundary.

At the first interim look $k = 1$, operating under the null hypothesis, the boundary value is calculated simply by finding the smallest integer c_1 such that

$$\sum_{i=c_1}^{N_1} C_1(i; \pi_0, N_1) \leq \alpha_1$$

However, the actual tail end probability defined by the cut-off value c_1 is

$$\alpha_1^* = \sum_{i=c_1}^{N_1} C_1(i; \pi_0, N_1)$$

The unused type 1 error $\theta_1 = \alpha_1 - \alpha_1^*$ can be carried over to be spent at stage 2.

More generally, define $\theta_0 = 0$ then at stage k , the available type 1 error is $\alpha_k + \theta_{k-1}$ where

$$\begin{aligned}\theta_{k-1} &= (\alpha_{k-1} + \theta_{k-2}) - \alpha_{k-1}^* \\ &= \sum_{i=1}^{k-1} (\alpha_i - \alpha_i^*)\end{aligned}$$

The boundary value is then calculated by finding the smallest integer c_k for which

$$\sum_{i=c_k}^{N_k} C_k(i; \pi_0, N_k) \leq \alpha_k + \theta_{k-1}$$

Repeating this process until the ultimate look K enables the full construction of the efficacy boundary. Note that at the last look, the cumulative and thus overall attained type 1 error of the design will be $\alpha^* i \leq \alpha$ where

$$\begin{aligned}\alpha^* &= \sum_{i=1}^K \alpha_i^* \\ &= \alpha - \theta_K.\end{aligned}$$

S.2.2 Power of the Exact Group Sequential Test

Just as in the case of the exact fixed sample test, the power and type 1 error of the group sequential test are intimately tied. The previous section provided calculation of the boundary c_1, \dots, c_K under the assumption that the null hypothesis was true. These defined an attained overall type 1 error $\alpha^* \leq \alpha$ of the group sequential design.

Considering the crossing probability of that same boundary under the alternative hypothesis provides the power of the group sequential test. That is

$$(1 - \beta) = \sum_{k=1}^K \sum_{i=c_k}^{N_k} C_k(i; \pi_1, N_k)$$

S Theory - Design - Binomial One-Sample Exact Test

S.2.3 Sample Size Calculation for the Exact Group Sequential Design

As in the case of the exact fixed sample design, calculation of the maximum sample size $N_{\max} = N_K$ for a pre-specified type 1 error α and power $(1 - \beta)$ is complicated by the fact that neither α nor β are attainable given the discreteness of the binomial distribution.

As a result, the choice of N_{\max} is not unique. Rather the decision of which sample size to choose for a particular trial will depend on the priorities given to type 1 and 2 errors by the investigator.

The sample size N_{\max} is calculated such that both attained type 1 and 2 errors α^* and β^* are controlled. A search of the parameter space of $N_{\max} = N_K$ must be performed to find those values satisfying both of the following equations

$$\alpha^* = \sum_{k=1}^K \sum_{i=c_k}^{N_k} C_k(i; \pi_0, N_k) \leq \alpha$$

and

$$\beta^* = \sum_{k=1}^K \sum_{i=0}^{(c_k-1)} C_k(i; \pi_1, N_k) \leq \beta$$

while either (1) primarily maximizing the attained type 1 error while maintaining it below α , (2) primarily maximizing the attained type 2 error while maintaining it below β , or (3) optimizing to get α^* and β^* as close to α and β , respectively as possible.

The most practical choice of sample size, however, may be that sample size above which power is guaranteed to be at least $(1 - \beta)$.

T Theory - Design - Binomial Paired-Sample Exact Test

This appendix presents the theory behind the computations of power and sample size for the conditional exact McNemar's test for the difference of proportions arising from paired binomial populations. East implements the methodology and numerical algorithms for the conditional version of McNemar's test, published by Duffy (1984) and Agresti (2002). Methods and algorithms for the unconditional test, used in previous versions of East, have been published by Suissa and Shuster (1991).

Exact conditional methods are considerably faster to execute than the exact unconditional methods. In the paired binomial case, the conditional approach simplifies to a single binomial model, allowing the computation of exact p-values and confidence intervals for arbitrarily large data sets with little difficulty. This is not the case for unconditional methods, where fairly long computing times are to be expected for larger sample sizes. In addition, the theory of exact unconditional inference is more complex and historically has not possessed as extensive a bibliography as the theory of exact conditional inference. Section (T.1) presents how to calculate the power and the sample size for the exact fixed sample conditional McNemar's test.

T.1 Power and Sample Size for the Exact Conditional Fixed Sample Test: McNemar's Test

Consider a trial in which the investigator's interest is in testing for a difference in success rates between paired binary responses. Such a test is typically used in a repeated measures setting, for example when each subject's response is recorded both before and after treatment. The test then determines if the pre and post treatment response rates are equivalent. Another example would be a study involving matched pairs, such as siblings, where each member of the pair is measured for an outcome of interest and tests for the same probability of response. Here, the inference is complicated by the fact that the observations are correlated, even though there is independence across the different pairs being studied.

Suppose that two binomial responses are observed on either N individuals (pre and post event), or N matched pairs. Let y_{11} be the count of the number of individuals whose first and second responses are both positive, or in the case of matched pairs where both responses are positive. In a similar manner let y_{22} be the count where both first and second responses are negative. Let y_{12} be the count of pairs where the first response is positive and second response is negative and let y_{21} be the count where the first response is negative and second response is positive. McNemar's test is based on

T Theory - Design - Binomial Paired-Sample Exact Test

the 2×2 table of the form

$$y = \begin{matrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{matrix} \quad (\text{T.1})$$

Again, interest is in the equality of binary response rates from two populations, where the data consist of paired, dependent responses. The tests described here determine if the initial response rate is statistically equivalent to the final response rate.

Let $(\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$, denote the four cell probabilities for table (T.1). Let π_1 be the probability that the first response is positive and π_2 be the probability that the second response is positive.

Marginal probabilities can be expressed as

$$\pi_1 = \pi_{11} + \pi_{12}, \text{ and } \pi_2 = \pi_{21} + \pi_{22}. \quad (\text{T.2})$$

Therefore the null hypothesis can be expressed as

$$H_0: \pi_1 = \pi_2, \quad (\text{T.3})$$

versus the alternative

$$H_1: \pi_1 \neq \pi_2, \quad (\text{T.4})$$

Using (T.2), $\pi_1 = \pi_2$ implies that $\pi_{12} = \pi_{21}$. The inference becomes focused on the probabilities of discordant pairs, and subsequent test statistics are all functions of the difference $y_{21} - y_{12}$.

East calculates the power for the exact conditional test of the null hypothesis:

$$H_0: \pi_{12} = \pi_{21}.$$

against the specific alternative

$$H_1: \pi_{21} - \pi_{12} = \Delta.$$

In both cases, the user inputs the probability of a discordant pair, which is $\Psi = \pi_{12} + \pi_{21}$, and the difference of interest, Δ . From this information, East determines the original cell probabilities

$$\pi_{12} = \frac{\Psi - \Delta}{2}, \quad \pi_{21} = \Psi - \pi_{12}.$$

Exact Conditional Test - Power

The unconditional power for the exact conditional test uses the fact that, conditional on the number of discordant pairs $N_d = y_{12} + y_{21}$, Y_{12} has a binomial distribution with number of trials N_d and success probability $\pi_{12}/(\pi_{12} + \pi_{21})$. Thus, one can use the power calculation for a single binomial proportion to obtain the exact conditional power for McNemar's test. Let y_α be the cut-off value for rejecting the null hypothesis with a level- α one-sided exact McNemar test conditional on N_d . Thus, y_α is the smallest integer such that, under the null hypothesis,

$$\Pr(Y_{12} \geq y_\alpha | N_d, H_0) \leq \alpha, \quad (\text{T.5})$$

where

$$\Pr(Y_{12} = y | N_d, H_0) = \frac{(0.5)^{N_d} N_d!}{y_{12}!(N_d - y_{12})!}. \quad (\text{T.6})$$

The conditional power of the one-sided exact conditional McNemar test is thus

$$\Pr(Y_{12} \geq y_\alpha | N_d, H_1) = \sum_{y \geq y_\alpha} \binom{N_d}{y} \left(\frac{\pi_{12}}{\pi_{12} + \pi_{21}} \right)^y \left(\frac{\pi_{21}}{\pi_{12} + \pi_{21}} \right)^{N_d - y}. \quad (\text{T.7})$$

Exact Conditional Test - Sample Size

The exact conditional sample size for fixed parameter and power values are obtained by evaluating the exact conditional power functions over a range of sample sizes until the resulting N is found that obtains the desired power. Since neither α nor β are guaranteed to be attainable due to discreteness of the binomial distribution, the solution to this parameter space search N is not unique. The choice of sample size for a particular trial should depend on the priorities given to type 1 and 2 errors by the investigator. Possible prioritization may involve:

- Primarily maximizing the attained type 1 error while maintaining it below α
- Primarily maximizing the attained type 2 error while maintaining it below β
- Optimizing to get α^* and β^* as close to α and β as possible.

The most practical choice of sample size, however, may be that sample size above which power is guaranteed to be at least $(1 - \beta)$.

U Theory - Design - Simon's Two-Stage Design

In this appendix, we describe the theory behind the two-stage optimal design for phase 2 clinical trials developed by Simon (1989). This design is optimal in the sense that it minimizes the maximum expected sample size under the null hypothesis. It was developed for oncology trials to ensure that patients do not receive a treatment that is clearly inferior to other available options. East also supports Simon's minimax approach as well as an admissible two-stage design, which is a graphical method used to search for an alternative with more favorable features (Jung, et al. 2004).

Simon's Optimal design

Of primary interest is testing the null hypothesis $H_0 : \pi \leq \pi_0$ that the true response probability is less than some uninteresting level π_0 . If the null hypothesis is indeed true, then the probability of a false positive should be controlled at level α . This means that the decision to carry the drug into later phases of clinical development should be less than α .

Suppose an alternative hypothesis $H_1 : \pi \geq \pi_1$ is also specified, which claims that the true response probability is at least some desirable target level π_1 . If this hypothesis is true, then the probability of a false negative should be controlled to be less than a pre-specified value β .

Finally, in addition to these two constraints, the design should be optimal in the sense that it minimizes the number of patients treated with a drug of low activity.

Define n_1 and n_2 to be the number of patients studied in the first and second stage of the trial, respectively. The expected sample size n can be computed as

$$E[n] = n_1 + (1 - PET)n_2$$

where

$$PET = \sum_{i=0}^{s_1} Bin(i; \pi, n_1)$$

Here, PET represents the probability of early termination after the first stage, a decision based on the number of responses observed for the n_1 patients in that stage of the trial. Terminating the experiment at the end of the first stage for futility is based on the herein implicit rule that the treatment is dropped if s_1 or fewer responses are observed.

At the end of the second stage, the treatment is considered ineffective if a total of s responses are observed in all $n = n_1 + n_2$ patients of the trial. Thus, the probability of

concluding the treatment is ineffective is given by

$$\sum_{i=0}^{s_1} \text{Bin}(i; \pi, n_1) + \sum_{j=(s_1+1)}^{\min[n_1, s]} \text{Bin}(j; \pi, n_1) \sum_{k=0}^{(s-j)} \text{Bin}(k; \pi, n_2).$$

To optimally design the trial given parameters π_0 , π_1 , α , and β , this probability statement must be evaluated under the null hypothesis that $\pi = \pi_0$ over all values of n_1 and n_2 as well as s_1 and s .

Note that early termination of the trial for efficacy is not permitted in this design. If it were, it would be possible to further reduce the expected sample size of the trial. However, the ethical imperative of this type of trials is to terminate early for futility.

East optimizes the two-stage design using exact binomial probabilities. For each value of total sample size n and each value of stage 1 sample size n_1 in the range $(1, n - 1)$, integer values s_1 and s are found that satisfy the type 1 and type 2 error constraints and minimize the expected sample size when $\pi = \pi_0$. The search occurs over the range $s_1 \in (0, n_1)$. For each value of s_1 the maximum value of s satisfying the type 2 error constraint is determined. Next the set of parameters (n, n_1, s_1, s) is examined to see whether it satisfies the type 1 error constraint. If it does, then the expected sample size of the corresponding design is compared to the minimum expected sample size previously achieved by the search algorithm. The search continues over the entire range of s_1 . This is repeated for values in the range of n_1 while keeping n fixed.

The search over the range of n begins from the lower value of

$$\bar{\pi}(1 - \bar{\pi}) \left[\frac{z_{1-\alpha} + z_{1-\beta}}{\pi_1 - \pi_0} \right]$$

where $\bar{\pi} = (\pi_0 + \pi_1)/2$. A check must be performed below this starting point to ensure that this is indeed the smallest maximum sample size n for which there is a nontrivial $(n_1, n_2 > 0)$ two-stage design satisfying the type 1 and type 2 error constraints. The enumeration procedure then searches upwards from this minimum value of n until it is clear that the optimum had been determined.

The minimum expected sample size for fixed n is not a unimodal function of n because of the discreteness of the underlying binomial distributions. Nevertheless, eventually as n increases the value of the local minima increase and it becomes clear that a global minimum has been found.

U Theory - Design - Simon's Two-Stage Design

Simon's Minimax and Admissible designs

In addition to the optimal design, East offers Simon's minimax approach, which minimizes the total sample size while satisfying both type-I and type-II constraints. The admissible two-stage design (Jung, et al. 2004), employs a graphical method geared to search for an alternative with more favorable features. This approach provides a compromised solution between the minimax and the optimal designs, that also satisfy type-I and type-II constraints. Resulting designs yield the same total sample sizes, as well as having the minimum expected sample size under the Null.

V Theory-Design - Binomial Two-Sample Exact Tests

This appendix deals with exact power and sample size computations for comparing two independent binomials. Exact power and sample size calculations are considered for the two-sided Fisher's test, the unconditional one-sided tests of superiority, non-inferiority test, and two one-sided tests of equivalence.

Exact tests on categorical data are usually computed conditionally, by fixing the margins of the contingency table at their observed values. Corresponding power computations are, however, more useful if they are performed unconditionally, before these table margins have been observed. Only then can they aid in determining if the sample size proposed for the study is adequate. This appendix shows how to obtain exact unconditional power as a weighted sum of exact conditional powers, and applies the results to exact conditional tests on 2×2 contingency tables. It also covers the exact power and sample size computations for exact unconditional tests of non-inferiority and equivalence of two binomial populations.

The methods used by East to compute power and sample size of these two-sample exact tests are based on Fleiss (1981) for Fisher's exact test and the conditional exact superiority test, Suissa and Shuster (1985) for the unconditional exact superiority test, Chan (1988) for the unconditional exact non-inferiority test, and finally Dunnett and Gent (1977) for the exact equivalence test. In the equivalence testing of two binomials, power of

In all that follows, consider sampling from two independent binomial populations. Suppose x_c responses out of n_c subjects are observed in the control group. The mean response rate in this group is denoted π_c . Similarly define x_t , n_t , and π_t for the treatment group. The observed data may be represented in a 2×2 contingency table \mathbf{x} of the form

| | | |
|-------------|-------------|---------|
| x_c | x_t | m |
| $n_c - x_c$ | $n_t - x_t$ | $N - m$ |
| n_c | n_t | N |

Section (V.1) explains computation of the power of Fisher's exact test. In section (V.2) power of Barnard's unconditional test of superiority is described. Section (V.3) continues with the power of the unconditional test of non-inferiority. Power for the unconditional test of equivalence between two binomial proportions is considered in section (V.4). Finally, section (V.5) briefly describes the computation of sample size for all these tests.

V Theory-Design - Binomial Two-Sample Exact Tests

V.1 Fisher's Exact Test

V.1.1 Power

Fisher's exact test is concerned with testing the null hypothesis

$$H_0 : \pi_c = \pi_t \equiv \pi \quad (\text{V.1})$$

versus the two-sided alternative hypothesis

$$H_1 : \pi_c \neq \pi_t \quad (\text{V.2})$$

at fixed sample sizes n_c and n_t .

As is well known, the exact probability of \mathbf{x} under H_0 , conditional on $x_c + x_t = m$, is given by

$$\Pr(\mathbf{x}|\mathbf{m}, H_0) = \frac{\binom{n_c}{x_c} \binom{n_t}{x_t}}{\binom{N}{m}}. \quad (\text{V.3})$$

Notice that (V.3) does not depend on the common null response probability π . Thus this probability need not be specified for purposes of calculating power. The two response probabilities π_c and π_t are, however, needed to evaluate the probability of \mathbf{x} under H_1 .

Fisher's exact test is based on the exact distribution of the test statistic

$$T = -\log \left[\frac{\binom{n_c}{x_c} \binom{n_t}{x_t}}{\binom{N}{m}} \right]. \quad (\text{V.4})$$

V.1.1 Exact Unconditional Power for Fisher's Exact Test

Consider first the exact power of level- α tests based on the statistic T . Let

$$\Gamma_m = \{\mathbf{x} : \mathbf{x}_c + \mathbf{x}_t = \mathbf{m}\} \quad (\text{V.5})$$

and define the critical region

$$\Gamma_m(t) = \{\mathbf{x} \in \Gamma_m : T \geq t\}. \quad (\text{V.6})$$

The exact null distribution of T may then be obtained by evaluating

$$\Pr(T \geq t|m, H_0) = \sum_{\mathbf{x} \in \Gamma_m(t)} \left[\frac{\binom{n_c}{x_c} \binom{n_t}{x_t}}{\binom{N}{m}} \right], \quad (\text{V.7})$$

for each possible value of t .

Let α be the maximum allowable type-1 error and $t_\alpha(m)$ be the smallest possible cut-off such that

$$\Pr(T \geq t_\alpha(m) | m, H_0) \leq \alpha. \tag{V.8}$$

The **conditional** power of Fisher’s exact test is defined as

$$\Pr(T \geq t_\alpha(m) | m, H_1) = \sum_{\mathbf{x} \in \Gamma_m(t_\alpha(m))} \left[\frac{Q_c Q_t}{\sum_{\mathbf{x} \in \Gamma_m} Q_c Q_t} \right]. \tag{V.9}$$

where

$$Q_c = \binom{n_c}{x_c} \pi_c^{x_c} (1 - \pi_c)^{n_c - x_c} \tag{V.10}$$

$$Q_t = \binom{n_t}{x_t} \pi_t^{x_t} (1 - \pi_t)^{n_t - x_t} \tag{V.11}$$

Denote this two-sided conditional power by $(1 - \beta(m))$. Then the two-sided **unconditional** power of Fisher’s exact test is defined as

$$(1 - \beta) = \sum_{m=0}^N (1 - \beta(m)) P(m) \tag{V.12}$$

where

$$P(m) = \Pr(x_c + x_t = m | H_1), \tag{V.13}$$

is a convolution of two binomials under H_1 . It is relatively straightforward to compute equation (V.12) as only 2×2 tables are involved.

V.2 Power of Unconditional Test of Superiority

V.2.1 Diff.of Proportions

V.2.2 Ratio of Proportions

V.2.1 Superiority Test: Difference of Proportions

Superiority for Difference of Proportions – Case 1 Suppose it is desired to test $H_0 : \pi_t - \pi_c \leq 0$ against the one-sided alternative $H_1 : \pi_t - \pi_c > 0$. Let π_t and π_c

V Theory-Design - Binomial Two-Sample Exact Tests

denote the binomial probabilities for the treatment and control arms, respectively. Let x_t and x_c be the observed numbers of responses for the treatment and control arms, respectively. Let $\delta = \pi_t - \pi_c$. It is of interest to test the null hypothesis $H_0 : \delta \leq 0$ against one-sided alternative $H_1 : \delta > 0$. Let $\hat{\pi}_i$ denote the estimate of π_i based on n_i observations from treatment i . The test statistic can be defined by

$$T(x_t, x_c) = \frac{\hat{\pi}_t - \hat{\pi}_c}{\sqrt{\tilde{\pi} (1 - \tilde{\pi}) \left(\frac{1}{n_c} + \frac{1}{n_t} \right)}} \quad (\text{V.14})$$

where $\hat{\pi}_t, \hat{\pi}_c$ and $\tilde{\pi}$ are given by

$$\hat{\pi}_c = \frac{x_c}{n_c}, \hat{\pi}_t = \frac{x_t}{n_t}, \tilde{\pi} = \frac{x_t + x_c}{n_t + n_c} \quad (\text{V.15})$$

Let $\mathcal{X} = \{(x_t, x_c) : 0 \leq x_t \leq n_t, 0 \leq x_c \leq n_c\}$ denote the sample space for the 2×2 table. Let $f_{\pi_t, \pi_c}(x_t, x_c)$ denote the probability of observing the data $(x_t, x_c) \in \mathcal{X}$ when the response rates for the treatment and control arms are π_t and π_c , respectively, which is given by

$$f_{\pi_t, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} \pi_t^{x_t} (1 - \pi_t)^{n_t - x_t} \pi_c^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (\text{V.16})$$

For a given π_c and nominal significance level α , let b_{π_c} be defined by

$$b_{\pi_c} = \sup \{b : P_{\pi_c}(T(x_t, x_c) < b \mid H_0) \leq \alpha\} \quad (\text{V.17})$$

This probability $P_{\pi_c}(T(x_t, x_c) < b \mid H_0)$ is calculated based on the exact distribution of $T(x_t, x_c)$ under the null hypothesis $\pi_c = \pi_t$. This implies that, for a given π_c , b_{π_c} is defined such that

$$\sup \left\{ b : P_{\pi_c}(T(x_t, x_c) < b \mid H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_b(x_t, x_c) f_{\pi_c, \pi_c}(x_t, x_c) \leq \alpha \right\} \quad (\text{V.18})$$

where the indicator function is defined by

$$I_b(x_t, x_c) = \begin{cases} 1 & \text{if } T(x_t, x_c) < b \\ 0 & \text{otherwise} \end{cases} \quad (\text{V.19})$$

Note that there is a one-to-one correspondence between the critical value b_{π_c} and the control rate π_c . Let $b^* = \inf \{b_{\pi_c} : \pi_c \in (0, 1)\}$ and suppose that this infimum takes place at π_c^* .

The decision rule of the exact test is to reject H_0 if $T(x_t, x_c) < b^*$. Since b^* is the infimum of the critical values over the possible range of π_c , this test guarantees the type I error control regardless of the underlying true response rate for the control arm.

The attained significance level of this test is the exact probability of rejecting the null hypothesis when the underlying control rate equals π_c^* which is given by

$$P_{\pi_c^*} (T(x_t, x_c) < b^* | H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\pi_c^*, \pi_c^*}(x_t, x_c) \quad (\text{V.20})$$

Note that the attained significance level is the maximum type I error one can actually commit using this test given the desired significance level α . Due to the discreteness of the distributions, the attained significance level is always bounded above by α .

Next we will show how the unconditional power of this exact test is calculated. The unconditional power is the probability of rejecting the null hypothesis under the alternative hypothesis. Suppose that one is interested in the power of this test under $\delta = \delta_1 (< 0)$ and π_c . Under the alternative $\pi_t = \pi_c + \delta_1$. Then the unconditional power is given by

$$P_{\pi_c} (T(x_t, x_c) < b^* | H_1) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\pi_c + \delta_1, \pi_c}(x_t, x_c) \quad (\text{V.21})$$

where

$$f_{\pi_c + \delta_1, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} (\pi_c + \delta_1)^{x_t} (1 - \pi_c - \delta_1)^{n_t - x_t} (\pi_c)^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (\text{V.22})$$

Superiority for Difference of Proportions – Case 2 Suppose it is desired to test $H_0 : \pi_t - \pi_c \geq 0$ against the one-sided alternative $H_1 : \pi_t - \pi_c < 0$. Let π_t and π_c denote the binomial probabilities for the treatment and control arms, respectively. Let x_t and x_c be the observed numbers of responses for the treatment and control arms, respectively. Let $\delta = \pi_t - \pi_c$. It is of interest to test the null hypothesis $H_0 : \delta \geq 0$

V Theory-Design - Binomial Two-Sample Exact Tests

against one-sided alternative $H_1 : \delta < 0$. Let $\hat{\pi}_i$ denote the estimate of π_i based on n_i observations from treatment i . The test statistic can be defined by

$$T(x_t, x_c) = \frac{\hat{\pi}_t - \hat{\pi}_c}{\sqrt{\tilde{\pi}(1 - \tilde{\pi})\left(\frac{1}{n_c} + \frac{1}{n_t}\right)}} \quad (\text{V.23})$$

where $\hat{\pi}_t, \hat{\pi}_c$ and $\tilde{\pi}$ are given by

$$\hat{\pi}_c = \frac{x_c}{n_c}, \hat{\pi}_t = \frac{x_t}{n_t}, \tilde{\pi} = \frac{x_c + x_t}{n_t + n_c} \quad (\text{V.24})$$

Let $\mathcal{X} = \{(x_t, x_c) : 0 \leq x_t \leq n_t, 0 \leq x_c \leq n_c\}$ denote the sample space for the 2×2 table. Let $f_{\pi_t, \pi_c}(x_t, x_c)$ denote the probability of observing the data $(x_t, x_c) \in \mathcal{X}$ when the response rates for the treatment and control arms are π_t and π_c , respectively, which is given by

$$f_{\pi_t, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} \pi_t^{x_t} (1 - \pi_t)^{n_t - x_t} \pi_c^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (\text{V.25})$$

For a given π_c and nominal significance level α , let b_{π_c} be defined by

$$b_{\pi_c} = \inf \{b : P_{\pi_c}(T(x_t, x_c) > b \mid H_0) \leq \alpha\} \quad (\text{V.26})$$

This probability $P_{\pi_c}(T(x_t, x_c) > b \mid H_0)$ is calculated based on the exact distribution of $T(x_t, x_c)$ under the null hypothesis $\pi_c = \pi_t$. This implies that, for a given π_c , b_{π_c} is defined such that

$$\inf \left\{ b : P_{\pi_c}(T(x_t, x_c) > b \mid H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_b(x_t, x_c) f_{\pi_c, \pi_c}(x_t, x_c) \leq \alpha \right\} \quad (\text{V.27})$$

where the indicator function is defined by

$$I_b(x_t, x_c) = \begin{cases} 1 & \text{if } T(x_t, x_c) > b \\ 0 & \text{otherwise} \end{cases} \quad (\text{V.28})$$

Note that there is a one-to-one correspondence between the critical value b_{π_c} and the control rate π_c . Let $b^* = \sup \{b_{\pi_c} : \pi_c \in (0, 1)\}$ and suppose that this supremum takes place at π_c^* .

The decision rule of the exact test is to reject H_0 if $T(x_t, x_c) > b^*$. Since b^* is the supremum of the critical values over the possible range of π_c , this test guarantees the type I error control regardless of the underlying true response rate for the control arm.

The attained significance level of this test is the exact probability of rejecting the null hypothesis when the underlying control rate equals π_c^* which is given by

$$P_{\pi_c^*}(T(x_t, x_c) > b^* | H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\pi_c^*, \pi_c^*}(x_t, x_c) \quad (\text{V.29})$$

Note that the attained significance level is the maximum type I error one can actually commit using this test given the desired significance level α . Due to the discreteness of the distributions, the attained significance level is always bounded above by α .

Next we will show how the unconditional power of this exact test is calculated. The unconditional power is the probability of rejecting the null hypothesis under the alternative hypothesis. Suppose that one is interested in the power of this test under $\delta = \delta_1 (> 0)$ and π_c . Under the alternative $\pi_t = \pi_c + \delta_1$. Then the unconditional power is given by

$$P_{\pi_c}(T(x_t, x_c) > b^* | H_1) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\pi_c + \delta_1, \pi_c}(x_t, x_c) \quad (\text{V.30})$$

where

$$f_{\pi_c + \delta_1, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} (\pi_c + \delta_1)^{x_t} (1 - \pi_c - \delta_1)^{n_t - x_t} (\pi_c)^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (\text{V.31})$$

V.2.2 Superiority Test: Ratio of Proportions

Superiority for Ratio of Proportions – Case 1 Suppose that it is desired to test $H_0 : \frac{\pi_t}{\pi_c} \leq 1$ against $H_1 : \frac{\pi_t}{\pi_c} > 1$. Let π_t and π_c denote the binomial probabilities for the treatment and control arms, respectively, and let $\rho = \frac{\pi_t}{\pi_c}$. Let x_t and x_c be the observed number of responses for the treatment and control arms, respectively. It is of

V Theory-Design - Binomial Two-Sample Exact Tests

interest to test the null hypothesis $H_0 : \rho \leq 1$ against the one-sided alternative $H_1 : \rho > 1$. Let $\delta = \ln(\pi_t) - \ln(\pi_c)$. Then it is equivalent to test $H_0 : \delta \leq 0$ against $H_1 : \delta > 0$. Let $\hat{\pi}_i$ denote the estimate of π_i based on n_i observations from treatment i . The test statistic is defined by

$$T = \frac{\ln(\hat{\pi}_t) - \ln(\hat{\pi}_c)}{\sqrt{\frac{1-\tilde{\pi}}{\tilde{\pi}} \left(\frac{1}{n_t} + \frac{1}{n_c}\right)}} \quad (\text{V.32})$$

where $\hat{\pi}_t, \hat{\pi}_c$ and $\tilde{\pi}$ are given by

$$\hat{\pi}_t = \frac{x_t}{n_t}, \hat{\pi}_c = \frac{x_c}{n_c}, \tilde{\pi} = \frac{x_t + x_c}{n_t + n_c} \quad (\text{V.33})$$

Note that $\frac{1-\tilde{\pi}}{\tilde{\pi}} \left(\frac{1}{n_t} + \frac{1}{n_c}\right)$ is the maximum likelihood estimate of the variance of $\ln(\hat{\pi}_t) - \ln(\hat{\pi}_c)$ restricted under the null hypothesis. Let $\mathcal{X} = \{(x_t, x_c) : 0 \leq x_t \leq n_t, 0 \leq x_c \leq n_c\}$ denote the sample space for the 2×2 table. Let $f_{\pi_t, \pi_c}(x_t, x_c)$ denote the probability of observing the data $(x_t, x_c) \in \mathcal{X}$ when the response rates for the treatment and control arms are π_t and π_c , respectively, which is given by

$$f_{\pi_t, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} \pi_t^{x_t} (1 - \pi_t)^{n_t - x_t} \pi_c^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (\text{V.34})$$

For a given π_c and nominal significance level α , let b_{π_c} be defined by

$$b_{\pi_c} = \inf \{b : P_{\pi_c}(T(x_t, x_c) > b \mid H_0) \leq \alpha\} \quad (\text{V.35})$$

This probability $P_{\pi_c}(T(x_t, x_c) > b \mid H_0)$ is calculated based on the exact distribution of $T(x_t, x_c)$. This implies that, for a given π_c, c_{π_c} is such that

$$\inf \left\{ b : P_{\pi_c}(T(x_t, x_c) > b \mid H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_b(x_t, x_c) f_{\pi_c, \pi_c}(x_t, x_c) \leq \alpha \right\} \quad (\text{V.36})$$

where the indicator function is defined by

$$I_b(x_t, x_c) = \begin{cases} 1 & \text{if } T(x_t, x_c) > b \\ 0 & \text{otherwise} \end{cases} \quad (\text{V.37})$$

Note that there is a one-to-one correspondence between the critical value b_{π_c} and the control rate π_c . Let $b^* = \sup \{b_{\pi_c} : \pi_c \in (0, 1)\}$ and this supremum takes place at π_c^* . The decision rule of the exact test is to reject H_0 if $T(x_t, x_c) > b^*$. Since b^* is the supremum of the critical values over the possible range of π_c , this test will guarantee the type I error control regardless of the underlying true response rate for the control arm.

The attained significance level of this test is the exact probability of rejecting the null hypothesis when the control rate equals π_c^* which is given by

$$P_{\pi_c^*}(T(x_t, x_c) > b^* | H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\pi_c^*, \pi_c^*}(x_t, x_c) \quad (V.38)$$

Note that the attained significance level is the maximum type I error one can actually commit using this test given the desired significance level. Due to the discreteness of the distributions, the attained significance level is always bounded above by α .

Next we will show how the unconditional power of this exact test is calculated. The unconditional power is the probability of rejecting the null hypothesis under the alternative hypothesis. Suppose that one is interested in the power of this test at $\rho = \rho_1 > 1$ and π_c . Under the alternative $\pi_t = \rho_1 \pi_c$, then the unconditional power is given by

$$P_{\pi_c}(T(x_t, x_c) > b^* | H_1) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\rho_1 \pi_c, \pi_c}(x_t, x_c) \quad (V.39)$$

where

$$f_{\rho_1 \pi_c, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} (\rho_1 \pi_c)^{x_t} (1 - \rho_1 \pi_c)^{n_t - x_t} \pi_c^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (V.40)$$

Superiority for Ratio for Proportions – Case 2 Suppose that it is desired to test $H_0 : \frac{\pi_t}{\pi_c} \geq 1$ against the one-sided alternative $H_1 : \frac{\pi_t}{\pi_c} < 1$. In this case, we use the same test statistic as in Case 1

$$T = \frac{\ln(\hat{\pi}_t) - \ln(\hat{\pi}_c)}{\sqrt{\frac{1 - \hat{\pi}}{\hat{\pi}} \left(\frac{1}{n_t} + \frac{1}{n_c} \right)}} \quad (V.41)$$

V Theory-Design - Binomial Two-Sample Exact Tests

where $\hat{\pi}_t, \hat{\pi}_c, \tilde{\pi}$ are defined in the same way as the above.

Let $\mathcal{X} = \{(x_t, x_c) : 0 \leq x_t \leq n_t, 0 \leq x_c \leq n_c\}$ denote the set of all possible data values that could possibly be observed for the 2×2 table. Let $f_{\pi_t, \pi_c}(x_t, x_c)$ denote the probability of observing the data $(x_t, x_c) \in \mathcal{X}$ when the response rates for the treatment and control arms are π_t and π_c , respectively, which is given by

$$f_{\pi_t, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} \pi_t^{x_t} (1 - \pi_t)^{n_t - x_t} \pi_c^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (\text{V.42})$$

For a given π_c and nominal significance level α , let b_{π_c} be defined by

$$b_{\pi_c} = \sup \{b : P_{\pi_c}(T(x_t, x_c) < b \mid H_0) \leq \alpha\} \quad (\text{V.43})$$

This probability $P_{\pi_c}(T(x_t, x_c) < b \mid H_0)$ is calculated based on the exact distribution of $T(x_t, x_c)$. This implies that, for a given π_c , b_{π_c} is such that

$$\sup \left\{ b : P_{\pi_c}(T(x_t, x_c) < b \mid H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_b(x_t, x_c) f_{\pi_c, \pi_c}(x_t, x_c) \leq \alpha \right\} \quad (\text{V.44})$$

where the indicator function is defined by

$$I_b(x_t, x_c) = \begin{cases} 1 & \text{if } T(x_t, x_c) < b \\ 0 & \text{otherwise} \end{cases} \quad (\text{V.45})$$

Note that there is a one-to-one correspondence between the critical value b_{π_c} and the control rate π_c . Let $b^* = \inf \{b_{\pi_c} : \pi_c \in (0, 1)\}$ and this infimum takes place at π_c^* . The decision rule of the exact test is to reject H_0 if $T(x_t, x_c) < b^*$. Since b^* is the infimum of the critical values over the possible range of π_c , this test will guarantee the type I error control regardless of the underlying true response rate for the control arm.

The attained significance level of this test is the exact probability of rejecting the null hypothesis when the control rate equals π_c^* which is given by

$$P_{\pi_c^*}(T(x_t, x_c) < b^* \mid H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\pi_c^*, \pi_c^*}(x_t, x_c) \quad (\text{V.46})$$

Note that the attained significance level is the maximum type I error one can actually commit using this test given the desired significance level. Due to the discreteness of the distributions, the attained significance level is always bounded above by α .

The unconditional power is the probability of rejecting the null hypothesis under the alternative hypothesis. Suppose that one is interested in the power of this test at $\rho = \rho_1 < 1$ and π_c . Under the alternative $\rho = \rho_1, \pi_t = \rho_1 \pi_c$, then the unconditional power is given by

$$P_{\pi_c}(T(x_t, x_c) < b^* | H_1) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\rho_1 \pi_c, \pi_c}(x_t, x_c) \quad (V.47)$$

where

$$f_{\rho_1 \pi_c, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} (\rho_1 \pi_c)^{x_t} (1 - \rho_1 \pi_c)^{n_t - x_t} \pi_c^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (V.48)$$

V.3 Power of the Unconditional Test of Non-Inferiority

V.3.1 Diff.of Proportions

V.3.2 Ratio of Proportions

V.3.1 Non-inferiority Test: Difference of Proportions

Non-inferiority for Difference of Proportions – Case 1 Suppose it is desired to test $H_0 : \pi_t - \pi_c \leq \delta_0 (\delta_0 < 0)$ against the one-sided alternative $H_1 : \pi_t - \pi_c > \delta_0$. Let π_t and π_c denote the binomial probabilities for the treatment and control arms, respectively. Let x_t and x_c be the observed numbers of responses for the treatment and control arms, respectively. Let $\delta = \pi_t - \pi_c$. It is of interest to test the null hypothesis $H_0 : \delta \leq \delta_0$ against one-sided alternative $H_1 : \delta > \delta_0$. Let $\hat{\pi}_i$ denote the estimate of π_i based on n_i observations from treatment i . The test statistic can be defined by

$$T(x_t, x_c) = \frac{\hat{\pi}_t - \hat{\pi}_c - \delta_0}{\sqrt{\frac{\tilde{\pi}_c(1-\tilde{\pi}_c)}{n_c} + \frac{+\tilde{\pi}_t(1-\tilde{\pi}_t)}{n_t}}} \quad (V.49)$$

where $\hat{\pi}_t$ and $\hat{\pi}_c$ are given by

$$\hat{\pi}_c = \frac{x_c}{n_c}, \hat{\pi}_t = \frac{x_t}{n_t} \quad (V.50)$$

and $\tilde{\pi}_t$ and $\tilde{\pi}_c$ are the maximum likelihood estimates of π_t and π_c , respectively, restricted under the null hypothesis such that $\tilde{\pi}_t - \tilde{\pi}_c = \delta_0$. Miettinen and Nurminen

V Theory-Design - Binomial Two-Sample Exact Tests

(1985) have shown that one may obtain these restricted maximum likelihood estimates by solving the third degree likelihood equation

$$\sum_{k=0}^3 L_k \tilde{\pi}_c^k = 0 \tag{V.51}$$

for $\tilde{\pi}_c$ and setting $\tilde{\pi}_t = \tilde{\pi}_c + \delta_0$, where

$$\begin{aligned} L_3 &= N = n_c + n_t \\ L_2 &= (n_t + 2n_c) \delta_0 - N - x_c - x_t \\ L_1 &= (n_c \delta_0 - N - 2x_c) \delta_0 + x_c + x_t \\ L_0 &= x_c \delta_0 (1 - \delta_0) \end{aligned} \tag{V.52}$$

Let $\mathcal{X} = \{(x_t, x_c) : 0 \leq x_t \leq n_t, 0 \leq x_c \leq n_c\}$ denote the sample space for the 2×2 table. Let $f_{\pi_t, \pi_c}(x_t, x_c)$ denote the probability of observing the data $(x_t, x_c) \in \mathcal{X}$ when the response rates for the treatment and control arms are π_t and π_c , respectively, which is given by

$$f_{\pi_t, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} \pi_t^{x_t} (1 - \pi_t)^{n_t - x_t} \pi_c^{x_c} (1 - \pi_c)^{n_c - x_c} \tag{V.53}$$

For a given π_c and nominal significance level α , let b_{π_c} be defined by

$$b_{\pi_c} = \sup \{b : P_{\pi_c}(T(x_t, x_c) < b \mid H_0) \leq \alpha\} \tag{V.54}$$

This probability $P_{\pi_c}(T(x_t, x_c) < b \mid H_0)$ is calculated based on the exact distribution of $T(x_t, x_c)$ under the null hypothesis $\pi_t - \pi_c = \delta_0$. This implies that, for a given π_c , b_{π_c} is defined such that

$$\sup \left\{ b : P_{\pi_c}(T(x_t, x_c) < b \mid H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_b(x_t, x_c) f_{\pi_c - \delta_0, \pi_c}(x_t, x_c) \leq \alpha \right\} \tag{V.55}$$

where the indicator function is defined by

$$I_b(x_t, x_c) = \begin{cases} 1 & \text{if } T(x_t, x_c) < b \\ 0 & \text{otherwise} \end{cases} \tag{V.56}$$

Note that there is a one-to-one correspondence between the critical value b_{π_c} and the control rate π_c . Let $b^* = \inf \{b_{\pi_c} : \pi_c \in (0, 1)\}$ and suppose that this infimum takes place at π_c^* .

The decision rule of the exact test is to reject H_0 if $T(x_t, x_c) < b^*$. Since b^* is the infimum of the critical values over the possible range of π_c , this test guarantees the type I error control regardless of the underlying true response rate for the control arm.

The attained significance level of this test is the exact probability of rejecting the null hypothesis when the null hypothesis is true and the underlying control rate equals π_c^* which is given by

$$P_{\pi_c^*} (T(x_t, x_c) < b^* | H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\pi_c^* - \delta_0, \pi_c^*}(x_t, x_c) \quad (V.57)$$

Note that the attained significance level is the maximum type I error one can actually commit using this test given the desired significance level α . Due to the discreteness of the distributions, the attained significance level is always bounded above by α .

Next we will show how the unconditional power of this exact test is calculated. The unconditional power is the probability of rejecting the null hypothesis under the alternative hypothesis. Suppose that one is interested in the power of this test under $\delta = \delta_1 (< \delta_0)$ and π_c . Under the alternative we have $\pi_c = \pi_t + \delta_1$. Then the unconditional power is given by

$$P_{\pi_c} (T(x_t, x_c) > b^* | H_1) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\pi_c + \delta_1, \pi_c}(x_t, x_c) \quad (V.58)$$

where

$$f_{\pi_c + \delta_1, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} (\pi_c + \delta_1)^{x_t} (1 - \pi_c - \delta_1)^{n_t - x_t} (\pi_c)^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (V.59)$$

Non-inferiority for Difference of Proportions – Case 2 Suppose it is desired to test $H_0 : \pi_t - \pi_c \geq \delta_0 (> 0)$ against the one-sided alternative $H_1 : \pi_t - \pi_c < \delta_0$. Let π_t and π_c denote the binomial probabilities for the treatment and control arms, respectively. Let x_t and x_c be the observed numbers of responses for the treatment and

V Theory-Design - Binomial Two-Sample Exact Tests

control arms, respectively. Let $\delta = \pi_t - \pi_c$. It is of interest to test the null hypothesis $H_0 : \delta \geq \delta_0$ against one-sided alternative $H_1 : \delta < \delta_0$. Let $\hat{\pi}_i$ denote the estimate of π_i based on n_i observations from treatment i . The test statistic can be defined by

$$T(x_t, x_c) = \frac{\hat{\pi}_t - \hat{\pi}_c - \delta_0}{\sqrt{\frac{\tilde{\pi}_c(1-\tilde{\pi}_c)}{n_c} + \frac{\tilde{\pi}_t(1-\tilde{\pi}_t)}{n_t}}} \quad (\text{V.60})$$

where $\hat{\pi}_t, \hat{\pi}_c, \tilde{\pi}_t$ and $\tilde{\pi}_c$ are defined in the same way as in Case 1.

Let $\mathcal{X} = \{(x_t, x_c) : 0 \leq x_t \leq n_t, 0 \leq x_c \leq n_c\}$ denote the sample space for the 2×2 table. Let $f_{\pi_t, \pi_c}(x_t, x_c)$ denote the probability of observing the data $(x_t, x_c) \in \mathcal{X}$ when the response rates for the treatment and control arms are π_t and π_c , respectively, which is given by

$$f_{\pi_t, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} \pi_t^{x_t} (1 - \pi_t)^{n_t - x_t} \pi_c^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (\text{V.61})$$

For a given π_c and nominal significance level α , let b_{π_c} be defined by

$$b_{\pi_c} = \inf \{b : P_{\pi_c}(T(x_t, x_c) > b \mid H_0) \leq \alpha\} \quad (\text{V.62})$$

This probability $P_{\pi_c}(T(x_t, x_c) > b \mid H_0)$ is calculated based on the exact distribution of $T(x_t, x_c)$ under the null hypothesis $\pi_t - \pi_c = \delta_0$. This implies that, for a given π_c , b_{π_c} is defined such that

$$\inf \left\{ b : P_{\pi_c}(T(x_t, x_c) > b \mid H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_b(x_t, x_c) f_{\pi_c - \delta_0, \pi_c}(x_t, x_c) \leq \alpha \right\} \quad (\text{V.63})$$

where the indicator function is defined by

$$I_b(x_t, x_c) = \begin{cases} 1 & \text{if } T(x_t, x_c) > b \\ 0 & \text{otherwise} \end{cases} \quad (\text{V.64})$$

Note that there is a one-to-one correspondence between the critical value b_{π_c} and the control rate π_c . Let $b^* = \sup \{b_{\pi_c} : \pi_c \in (0, 1)\}$ and suppose that this supremum takes place at π_c^* .

The decision rule of the exact test is to reject H_0 if $T(x_t, x_c) > b^*$. Since b^* is the supremum of the critical values over the possible range of π_c , this test guarantees the type I error control regardless of the underlying true response rate for the control arm.

The attained significance level of this test is the exact probability of rejecting the null hypothesis when the null hypothesis is true and the underlying control rate equals π_c^* which is given by

$$P_{\pi_c^*}(T(x_t, x_c) > b^* | H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\pi_c^* - \delta_0, \pi_c^*}(x_t, x_c) \quad (\text{V.65})$$

Note that the attained significance level is the maximum type I error one can actually commit using this test given the desired significance level α . Due to the discreteness of the distributions, the attained significance level is always bounded above by α .

Next we will show how the unconditional power of this exact test is calculated. The unconditional power is the probability of rejecting the null hypothesis under the alternative hypothesis. Suppose that one is interested in the power of this test under $\delta = \delta_1 (> \delta_0)$ and π_c . Under the alternative we have $\pi_c = \pi_t + \delta_1$. Then the unconditional power is given by

$$P_{\pi_c}(T(x_t, x_c) > b^* | H_1) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\pi_c + \delta_1, \pi_c}(x_t, x_c) \quad (\text{V.66})$$

where

$$f_{\pi_c + \delta_1, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} (\pi_c + \delta_1)^{x_t} (1 - \pi_c - \delta_1)^{n_t - x_t} (\pi_c)^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (\text{V.67})$$

V.3.2 Non-inferiority Test: Ratio of Proportions

Non-inferiority for Ratio of Proportions – Case 1 Suppose it is desired to test $H_0 : \frac{\pi_t}{\pi_c} \leq \rho_0 (< 1)$ against the one-side alternative $H_1 : \frac{\pi_t}{\pi_c} > \rho_0$. An alternative approach to establishing non-inferiority of an experimental treatment to the control treatment with respect to the ratio of probabilities was proposed by Farrington and Manning (1990). Let π_t and π_c denote the binomial probabilities for the treatment and control arms, respectively. Let x_t and x_c be the observed numbers of responses for the treatment and control arms, respectively. Let $\rho = \frac{\pi_t}{\pi_c}$. Suppose that, for some $\rho_0 < 1$,

V Theory-Design - Binomial Two-Sample Exact Tests

one is interested in testing the null hypothesis $H_0 : \rho \leq \rho_0$ against one-sided alternative $H_1 : \rho > \rho_0$. Let $\hat{\pi}_i$ denote the estimate of π_i based on n_i observations from treatment i . The test statistic can be defined by

$$T(x_t, x_c) = \frac{\hat{\pi}_t - \rho_0 \hat{\pi}_c}{\sqrt{\frac{\hat{\pi}_t(1-\hat{\pi}_t)}{n_t} + \frac{\rho_0^2 \hat{\pi}_c(1-\hat{\pi}_c)}{n_c}} \quad (\text{V.68})$$

where $\hat{\pi}_t$ and $\hat{\pi}_c$ are given by

$$\hat{\pi}_t = \frac{x_t}{n_t}, \hat{\pi}_c = \frac{x_c}{n_c} \quad (\text{V.69})$$

and $\tilde{\pi}_t$ and $\tilde{\pi}_c$ are the maximum likelihood estimates of π_t and π_c , respectively, restricted under the null hypothesis such that $\frac{\tilde{\pi}_t}{\tilde{\pi}_c} = \rho_0$. Miettinen and Nurminen (1985) have shown that one may obtain these restricted maximum likelihood estimates by solving a quadratic likelihood equation. Thus

$$\tilde{\pi}_c = \frac{-B - \sqrt{B^2 - 4AC}}{2A} \quad (\text{V.70})$$

and

$$\tilde{\pi}_t = \rho_0 \tilde{\pi}_c \quad (\text{V.71})$$

where

$$\begin{aligned} A &= \rho_0(n_t + n_c) \\ B &= -(\rho_0 n_c + x_c + n_t + \rho_0 x_t) \\ C &= x_c + x_t \end{aligned} \quad (\text{V.72})$$

Let $\mathcal{X} = \{(x_t, x_c) : 0 \leq x_t \leq n_t, 0 \leq x_c \leq n_c\}$ denote the sample space for the 2×2 table. Let $f_{\pi_t, \pi_c}(x_t, x_c)$ denote the probability of observing the data $(x_t, x_c) \in \mathcal{X}$ when the response rates for the treatment and control arms are π_t and π_c , respectively, which is given by

$$f_{\pi_t, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} \pi_t^{x_t} (1 - \pi_t)^{n_t - x_t} \pi_c^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (\text{V.73})$$

For a given π_c and nominal significance level α , let b_{π_c} be defined by

$$b_{\pi_c} = \inf \{b : P_{\pi_c}(T(x_t, x_c) > b \mid H_0) \leq \alpha\} \quad (\text{V.74})$$

This probability $P_{\pi_c}(T(x_t, x_c) > b \mid H_0)$ is calculated based on the exact distribution of $T(x_t, x_c)$ under the null hypothesis $\pi_t = \rho_0\pi_c$. This implies that, for a given π_c, b_{π_c} is such that

$$\inf \left\{ b : P_{\pi_c}(T(x_t, x_c) > b \mid H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_b(x_t, x_c) f_{\rho_0\pi_c, \pi_c}(x_t, x_c) \leq \alpha \right\} \quad (\text{V.75})$$

where the indicator function is defined by

$$I_b(x_t, x_c) = \begin{cases} 1 & \text{if } T(x_t, x_c) > b \\ 0 & \text{otherwise} \end{cases} \quad (\text{V.76})$$

Note that there is a one-to-one correspondence between the critical value b_{π_c} and the control rate π_c . Let $b^* = \sup \{b_{\pi_c} : \pi_c \in (0, 1)\}$ and suppose that this supremum takes place at π_c^* . The decision rule of the exact test is to reject H_0 if $T(x_t, x_c) > b^*$. Since b^* is the supremum of the critical values over the possible range of π_c , this test guarantees the type I error control regardless of the underlying true response rate for the control arm.

The attained significance level of this test is the exact probability of rejecting the null hypothesis when the underlying control rate equals π_c^* which is given by

$$P_{\pi_c^*}(T(x_t, x_c) > b^* \mid H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\rho_0\pi_c^*, \pi_c^*}(x_t, x_c) \quad (\text{V.77})$$

Note that the attained significance level is the maximum type I error one can actually commit using this test given the desired significance level α . Due to the discreteness of the distributions, the attained significance level is always bounded above by α .

Next we will show how the unconditional power of this exact test is calculated. The unconditional power is the probability of rejecting the null hypothesis under the alternative hypothesis. Suppose that one is interested in the power of this test when $\rho = \rho_1 (> \rho_0)$ and the response rate for the control arm is π_c . Under the alternative $\rho = \rho_1$, we have $\pi_t = \rho_1\pi_c$. Then the unconditional power is given by

$$P_{\pi_c}(T(x_t, x_c) > b^* \mid H_1) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\rho_1\pi_c, \pi_c}(x_t, x_c) \quad (\text{V.78})$$

where

$$f_{\rho_1\pi_c, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} (\rho_1\pi_c)^{x_t} (1 - \rho_1\pi_c)^{n_t - x_t} (\pi_c)^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (\text{V.79})$$

V Theory-Design - Binomial Two-Sample Exact Tests

Non-inferiority for Ratio of Proportions – Case 2 Suppose it is desired to test $H_0 : \frac{\pi_t}{\pi_c} \geq \rho_0 (> 1)$ against the one-sided alternative $H_1 : \frac{\pi_t}{\pi_c} < \rho_0$. In this case, the same test statistic can be used

$$T(x_t, x_c) = \frac{\hat{\pi}_t - \rho_0 \hat{\pi}_c}{\sqrt{\frac{\tilde{\pi}_t(1-\tilde{\pi}_t)}{n_t} + \frac{\rho_0^2 \tilde{\pi}_c(1-\tilde{\pi}_c)}{n_c}}} \quad (\text{V.80})$$

where $\hat{\pi}_t, \hat{\pi}_c, \tilde{\pi}_t$ and $\tilde{\pi}_c$ are defined in the same way as in Case 1.

Let $\mathcal{X} = \{(x_t, x_c) : 0 \leq x_t \leq n_t, 0 \leq x_c \leq n_c\}$ denote the sample space for the 2×2 table. Let $f_{\pi_t, \pi_c}(x_t, x_c)$ denote the probability of observing the data $(x_t, x_c) \in \mathcal{X}$ when the response rates for the treatment and control arms are π_t and π_c , respectively, which is given by

$$f_{\pi_t, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} \pi_t^{x_t} (1 - \pi_t)^{n_t - x_t} \pi_c^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (\text{V.81})$$

For a given π_c and nominal significance level α , let b_{π_c} be defined by

$$\sup \left\{ b : P_{\pi_c} (T(x_t, x_c) < b \mid H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_b(x_t, x_c) f_{\rho_0 \pi_c, \pi_c}(x_t, x_c) \leq \alpha \right\} \quad (\text{V.82})$$

where the indicator function is defined by

$$I_b(x_t, x_c) = \begin{cases} 1 & \text{if } T(x_t, x_c) < b \\ 0 & \text{otherwise} \end{cases} \quad (\text{V.83})$$

Let $b^* = \inf \{b_{\pi_c} : \pi_c \in (0, 1)\}$ and suppose that this infimum takes place at π_c^* . The decision rule of the exact test is to reject H_0 if $T(x_t, x_c) < b^*$. Since b^* is the infimum of the critical values over the possible range of π_c , this test will guarantee the type I error control regardless of the underlying true response rate for the control arm.

The attained significance level of this test is the exact probability of rejecting the null hypothesis when the underlying control rate equals π_c^* which is given by

$$P_{\pi_c^*} (T(x_t, x_c) < b^* \mid H_0) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\rho_0 \pi_c^*, \pi_c^*}(x_t, x_c) \quad (\text{V.84})$$

Note that the attained significance level is the maximum type I error one can actually commit using this test given the desired significance level α . Due to the discreteness of the distributions, the attained significance level is always bounded above by α .

The unconditional power under the specific alternative $\rho = \rho_1$ and π_c is given by

$$P_{\pi_c}(T(x_t, x_c) < b^* | H_1) = \sum_{(x_t, x_c) \in \mathcal{X}} I_{b^*}(x_t, x_c) f_{\rho_1 \pi_c, \pi_c}(x_t, x_c) \quad (\text{V.85})$$

where

$$f_{\rho_1 \pi_c, \pi_c}(x_t, x_c) = \binom{n_t}{x_t} \binom{n_c}{x_c} (\rho_1 \pi_c)^{x_t} (1 - \rho_1 \pi_c)^{n_t - x_t} (\pi_c)^{x_c} (1 - \pi_c)^{n_c - x_c} \quad (\text{V.86})$$

V.4 Power of the Unconditional Test of Equivalence

V.4.1 Power

Equivalence testing usually arises in the context of a clinical trial comparing two treatments in which the goal is to assess whether the two treatments are equally efficacious rather than attempting to assess whether one treatment is more efficacious than the other. This implies an inversion of the conventional formulation of null and alternative hypotheses. The statistical formulation proposed by Dunnett and Gent (1977) is used to describe this procedure. First define the true underlying treatment difference

$$\delta = |\pi_t - \pi_c| \quad (\text{V.87})$$

and specify an equivalence margin, $\delta_0 > 0$, such that if $\delta < \delta_0$ the two treatments are considered equivalent while if $\delta \geq \delta_0$, they are not. Interest resides in testing the null hypothesis

$$H_0: \delta = \delta_0 \quad (\text{V.88})$$

against the alternative hypothesis

$$H_1: \delta < \delta_0 . \quad (\text{V.89})$$

The null hypothesis (V.88) really consists of the two possibilities

$$H_{01}: \pi_c - \pi_t = \delta_0 \quad (\text{V.90})$$

and

$$H_{02}: \pi_t - \pi_c = \delta_0. \quad (\text{V.91})$$

V Theory-Design - Binomial Two-Sample Exact Tests

In order to cater to both possibilities two one-sided level- α tests are performed using the test statistics

$$T_1 = \frac{\hat{\pi}_c - \hat{\pi}_t - \delta_0}{\sqrt{\frac{(\hat{\pi}_c)(1-\hat{\pi}_c)}{n_c} + \frac{(\hat{\pi}_t)(1-\hat{\pi}_t)}{n_t}}} \quad (\text{V.92})$$

and

$$T_2 = \frac{\hat{\pi}_t - \hat{\pi}_c - \delta_0}{\sqrt{\frac{(\hat{\pi}_c)(1-\hat{\pi}_c)}{n_c} + \frac{(\hat{\pi}_t)(1-\hat{\pi}_t)}{n_t}}} . \quad (\text{V.93})$$

Clearly $T_1 \sim N(0, 1)$ conditional on H_{01} and $T_2 \sim N(0, 1)$ conditional on H_{02} . In order to reject the null hypothesis (V.88) and declare equivalence, both H_{01} and H_{02} must be rejected. The rejection region is thus the joint event $\{(T_1 \leq z_\alpha) \cap (T_2 \leq z_\alpha)\}$. It can be shown that under the null hypothesis (V.88), regardless of whether H_{01} or H_{02} holds,

$$\Pr\{(T_1 \leq z_\alpha) \cap (T_2 \leq z_\alpha)\} \leq \alpha \quad (\text{V.94})$$

thereby preserving the type-1 error.

V.4.1 Exact Unconditional Power for Equivalence Tests

Suppose it is desired to obtain the exact power of the two one-sided equivalence test at specific values of π_c and π_t with $|\pi_c - \pi_t| = \delta_1$ where $0 \leq \delta_1 \leq \delta_0$. The exact unconditional power is then readily evaluated as the probability, $\Pr\{(T_1 \leq z_\alpha) \cap (T_2 \leq z_\alpha) | \pi_c, \pi_t\}$, of falling in the rejection region under the alternative hypothesis. Denote this probability by $(1 - \beta)$. Then

$$(1 - \beta) = \sum_{x_c=0}^{n_c} \sum_{x_t=0}^{n_t} I_\alpha(x_c, x_t) \binom{n_c}{x_c} \pi_c^{x_c} (1 - \pi_c)^{n_c - x_c} \binom{n_t}{x_t} \pi_t^{x_t} (1 - \pi_t)^{n_t - x_t} , \quad (\text{V.95})$$

where the indicator function, $I_\alpha(x_c, x_t)$, assumes the value 1 if $(T_c \leq z_\alpha) \cap (T_t \leq z_\alpha) \leq \alpha$ and assumes the value 0 otherwise.

V.5 Sample Size Computations

For all tests discussed in this section, the sample size for a fixed unconditional power value is obtained by evaluating the null and alternative power functions over a range of sample sizes until an N is found that gives the desired power.

Since neither α nor β are guaranteed to be attainable due to discreteness of the binomial distribution, the solution to this parameter space search N is not unique. The choice of sample size for a particular trial should depend on the priorities given to type 1 and 2 errors by the investigator. Possible prioritization may involve (1) primarily

maximizing the attained type 1 error while maintaining it below α , (2) primarily maximizing the attained type 2 error while maintaining it below β , or (3) optimizing to get α^* and β^* as close to α and β , respectively as possible.

The most practical choice of sample size, however, may be that sample size above which power is guaranteed to be at least $(1 - \beta)$.

W *Classification Table*

Under usual notation, the formulas used in computing classification errors are listed below.

$$h_i^1 = \text{Hat Diagonal element assuming } y = 1, G_i = 1;$$

$$h_i^0 = \text{Hat Diagonal element assuming } y = 0, G_i = 1;$$

$$V_i = \mathbf{Cov} \times \mathbf{X}_i'$$

$$\beta_i^0 = \beta - \frac{(1 - \hat{\pi}_i)}{1 - h_i^1} \times V_i$$

$$\hat{\pi}_i^1 = \mathbf{X}_i \beta_i^1$$

$$\beta_i^0 = \beta - \frac{(-\hat{\pi}_i)}{1 - h_i^0} \times V_i$$

$$\hat{\pi}_i^0 = \mathbf{X}_i \beta_i^0$$

$$P(A|\bar{B}) = \frac{1}{n_2} \sum_{i \in C_2} I(\hat{\pi}_i^0 > z)$$

| Name | Formula | Comment |
|-----------------------------|---|---|
| Prob_event | P_e | |
| Cut-off prob | z | |
| Correct events (CE) | $\sum_{i \in C_1} I(\hat{\pi}_i^1 \geq z)$ | |
| Correct noevents(CN) | $\sum_{i \in C_2} I(\hat{\pi}_i^0 \leq z)$ | |
| Incorrect events (IE) | $\sum_{i \in C_2} I(\hat{\pi}_i^0 > z)$ | |
| Incorrect noevents (IN) | $\sum_{i \in C_1} I(\hat{\pi}_i^1 < z)$ | |
| Percent correct | $\frac{(CE+CN)}{(n_1+n_2)}$ | |
| Sensitivty | $\frac{(CE)}{(n_1)}$ | |
| Specifcty | $\frac{(CN)}{(n_2)}$ | |
| False_pos | $\left(\frac{IE}{n_2}\right) \times \frac{1-P_e}{v1}$ | $v1 = \frac{IE}{n_2} + P_e \times \left(\frac{CE}{n_1} - \frac{IE}{n_2}\right)$ |
| False_neg | $\left(1 - \frac{CE}{n_1}\right) \times \frac{P_e}{1-v1}$ | |

X

Glossary

Accrual rate

The number of subjects entering the study per unit of time.

Adaptive study design

In an adaptive design estimated treatment differences at interim analyses can be used to make mid-course data-dependent alterations to the trial design – changes in sample size, error spending function, and number and spacing of interim looks – while preserving the type-1 error.

Alpha spending function

The spending function to be used for allocating the type-1 cumulative error probability as a function of the information fraction.

Alpha spent

The cumulative amount of type-1 error probability spent up to and including a given look.

ASN (Average Sample Number) chart

This plot provides a graphical rendition of how the ASN (Average Sample Number, the expected sample size) varies as a function of a range of possible values for the effect size or non-inferiority margin (e.g. standardized difference, difference in proportions, etc.).

Assigned fraction (treatment)

The proportion, r , of subjects assigned (randomized) to the treatment (experimental) arm over the total number of subjects in the trial.

Beta spending function

The spending function to be used for allocating the type-2 cumulative error probability as a function of the information fraction.

Beta spent

The cumulative amount of type-2 error probability spent up to and including a given look.

X Glossary

Binding boundaries

Binding boundaries require the termination of the trial if the test statistic crosses the futility boundary; otherwise the type-1 error might be inflated. Contrariwise, non-binding boundaries produce the desired power and preserve the type-1 error so that the crossing of the futility boundary may be overruled.

Bioequivalence

A test formulation of a drug (t) and the control (or reference) formulation of the same drug (c) are considered to be bioequivalent if the rate and extent of absorption are similar. The goal is to establish that the difference or log-ratio of the means of the observations from the test formulation and the control is within a specified equivalence margin.

Boundaries

Boundaries are the generalization to group sequential methods of the critical values of a test, the values beyond which the standardized test statistic supplies enough evidence to reject H_0 or H_1 . Boundary families allow the user to specify how conservatively or aggressively tests are performed at each analysis point, while preserving the type-1 error, the probability of accepting H_1 when H_0 is in fact true. Available boundary families are p-value, Haybittle-Peto Power, Wang-Tsiatis Spending Functions, Published Spending Function, and Interpolated.

Boundary chart

This plot provides a graphical rendition of the stopping boundaries ("Nominal critical point") corresponding to each look, the latter being indexed by the cumulative information (e.g. sample size, number of events, etc. depending on the endpoint). For the meaning of the various "Boundary Scales" please refer to other sections of the manual.

Boundary family

The boundaries at the design stage can be derived with reference to one of several approaches, depending also on whether early stopping is allowed in favor of the null only, of the alternative only, or of both. The Haybittle-Peto boundaries (p-value family) are specified in terms of a constant p-value for all interim analyses; East will compute the p-value to be used at the final analysis in order to satisfy the desired significance level of the procedure or the user can specify it and then East computes

the achieved significance level of the procedure. The Wang and Tsiatis (early stopping for H_0 only) and the Pampallona and Tsiatis (early stopping for H_0 or H_1) families are direct application of the respective power boundary families, indexed by a boundary shape parameter, Delta, in the range -0.5 to 0.5: small values of Delta yield boundaries with a small probability of early stopping and a correspondingly low average sample size, vice versa for large values of Delta. Spending Function Boundaries (published) are defined by published error spending functions (e.g. Lan-DeMets, Rho family, Gamma family). Spending Functions Interpolated are defined by the user by specifying cumulative error probabilities at various looks. When interim looks are different from the design, linear interpolation is used for computing cumulative end probabilities spent.

Boundary scale

See Boundary chart.

Boundary shape parameter

The Wang-Tsiatis and the Pampallona-Tsiatis power boundaries are indexed by a shape parameter varying between -0.5 and 0.5. Smaller values of the shape parameter correspond to boundaries with reduced probability of early stopping but also to a smaller maximum sample size, vice-versa for larger values of the shape parameter. For designs allowing for early stopping in favor of either H_0 or H_1 East allows for different shape parameters to govern the boundary for early rejection of H_1 (denoted in the East worksheets as "Boundary shape parameter to reject H_1 ") and the boundary for early rejection of H_0 (denoted in the East worksheets as "Boundary shape parameter to reject H_0 ").

Coefficient of variation

The coefficient of variation is a summary measure of variability. It is calculated by taking the ratio of the standard deviation to the mean

Committed accrual (duration or subjects)

The committed number of subjects that can be accrued into the study (or equivalently, since the accrual rate is constant, the maximum accrual duration). In time to event studies, the power of the study is not determined by the number of subjects enrolled but by the number of events observed. Thus, there exists a range of accrual (bounded by the quantities

X Glossary

Min and Max), combined with a range of study durations, that would all produce the desired power. The lower bound of the range (Minimum committed number of subjects to accrue) corresponds to an initial estimate of the number of events to be observed for the study to have the desired power: with such a low accrual, the study will however be very long since all subjects accrued will have to fail before the final analysis can be performed. On the other hand, there is no need for the study to accrue more than the upper bound of the range (Maximum committed number of subjects to accrual), that is to keep the study open to accrual beyond the point in time when the required number of events has been observed. The user can input values for the accrual within the suggested range remembering that the larger the accrual the shorter the total study duration.

Conditional power

The conditional power is the probability of rejecting the null at one of the future looks given the data accumulated so far. This quantity can contribute, together with any other relevant information, to the decision to terminate or continue the study with a possible increase of the study's sample size.

Conditional power at ideal next look position (CP at INLP)

The conditional power at ideal next look position is the probability of rejecting the null at the next and final look given the data accumulated so far and if the next and final look was performed at the recommended "Ideal next look position". This quantity can contribute, together with any other relevant information, to the decision to terminate or continue the study.

Conditional power chart

This plot provides a graphical rendition of how the conditional power of the study at the current look varies as a function of the effect size (e.g. standardized difference, difference in proportions, etc.).

Confidence interval adjusted

The method suggested by Kim and DeMets (1987) is applied to derive the adjusted confidence interval at the end of the study allowing for repeated significance testing. This method was generalized by Brannath, Mehta and Posch (2008) for the parameter estimation in the adaptive trial.

Crossover ANOVA sqrt(MSE)

In a crossover design trial, the square root of the Mean Squared Error (MSE) from an ANOVA analysis is an estimate of the standard deviation of the error.

Chen-DeMets-Lan (CDL) method

The method for making sample size modifications to an ongoing trial and then performing the interim monitoring and final analysis with the classical Wald statistic. The method is further extended to a more general setting by Gao, Ware and Mehta (2008).

Cui, Hung, and Wang (CHW) method

The CHW method is a procedure for adaptive sample size modification of an on-going two-arm, K -look group sequential clinical trial. It is based on the examination of data at any interim look $L < K$, making a sample size modification if required, and continuing with the interim monitoring, using a modified test statistic that combines the standardized treatment effects before and after the modification as a weighted sum, with appropriate weights so as to preserve the type-1 error.

Cumulative accrual

The cumulative number of subjects accrued up to a given look.

Cumulative events

The cumulative number of events observed up to a given look.

Design proportion

When designing a study to compare binomial proportions, the expected value of the difference between the two groups being compared is expressed in terms of the expected proportion of success in the Treatment and in the Control groups respectively. In non-inferiority studies this difference represents the non-inferiority margin (the treatment arm should not be worse than the control arm by more than the non-inferiority margin). A setting of particular importance for binomial studies is the Casagrande, Pike, and Smith (1978) correction factor for the normal approximation to the binomial. It may be enabled and disabled by checking the appropriate checkbox located in Settings-Binomial. By default this correction is disabled.

X Glossary

Duration-accrual chart

For time to event endpoint. For a range of values of the committed accrual duration (or committed number of subjects) this chart shows the corresponding total study duration, that is the expected time by which the number of events needed to satisfy power considerations will be observed.

Effect size

The Information based module is not sensitive to the actual measurement scale in which the parameter of interest is expressed but only to its magnitude, the Effect Size: its value can express a difference in means or in proportions or even the coefficient from a complex regression model.

Equivalence

An equivalence trial aims to determine if two treatments have similar consequence. It aims to reject the null hypothesis that the difference between the two treatments falls outside the pre-specified lower and upper equivalence boundaries in favor of the alternative hypothesis that the difference between the two treatments falls within these boundaries.

Equivalence limits

In an equivalence trial for the difference of two normal means, the goal is to establish that the treatment mean and control mean are within an equivalence range. This range is delimited by the lower and upper equivalence limits δ_l and δ_u , which need not be equidistant from the value specified for the difference of means under the alternative hypothesis δ_1 .

Equivalence margin (δ_0)

In an equivalence trial, the goal is to establish that the treatment and control parameters are within a specified value δ_0 . This δ_0 value is the equivalence margin and is often defined as a proportion, such as 25% of the control mean for the comparison of the mean normal distributions.

Error spending chart

This plot provides a graphical rendition of the error probability spending functions as functions of the cumulative information fraction.

Events/Accruals vs. Time chart

For time to event endpoints, this chart shows how accrual increases (at a constant rate) until the end of the accrual period and how events will

accumulate on each treatment arm (depending on the corresponding failure rate) as the study progresses in chronological time (horizontal axis).

Expected values under H_0 , H_1 and $H_{1/2}$

The probability to stop the trial at any of the planned looks can be computed under the null (H_0), the alternative (H_1) or the mid-alternative ($H_{1/2}$). These probabilities can be used to compute several expected quantities at study termination. The expected accrual, for instance, can be computed as the sum, over all looks, of the probability of stopping at the given look times the accumulated accrual (sample size) at that look.

Fixed sample study information

In the design worksheet of the General module, one of the needed input parameters is the information (e.g. number of subjects) required for the fixed-sample-study. This quantity can be obtained from any sample size software and on that basis East generates a group sequential study with the same size and power to detect the same alternative. See also Inflation Factor.

Group sequential designs

Group sequential designs allow the investigator to take early interim looks at the data for evidence of efficacy, harm, and/or futility with the aim of possibly stopping the trial early. The planned number of looks describes the number of time points, including the closing date of the study, at which the investigator plans to analyze the thus far collected data. The value 1 corresponds to a classic fixed-sample-size design with a single look at the end of the study when all data have been collected. The planned number of looks K can vary from 1 to 10. The number eventually performed may differ from K .

Hypothesis to be rejected

Early stopping can be allowed for in favor of H_1 only (early stopping with rejection of H_0) or in favor of either H_0 (futility) or H_1 or in favor of H_0 only (futility only).

Ideal next look position

After each look East revises the maximum information (e.g. sample size, number of failures etc. depending on the endpoint) to be achieved for the study to satisfy the desired type-1 and type-2 error probabilities allowing

X Glossary

for the actually adopted schedule of analyses (which may be different from the tentative number and relative spacing assumed at design). This quantity can contribute, together with any other relevant information, to decide when to perform a further analysis of the accumulating data.

Inflation factor

More information (e.g. number of subjects) is required for a group sequential study than for the corresponding fixed-sample study with the same operating characteristics. This is the penalty associated with repeated significance testing. The inflation factor is the proportionality constant (ratio) relating the information requirements of group sequential trial to its corresponding fixed- sample study. This ratio is independent of the test, the endpoint of interest or the actual magnitude of the effect size of interest. East uses this result in the General module to set up a group sequential study on the basis of the information requirements of a fixed-sample study. See also Fixed Sample Study Information.

Information calculator

The calculator applies to parallel two-arm randomized designs with normal or binomial endpoints. During interim monitoring of an information based study the accumulated information up to the current look can be computed on the basis of the current values of the sample size and of the observed sample mean and standard deviation (if the underlying endpoint follows a normal distribution) or number of responses (if the underlying endpoint follows a binomial distribution) in the control and treatment arm respectively. It computes the achieved statistical information, the value of the current test statistic and a new estimate of the maximum sample size required. This latter quantity may differ from the value obtained at design (using the Sample Size Calculator) if the statistical information actually accumulates at a higher or lower pace than anticipated (i.e. if for normal data the actual standard deviation of the observation is different from the value used at design and for binomial data if the observed success rate in the control group is different from the value used at design).

Information fraction

This is defined as the ratio of the information at the current time-point to the maximum information committed to the study. For a large number of studies, including studies with normal and binomial end points the information fraction is simply the ratio of the current sample size to the

maximum sample size committed to the study. For time to event end points it is the ratio of the current number of events (such as failures) to the total number of events committed to the study. For studies in which the monitoring will be performed on the Fisher information scale, the information is estimated as the square inverse of the estimate of the standard error of the parameter under investigation. Thus the information fraction is the ratio of the current inverse square estimate to the maximum inverse square estimate needed to achieve the goals of the study. Information fraction is also referred to sometimes as Process time.

Last look logic

When the trial has to come to an end for administrative reasons (i.e. not because one of the boundaries has been crossed or because the maximum information has been reached) the boundary for this last look should be determined by spending the remaining alpha so as to respect the desired size of the testing procedure. In interim monitoring, this is what happens when the **Tools-Last Look** menu item is selected in East before performing the next look.

Look number

The counter identifying successive analyses of the data.

Maximum accrual

The accrual to be reached if no early stopping occurred (i.e. if the study went on until the last look). This quantity satisfies the desired significance level and power of the design.

Maximum accrual duration

In studies with time to event endpoint, the time required to achieve the necessary maximum accrual.

Maximum events

See Maximum study duration.

Maximum information

The information to be achieved if the study does not stop at any interim analysis. This quantity computed at design is revised during interim monitoring to allow for the actual schedule of looks, since their number

X Glossary

and relative spacing may be different than assumed at design (see Ideal next look position).

Maximum study duration

In studies with time to event endpoint, the study duration and the corresponding number of events of the study to satisfy the desired operating characteristics of the study if no early stopping occurs.

Median survival

When designing a study to compare the distributions of the times to event, the expected relative advantage of the treatments being compared is expressed, by default, in terms of the expected median survival in the treatment and in the control groups respectively. Alternatively, the Design Wizard allows the specification of the relative survival experience in terms of expected percent survival at a specific time or in terms of hazard rates.

Median unbiased estimator (MUE), Adjusted

The method suggested by Kim (1989) is applied to derive the median unbiased estimator of the effect size at the end of the study allowing for repeated significance testing.

Mid-alternative

Studies where early stopping may occur either in favor of the null or of the alternative hypothesis, may extend until relatively large stopping times, if the alternative has been overestimated. In such cases, the test statistic will tend to fluctuate within the continuation region. East computes the expected quantities (e.g. sample size or accrual) at termination not only under the null and the alternative but also under an intermediate hypothesis. Due to the non-linearity of the transformation linking the scale in which the effect size of interest to the user is expressed and the internal standardized scale used by East, the mid-alternative does not correspond to half of the alternative. The expected quantities computed by East under the mid-alternative, however, express the worst case scenarios.

Muller and Schafer method

In adaptive trials, the Muller and Schafer method aims to preserve the conditional type-1 error computed at the time of the adaptation. It is permissible to make any desired data dependent change to an ongoing group sequential trial, possibly more than once, by the simple process of

preserving the conditional type-1 error of the remainder of the trial after each change.

Nominal critical point

A synonym for the boundary value against which the test statistic has to be compared. The nominal critical point is expressed in the same scale as a standard normal deviate in order to facilitate the comparison against the test statistic computed at each look. This explains the use of the adjective "Nominal". See also Test Statistic and Nominal Significance Level.

Nominal significance level

The probability of values more extreme than the Nominal Critical Point according to a standard normal distribution. See also Nominal Critical Point.

Non-binding boundaries

Non-binding boundaries produce the desired power and preserve the type-1 error so that the crossing of the futility boundary may be overruled. Contrariwise, binding boundaries require the termination of the trial if the test statistic crosses the futility boundary; otherwise the type-1 error might be inflated.

Non-inferiority margin

In non-inferiority designs for difference, the non-inferiority margin (δ_0) is the magnitude of the difference between the treatment and the control arm that should not be exceeded for the treatment arm to be considered non-inferior to the control arm. In non-inferiority designs for ratio, the non-inferiority margin (ρ_0) is the ratio between the treatment proportion response and the control proportion response that should not be exceeded for the treatment arm to be considered non-inferior to the control arm. In non-inferiority designs for odds ratio, the non-inferiority margin (Ψ_0) is the odds ratio between the treatment proportion response and the control proportion response that should not be exceeded for the treatment arm to be considered non-inferior to the control arm.

Non-inferiority trial

A non-inferiority trial aims to determine if the outcome of an experimental treatment is no worse than the outcome of the standard treatment. It aims to reject the null hypothesis that the experimental treatment exceeds a

X Glossary

pre-specified non-inferiority margin. The amount by which the mean response on the experimental arm is worse than the mean response on the control arm must fall within this non-inferiority margin for the claim of non-inferiority to be sustained.

Nuisance parameters

Nuisance parameters affect the results of mathematical and statistical models but there may be insufficient information about their magnitudes. In clinical trials inaccurate initial estimates of these parameters will lead to incorrect estimates of the sample size or other resources and the study will not have the correct operating characteristics. Adaptive trials may estimate nuisance parameters based on early results and, then given these more accurate estimates, the conclusions of the trial may be more accurate than those from traditional trials based on poorly estimated nuisance parameters.

Number of looks (K)

For design purposes, K represents the tentative number of analyses to be performed during the interim monitoring phase up to and including the last look. The number of analyses eventually performed during interim monitoring of the trial can be different from K .

Pampallona-Tsiatis boundaries

These power boundaries are characterized by two shape parameters: Δ_1 for the boundaries that facilitate early stopping for efficacy by rejecting H_0 ; and Δ_2 for the boundaries that facilitate early stopping for futility by rejecting H_1 .

Percent survival at Time t

This option in East specifies the survival curves for the control and treatment arms using their percentages surviving at Time t . Given this information East will calculate medians, hazard rates, and hazard ratios.

Post-hoc power

The post-hoc power is an a-posteriori characteristic of the actually adopted sequence of analyses: it is the probability of rejecting the null hypothesis using a testing strategy that corresponds to the analyses performed during the trial, up to and including the final one.

Post-hoc power chart

The post-hoc power is an a-posteriori characteristic of the actually adopted sequence of analyses: when computed after each interim analysis, it is the probability of rejecting the null hypothesis using a testing strategy that corresponds to the analyses performed up to and including the current look plus a hypothetical final analysis. This plot provides a graphical rendition of how the post-hoc power varies as a function of the cumulative information (e.g. sample size, number of failures etc., depending on the endpoint) at this hypothetical last look. Two special cases are worth noting: before the first analysis the post-hoc power curve corresponds to a power curve for a fixed sample study as a function of information rather than of the parameter of interest; after the actual last analysis the post-hoc power reduces to a single number (displayed in the "Post-Hoc Power" output box of the Interim Monitoring worksheet).

Power (1-beta)

The power of the study (or one minus beta, where beta is the type-2 error probability) is the probability of terminating the study with the rejection of the null hypothesis (H_0) when the alternative hypothesis (H_1) is indeed true. Usual choices of power are 0.9 and 0.8 (corresponding to 10% and 20% type-2 error probability, respectively, also known as Beta). Beta is the type-2 error, the probability of not rejecting H_0 when it is in fact false. An underpowered trial is extremely undesirable because it places human subjects at risk with a low probability of reaching a positive scientific conclusion and diverts resources that could be better utilized elsewhere.

Power chart

This plot provides a graphical rendition of how the power of the study varies as a function of the effect size or non-inferiority margin (e.g. standardized difference, difference in proportions, etc.).

p-value, adjusted

The method suggested by Fairbanks and Madsen (1982) is applied to derive the overall adjusted p- value at the end of the study allowing for repeated significance testing.

Repeated confidence interval

The sequence of repeated confidence intervals provided after each look has simultaneous coverage probability of $(1 - \alpha)100\%$. Each interval

X Glossary

provides a statistical summary of the information about the parameter of interest allowing for repeated looks at the accumulating data. This quantity can contribute, together with any other relevant information, to the decision to terminate or continue the study. The coverage probability of the procedure is maintained regardless of how the decision to terminate the study is taken.

Repeated P-value

At the k_{th} analysis, a two-sided repeated P-value for the null hypothesis $H_0 : \delta = \delta_0$ is defined as $p_k = \max(\alpha : \delta_0 \in I_k(\alpha))$, where $I_k(\alpha)$ is the current $(1 - \alpha)$ -level Repeated Confidence Interval (RCI). In other words, P_k is that value of α for which the k_{th} $(1 - \alpha)$ -level RCI contains the null value, δ_0 , as one of its endpoints. The repeated P-value provides protection against the effect due to multiple-looks.

Repeated significance test

The idea of a "repeated significance test" at a constant nominal significance level to analyze accumulating data at a number of times over the course of a study was developed by Pocock. Subject entry is divided into K equally sized groups containing m subjects on each treatment, and the data are analyzed after each new group of observations has been observed.

Sample size calculator

The calculator applies to parallel two-arm randomized designs with normal or binomial endpoints. For such studies it translates information into a sample size when supplied with the value of the nuisance parameter, namely the known and common standard deviation of the observations (if the underlying endpoint follows a normal distribution) or the success rate in the control group (if the underlying endpoint follows a binomial distribution).

Significance level (alpha)

Alpha (or type-1 error probability), is the probability of terminating the study with the rejection of the null hypothesis (H_0) when it is actually true. Usual choices of alpha are 0.05 and 0.10 (corresponding to 5% and 10% type-1 error probability, respectively).

Spacing of looks

Two options are available in East to specify the relative spacing of looks. If "Equal Spacing" is selected, East assumes, at design, that analyses are performed after equal increments of physical resources (e.g. subjects for normal or binomial endpoint, failures for survival) or of statistical information. If "Unequal Spacing" is selected, the user specifies the timing of the analyses in terms of fractions (in the range 0 to 1) of cumulative information. The actual spacing of analyses adopted during the trial can be different from the one tentatively chosen for design purposes.

Spending function

See alpha spending function or beta spending function or the next entry.

Spending Functions, Published (Pub)

These are single-parameter boundary families, the ρ (rho) or γ (gamma). $\rho = 1$ produces boundaries that resemble the Pocock; $\rho = 3$ produces boundaries that resemble the more conservative O'Brien-Fleming. When γ is negative its convex spending functions increase in conservatism as γ decreases; when γ is positive its concave spending functions increase in aggressiveness as γ increases. When $\gamma = 0$ the type-1 error is spent linearly. When $\gamma = 1$ the stopping boundaries resemble the Pocock.

Standardized difference

When designing a study to compare means of normally distributed observation the expected value of the difference between the means of the two groups being compared divided by the (common and assumed known) standard deviation of the observations is of relevance. This quantity is referred to as the standardized difference. In non-inferiority studies this difference represents the non-inferiority margin (the treatment arm should not be worse than the control arm by more than the non-inferiority margin). It can also be expressed as a function of its individual components (the two means and the common standard deviation), or of the difference in means and the standard deviation.

Stopping probabilities

The probability that the test statistic will cross a stopping boundary at a given look. These probabilities are different depending on which hypothesis is assumed to hold (for instance under the null, the alternative or an intermediate hypothesis).

Study duration

X Glossary

In studies with time to event endpoint, the study duration up to and including a given look (actual chronological time of the analysis relative to study start) computed under various hypotheses.

Superiority trial

A superiority trial aims to determine if the outcome of an experimental treatment is better than the outcome of the standard treatment. It aims to reject the null hypothesis that there is no difference between these two outcomes.

Test statistic

In any of the Interim Monitoring worksheets and in the Direct Monitoring worksheet the user is requested to input the value of the test statistic observed at the current analysis. This corresponds to the usual deviate, following a standard normal distribution under the null, as provided by statistical analysis packages. See also Nominal Critical Point.

Test statistic calculator

When at any of the interim analyses the value of the effect size of interest (delta) is known as well as its estimated standard error, the calculator computes the corresponding value of the Wald test statistic. The supplied values of delta and its estimated standard are then used to compute the repeated confidence interval at the given look instead of the design values.

Test type

The type of the test can be either one- or two-sided. A one-sided test assumes that under the alternative hypothesis the parameter of interest lies in a single direction away from the null hypothesis H_0 . A two-sided test assumes that under the alternative hypothesis the parameter of interest lies in either direction away from the null hypothesis H_0 , and the test searches in both directions for departures of the test statistic from H_0 .

Time of looks

The time at which the analyses are performed, in terms of the cumulative fraction of the maximum information (in the range 0 to 1). In particular, for Normal and Binomial endpoints the maximum information is given by the maximum accrual. For Survival type of data, it is given by the maximum number of events.

Traditional study designs

Preference for an experimental treatment can be demonstrated in terms of its improved efficacy with respect to control (a superiority trial), its equivalence to the control treatment (an equivalence trial), or its being not much worse than the control treatment (a non-inferiority trial). In an equivalence trial the goal is to establish equivalence between two treatments rather than the superiority in efficacy of one over the other. In a non-inferiority trial, the experimental treatment should be demonstrated not to be inferior by more than a tolerable non-inferiority margin.

Type of trial

Preference for an experimental treatment can be demonstrated in terms of its improved efficacy with respect to control (“Superiority” trial), its equivalence to the control treatment (“Equivalence” trial), or its being not much worse than the control treatment (“Non-inferiority” trial). In an equivalence trial, the goal is to establish equivalence between two treatments rather than the superiority in efficacy of one over the other. In a non-inferiority trial, the experimental treatment should be demonstrated not to be inferior by more than a tolerable non-inferiority margin.

Type-1 error

The type-1 error probability is the probability of selecting the alternative hypothesis H_1 when the null hypothesis H_0 is in fact true. The significance level α (alpha) quantifies the strength of the evidence against the null hypothesis $H_0 : \mu = \mu_0$. An $\alpha = .05$ implies that the test of significance would erroneously reject the null hypothesis when in fact it was true only five times in 100 tests (1 time in 20). Commonly used significance levels are: .05, .01 (1 time in 100), .025 (25 times in 1000) or .1 (1 time in 10).

Type-2 error

The type-2 error probability (β) (beta) is the probability of erroneously accepting the null hypothesis H_0 when H_1 is in fact true. Commonly used values of (β) are .10 and .20. The power of the test is defined as $1 - (\beta)$. It is the probability of correctly rejecting H_0 (the null hypothesis) when H_1 (the alternative hypothesis) is in fact true.

Wang-Tsiatis boundaries

The Wang-Tsiatis boundaries permit early stopping to reject H_0 . They are

X Glossary

used to stop a trial early for efficacy only (1-sided boundaries), safety only (1-sided boundaries), or to stop early either for efficacy or safety (two-sided boundaries).

Y *On validating the East Software*

Y.1 Group Sequential and Adaptive Designs

Y.1.1 East 6.4 Validation

Y.1.2 East 6.3 Validation

Y.1.3 East 6.2 Validation

*Y.1.4 East 6.0 and 6.1
Validation*

Y.1.5 East 5.4 Validation

Y.1.6 East 5.3 Validation

*Y.1.7 East 5 and East 4
Validation*

Y.1.8 East 3 Validation

Y.1.1 East 6.4 Validation

This section describes the extensive validating procedures carried out on all the features incorporated in East 6.4. East 6.4 will be referred to as East in this subsection. A summary table displaying the methods used for each statistical procedure is given below. Each row of the table corresponds to a statistical procedure and the columns C1-C8 correspond to the following methods:

- **C1 column: Validation using East5.4** - Most of the features which are implemented in East can be validated using the earlier version of East, version 5. Results from such features are compared and validated against East 5 and their consistency is ensured.
- **C2 column: Validation using in-house R codes** - We have developed and are using independent R scripts to validate results from East. These R codes, in some cases, can be used to validate the intermediate output quantities whereas in some cases to validate the complete feature.
- **C3 column: Validation using published R packages** - Some features in East are partially or completely available in published R packages. The results from such features are compared and validated against the results from these R packages.
- **C4 column: Validation using SAS** - Some features in East are partially or completely available in SAS. The results from such features are compared and validated against the results from these SAS procedures.
- **C5 column: Validation using SiZ 2.0** - Most of the features in East which related to Single look design come from SiZ 2.0 version. Results from such features are compared and validated against SiZ 2.0 and their consistency is ensured. SiZ 2.0 is fully validated released software. It has been thoroughly validated against external software like nQuery, PASS, SAS and R as well as with in-house validation programs in R/SAS.
- **C6 column: Using East for Internal Validation and Consistency** - All the features in East are validated by applying some internal consistency checks. These checks are generally carried out using different features within East.
- **C7 column: Validation using StatXact10** - Most of the features in East which related to Single look design come from StatXact 11 version. Results from such features are compared and validated against StatXact11.
- **C8 column: Validation using commercial software packages** - Features that are available in other commercial packages like nQuery, PASS and SAS have

Y On validating the East Software

been validated against East.

| N | East Feature | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 |
|----|---|----|----|----|----|----|----|----|----|
| 1 | Design-MCP for Survival Endpoint | - | 1 | - | - | - | 1 | - | - |
| 2 | Design-MEP for Discrete Endpoint | - | 1 | 1 | - | - | 1 | - | - |
| 3 | Analysis-MEP for Discrete, Continuous Endpoint | - | 1 | 1 | - | - | 1 | - | - |
| 4 | Analysis-MCP for Survival Endpoint | - | 1 | 1 | - | - | 1 | - | - |
| 5 | Assurance and Bayesian predictive power for Survival Endpoint | - | 1 | - | - | - | 1 | - | - |
| 6 | Dose Escalation Designs | - | 1 | 1 | - | - | 1 | - | - |
| 7 | Multi-arm Two-stage Designs based on p-value combination | - | 1 | 1 | - | - | 1 | - | - |
| 8 | MAMS for Continuous Endpoint | - | 1 | 1 | - | - | 1 | - | - |
| 9 | Predict Procedures | - | 1 | - | - | - | 1 | - | - |
| 10 | IM using Muller-Schafer Method | - | 1 | - | - | - | 1 | - | - |

Y.1.2 East 6.3 Validation

This section describes the extensive validating procedures carried out on all the features incorporated in East 6.3. East 6.3 will be referred to as East in this subsection. A summary table displaying the methods used for each statistical procedure is given below. Each row of the table corresponds to a statistical procedure and the columns C1-C8 correspond to the following methods:

- **C1 column: Validation using East5.4** - Most of the features which are implemented in East can be validated using the earlier version of East, version 5. Results from such features are compared and validated against East 5 and their consistency is ensured.
- **C2 column: Validation using in-house R codes** - We have developed and are using independent R scripts to validate results from East. These R codes, in some cases, can be used to validate the intermediate output quantities whereas in some cases to validate the complete feature.
- **C3 column: Validation using published R packages** - Some features in East are partially or completely available in published R packages. The results from such features are compared and validated against the results from these R packages.
- **C4 column: Validation using SAS** - Some features in East are partially or completely available in SAS. The results from such features are compared and validated against the results from these SAS procedures.
- **C5 column: Validation using SiZ 2.0** - Most of the features in East which related to Single look design come from SiZ 2.0 version. Results from such features are compared and validated against SiZ 2.0 and their consistency is

ensured. SiZ 2.0 is fully validated released software. It has been thoroughly validated against external software like nQuery, PASS, SAS and R as well as with in-house validation programs in R/SAS.

- **C6 column: Using East for Internal Validation and Consistency** - All the features in East are validated by applying some internal consistency checks. These checks are generally carried out using different features within East.
- **C7 column: Validation using StatXact10** - Most of the features in East which related to Single look design come from StatXact 10.1 version. Results from such features are compared and validated against StatXact10.1.
- **C8 column: Validation using commercial software packages** - Features that are available in other commercial packages like nQuery, PASS and SAS have been validated against East.

In the table below, the symbol "1" indicates that the method in that column was used for validation of the feature in corresponding row. The symbol "-" indicates that the method in that column was not applicable for that feature.

Y On validating the East Software

| N | East Feature | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 |
|-----|--|----|----|----|----|----|----|----|----|
| 1 | Fixed Sample Tests | | | | | | | | |
| 1.1 | Exact Design Module | 1 | – | – | – | 1 | 1 | 1 | 1 |
| 1.2 | Exact Analysis Module | – | 1 | – | – | – | 1 | 1 | – |
| 2 | Group Sequential Exact Probability Computation | 1 | – | – | – | – | 1 | – | – |
| 3 | Exact Adjusted Confidence Interval | 1 | – | – | – | – | 1 | – | – |
| 4 | Exact Conditional Power | 1 | – | – | – | – | 1 | – | – |
| 5 | Simon’s Two Stage Design | 1 | 1 | 1 | – | – | – | – | 1 |
| 6 | Dose Escalation Designs | – | 1 | 1 | – | – | 1 | – | – |
| 7 | Conditional Simulations | – | 1 | – | – | – | 1 | – | – |
| 8 | Site Info Simulations | – | 1 | – | – | – | 1 | – | – |
| 9 | Parallel Gatekeeping for Multiple Endpoints | – | 1 | 1 | – | – | 1 | – | – |
| 10 | Muller-Schafer for SSR | 1 | 1 | – | – | – | 1 | – | – |
| 11 | SSR for Ratio of Proportions | – | 1 | – | – | – | 1 | – | – |
| 12 | Predicted Interval Plots | – | 1 | – | – | – | 1 | – | – |
| 13 | Exact Inference Adaptive (BWCI) | 1 | 1 | – | – | – | 1 | – | – |
| 14 | Exact Inference Adaptive (RCI) | 1 | 1 | – | – | – | 1 | – | – |
| 15 | Arbitrary Weights CHW | – | 1 | – | – | – | 1 | – | – |
| 16 | Sample Size / Information Calculator | 1 | – | – | – | – | 1 | – | – |

Y.1.3 East 6.2 Validation

This section describes the extensive validating procedures carried out on all the features incorporated in East 6.2. East 6.3 will be referred to as East in this subsection. A summary table displaying the methods used for each statistical procedure is given below. Each row of the table corresponds to a statistical procedure and the columns C1-C5 correspond to the following methods:

- **C1 column: Validation using East5.4** - Most of the features which are implemented in East can be validated using the earlier version of East, version 5. Results from such features are compared and validated against East 5 and their consistency is ensured.
- **C2 column: Validation using in-house R codes** - We have developed and are using independent R scripts to validate results from East. These R codes, in some cases, can be used to validate the intermediate output quantities whereas in some cases to validate the complete feature.
- **C3 column: Validation using published R packages** - Some features in East are partially or completely available in published R packages. The results from such features are compared and validated against the results from these R packages.
- **C4 column: Using East for Internal Validation and Consistency** - All the features in East are validated by applying some internal consistency checks. These checks are generally carried out using different features within East.
- **C5 column: Validation using commercial software packages** - Features that are available in other commercial packages like nQuery, PASS and SAS have been validated against East.

In the table below, the symbol "1" indicates that the method in that column was used for validation of the feature in corresponding row. The symbol "-" indicates that the method in that column was not applicable for that feature.

| N | East Feature | C1 | C2 | C3 | C4 | C5 |
|---|---|----|----|----|----|----|
| 1 | Count Data Designs (Poisson / Negative Binomial) | - | 1 | - | - | 1 |
| 2 | Serial Gatekeeping for Multiple Endpoints | - | 1 | - | - | - |
| 3 | CI-based Designs | - | 1 | 1 | 1 | - |
| 4 | Kaplan-Meier Plots | - | 1 | - | - | 1 |
| 5 | CHW / CDL Methods for SSR | 1 | 1 | - | 1 | - |

Y On validating the East Software

Y.1.4 East Architect and East 6.1 Validation

This section describes the extensive validating procedures carried out on all the features incorporated in East Architect as well as East 6.1. East Architect and East 6.1 will be referred to as East in this subsection. A summary table displaying the methods used for each statistical procedure is given below. Each row of the table corresponds to a statistical procedure and the columns **C1–C6** correspond to the following methods:

- **C1 column: Validation using East5.4** - Most of the features which are implemented in East can be validated using the earlier version of East, version 5. Results from such features are compared and validated against East 5 and their consistency is ensured.
- **C2 column: Validation using in-house R codes** - We have developed and are using independent R scripts to validate results from East. These R codes, in some cases, can be used to validate the intermediate output quantities whereas in some cases to validate the complete feature.
- **C3 column: Validation using published R packages** - Some features in East are partially or completely available in some of published R packages. The results from such features are compared and validated against the results from these R packages.
- **C4 column: Validation using SAS** - Some features in East are partially or completely available in SAS. The results from such features are compared and validated against the results from these SAS procedures.
- **C5 column: Validation using SiZ 2.0** - Most of the features in East which related to Single look design come from SiZ 2.0 version. Results from such features are compared and validated against SiZ 2.0 and their consistency is ensured. SiZ 2.0 is fully validated released software. It has been thoroughly validated against external software like nQuery, PASS, SAS and R as well as with in-house validation programs in R/SAS.
- **C6 column: Using East for Internal Validation and Consistency** - All the features in East are validated by applying some internal consistency checks. These checks are generally carried out using different features within East.

In the table below, the symbol "1" indicates that the method in that column was used for validation of the feature in corresponding row. The symbol "-" indicates that the method in that column was not applicable for that feature.

| N | East Feature | C1 | C2 | C3 | C4 | C5 | C6 |
|-----|---|----|----|----|----|----|----|
| 1 | Response Lag, Accrual, and Dropouts for Continuous and Discrete Endpoints | – | 1 | – | – | – | 1 |
| 2 | Predictive Power | – | 1 | – | – | – | 1 |
| 3 | Fixed Sample Tests | | | | | | |
| 3.1 | Design Module | – | – | – | – | 1 | 1 |
| 3.2 | Simulation Module | – | – | – | – | – | 1 |
| 3.3 | Analysis Module | – | – | – | – | 1 | 1 |
| 4 | Multi-Arm Tests | | | | | | |
| 4.1 | Design Module | – | 1 | 1 | – | – | 1 |
| 4.2 | Analysis Module | – | 1 | – | – | 1 | 1 |
| 5 | Group Sequential Probability Computation | 1 | 1 | 1 | – | – | 1 |
| 6 | Rounded Sample Size | – | 1 | – | – | – | 1 |
| 7 | Flexibility in Setting up Boundaries | | | | | | |
| 7.1 | Efficacy and Futility Missing Boundaries | 1 | 1 | 1 | – | – | 1 |
| 7.2 | (Standardized) Treatment Scale Futility Boundary | – | 1 | – | – | – | 1 |
| 7.3 | Conditional Power Scale for Futility Boundary | – | 1 | – | – | – | 1 |
| 8 | Haybittle-Peto (p-value Scale) Boundary Computation | 1 | 1 | – | – | – | 1 |
| 9 | Adjusted Confidence Interval (ACI) | 1 | 1 | – | 1 | – | 1 |
| 10 | Conditional Power (CP) | 1 | 1 | – | – | – | 1 |
| | East 6.1 Features | | | | | | |
| 11 | Stratified Simulations | 1 | 1 | 1 | 1 | 1 | 1 |
| 12 | Assurance (Probability of Success) | 0 | 1 | 0 | 0 | 0 | 1 |
| 13 | Bayesian Predictive Power | 0 | 1 | 0 | 0 | 0 | 1 |

Y On validating the East Software

Y.1.5 East 5.4 Validation

This section describes the extensive validating procedures carried out on adaptive features incorporated in East5.4. A summary table displaying the methods used for each statistical procedure is given below. Each row of the table corresponds to a statistical procedure and the columns **C1–C4** correspond to the following methods:

- **C1 column: Using East for Internal Validation and Consistency**- In case of adaptive simulations, the final outcome is the 'Re-estimated Sample Size' and the 'Achieved Conditional Power for that sample size. To validate these two numbers we can use intermediate parameters like 'Estimate of Delta', Standard Error of that estimate, the sample size at the adapt look in East designs. The output from CHW IM like repeated p-value is also verified using the Design level features in East.
- **C2 column: Use of R code** - We have developed and are using independent R scripts to validate results from adaptive features like CHW, CDL Simulations and CHW IM. In case of simulations this code works to compute the re-estimated sample size and the power achieved. In case of CHW IM, the R code computes Weighted statistics, the RCI's, and the repeated p-values.
- **C3 column: Use of Excel Based Tools** - We have developed in-house Excel based tools to validate the results obtained from adaptive features. These tools also require information on the adapt look parameters like 'Delta Estimate', 'Standard Error' of that estimate. The outcomes validated are the re-estimated sample size and the conditional power achieved.
- **C4 column: Use of Excel Based Tools** - Using Excel based tool (Developed and recommended By Dr. Cyrus Mehta) to verify the alpha and Power preservation from adaptive simulations. We can run the simulations under the Null/Alternative hypothesis and verify whether the Type-I Error/Design Power is indeed preserved or not. On running 100000 or more simulations, accuracy is achieved. To verify whether the simulated rejection probability is actually close to the Design Alpha or Power, we use the excel based tool which gives us the confidence of preservation of probabilities. This tool in general can be used to verify whether the observed number in (0,1) is close to the actual number or not.

In the table below, the symbol "1" indicates that the method in that column was used for validation of the feature in corresponding row. The symbol "-" indicates that the method in that column was not applicable for that feature.

All the features in the table below are validated for the two tests under **Survival Endpoint: Superiority Trial Two sample Given Accrual Duration and Accrual Rates and Superiority Trial Two sample Given Accrual Duration and Study**

Duration

| N | East 5.4 Feature | C1 | C2 | C3 | C4 |
|---|------------------|----|----|----|----|
| 1 | CHW Simulations | 1 | 1 | 1 | 1 |
| 2 | CDL Simulations | 1 | 1 | 1 | 1 |
| 3 | CHW IM | 1 | 1 | – | 1 |
| 4 | CP Calculator | 1 | 1 | 1 | 1 |

Y On validating the East Software

Y.1.6 East 5.3 Validation

This section describes the extensive validating procedures carried out on adaptive features incorporated in East5.3. A summary table displaying the methods used for each statistical procedure is given below. Each row of the table corresponds to a statistical procedure and the columns **C1–C5** correspond to the following methods:

- **C1 column: Using East for Internal Validation and Consistency** - In case of adaptive simulations, the final outcome consists of 'Re-estimated Sample Size' and the 'Achieved Conditional Power' for that sample size. To validate these two numbers, we use intermediate parameters like 'Estimate of Delta', 'Standard Error' of that estimate and the sample size at the adapt look. Output quantities like weighted test statistic and repeated p value from CHW IM sheet are also verified using internal validation.
- **C2 column: Use of R code** - We have developed independent R scripts to validate results from adaptive features like CHW and CDL Simulations as well as CHW IM sheet. In case of afore mentioned Simulations this code works to compute the re-estimated sample size and the power achieved. In case of CHW IM, it computes Weighted statistics, Repeated Confidence Intervals, and repeated p-values. We have utilized R-packages like 'lbound', 'Adapt'.
- **C3 column: Use of Excel Based Tools** - We have developed in-house Excel based tools to validate the results obtained from adaptive simulations. These tools also require information on the adapt look parameters like 'Delta Estimate', 'Standard Error' of that estimate. The outcomes validated are the re-estimated sample size and the conditional power achieved.
- **C4 column: Use of ADDPLAN** - We have compared results from CHW IM sheet and CP calculator with ADDPLAN.
- **C5 column: Confidence Interval for Probabilities using Excel** - We have used in-house Excel based tool recommended by Dr. Cyrus Mehta to verify the Alpha and Power Preservation from adaptive simulations. This tool provides confidence interval for simulated probability.

In the table below, the symbol "1" indicates that the method in that column was used for validation of the feature in corresponding row. The symbol "-" indicates that the method in that column was not applicable for that feature.

All the features in the table below are validated for **Normal Endpoint: Superiority Trial Two sample Difference of Means** and **Binomial Endpoint: Superiority Trial Two sample Difference of Proportions**.

| Serial No. | East 5.3 Feature | C1 | C2 | C3 | C4 | C5 |
|------------|--------------------------------|----|----|----|----|----|
| 1 | CHW Simulations | 1 | 1 | 1 | – | 1 |
| 2 | CDL Simulations | 1 | 1 | 1 | – | 1 |
| 3 | MS Simulations | 1 | – | – | – | 1 |
| 4 | MS-RCI Estimations | 1 | – | – | – | 1 |
| 5 | MS-SWACI Estimations | 1 | – | – | – | 1 |
| 6 | MS-RCI Estimation Calculator | 1 | – | – | – | 1 |
| 7 | MS-SWACI Estimation Calculator | 1 | – | – | – | 1 |
| 8 | CHW IM | 1 | 1 | – | 1 | 1 |
| 9 | CP Calculator | 1 | – | – | 1 | – |

Y On validating the East Software

Y.1.7 East 5 and East 4 Validation

This manual discusses more than one hundred illustrative trial designs with simulation and interim monitoring. We used these designs to validate the internal and external consistencies of East. A summary table displaying the methods used for each statistical procedure is given below. Each row of the table corresponds to a statistical procedure and the columns **C1–C4** correspond to the following comparisons:

- **C1 column:** Comparisons of the sample sizes for single look designs obtained from East 5 with the analogous estimates from the nQuery(2005) and Egret Siz (1997) software. For the repeated measures design that is not supported by these software packages, we compared the estimates obtained from East 5 with the results reported by Fitzmaurice, Laird and Ware (2004).
- **C2 column:** Comparisons of the design values of significance level and power with the values obtained by simulation in a single look setting.
- **C3 column:** Comparisons of the design values of the probabilities of crossing the stopping boundaries, significance level and power with the values obtained in the simulation in a multiple-look setting.
- **C4 column:** Comparisons of the design boundary values with the boundary value estimates generated in the internal monitoring (IM) module.

In the table, the symbol “1” indicates that the comparison was made for the test and the symbol “-” denotes that a comparable test in other software was not available or the comparison was not applicable (e.g. a check of the boundary crossing probabilities for the East procedures that only support a single look design).

| N | Test Type | Setting | Test Name | C1 | C2 | C3 | C4 |
|----|-----------------|-------------|---------------------------------|----|----|----|----|
| | Normal | | | | | | |
| 1 | Superiority | One Sample | Single Mean | 1 | 1 | 1 | 1 |
| 2 | Superiority | One Sample | Paired Means | 1 | 1 | 1 | 1 |
| 3 | Superiority | One Sample | t-Test | 1 | 1 | - | - |
| 4 | Superiority | One Sample | Paired t-Test | 1 | 1 | - | - |
| 5 | Superiority | Two Samples | Difference of Means | 1 | 1 | 1 | 1 |
| 6 | Superiority | Two Samples | Difference of Means (t-Test) | 1 | 1 | - | - |
| 7 | Superiority | Regression | Single Slope | 1 | - | - | - |
| 8 | Superiority | Regression | Two Slopes | 1 | - | - | - |
| 9 | Superiority | Regression | Repeated Measures | 1 | - | - | - |
| 10 | Non-inferiority | Two Samples | Difference of Means | 1 | 1 | 1 | 1 |
| 11 | Non-inferiority | Two Samples | Difference of Means (t-test) | 1 | 1 | - | - |
| 12 | Equivalence | Two Samples | Difference of Means | 1 | 1 | - | - |
| 13 | Equivalence | Two Samples | Log-ratio of Means | 1 | 1 | - | - |
| 14 | Equivalence | Two Samples | Difference of Means (Crossover) | 1 | - | - | - |
| 15 | Equivalence | Two Samples | Log-ratio of Means (Crossover) | 1 | - | - | - |
| | Binomial | | | | | | |
| 16 | Superiority | One Sample | Single Proportion | 1 | 1 | 1 | 1 |
| 17 | Superiority | One Sample | Matched Pairs | - | 1 | 1 | 1 |
| 18 | Superiority | Two Samples | Difference of Proportions | 1 | 1 | 1 | 1 |
| 19 | Superiority | Two Samples | Ratio of Proportions | - | 1 | 1 | 1 |
| 20 | Superiority | Two Samples | Odds ratio of Proportions | - | 1 | 1 | 1 |
| 21 | Superiority | Two Samples | Stratified 2x2 Tables | - | 1 | 1 | 1 |
| 22 | Superiority | Two Samples | Fisher Exact Test | 1 | - | - | - |
| 23 | Superiority | > 2 Samples | Trend in K Ordered Proportions | 1 | - | - | - |
| 24 | Superiority | Regression | Logistic Regression | 1 | - | - | 1 |
| 25 | Non-inferiority | Two Samples | Difference of Proportions | 1 | 1 | 1 | 1 |

Y On validating the East Software

| N | Test Type | Setting | Test Name | C1 | C2 | C3 | C4 |
|----|----------------------|-------------|---|----|----|----|----|
| | Binomial | | | | | | |
| 26 | Non-inferiority | Two Samples | Ratio of Proportions (Wald) | – | 1 | 1 | 1 |
| 27 | Non-inferiority | Two Samples | Ratio of Proportions (Farrington and Manning) | – | 1 | 1 | 1 |
| 28 | Non-inferiority | Two Samples | Odds Ratio of Proportions | – | 1 | 1 | 1 |
| 29 | Equivalence | Two Samples | Equivalence | 1 | 1 | – | – |
| | Survival | | | | | | |
| 30 | Superiority | Two Samples | Logrank test | 1 | 1 | 1 | 1 |
| 31 | Superiority | Two Samples | Logrank test (Advanced Version) | 1 | 1 | 1 | 1 |
| 32 | Superiority | Regression | Cox Proportional Hazard | 1 | – | – | 1 |
| 33 | Non-inferiority | Two Samples | Logrank | 1 | 1 | 1 | 1 |
| 34 | Non-inferiority | Two Samples | Logrank (Advanced Version) | – | 1 | 1 | 1 |
| | General | | | | | | |
| 35 | Superiority | Two Samples | Convert Single to Multi look | – | 1 | 1 | 1 |
| | Information | | | | | | |
| 36 | Superiority | Two Samples | Design and monitor Maximum Information Trials | – | 1 | 1 | 1 |
| | Nonparametric | | | | | | |
| 37 | Superiority | Two Samples | Wilcoxon, Mann and Whitney | 1 | – | – | – |
| 38 | Superiority | Two Samples | Wilcoxon Rank Sum | 1 | – | – | – |

Y.1.8 East 3 Validation

The statistical results computed by East 3 have been subjected to rigorous and extensive quality-assurance testing for purposes of validation. A database consisting of a large number of studies has been compiled at Cytel Software Corporation. These studies have been gathered from published articles, from East-2000 software and from East-2000 beta testers. Several additional studies have been constructed by us since the release of East-2000. We have also constructed studies using the University of Wisconsin software package. We have thereby tested the software across a broad range of possible input values. The results were checked by five different methods.

1. **Checks against East-2000 and East-DOS.** The results in East 3 have been checked against East-2000, which in turn was tested against East-DOS. The East-2000 and East-DOS software were collectively tested extensively over a period of ten years both in-house and by end-users at commercial sites, academic sites and the FDA.
2. **Checks against Published Tables.** East 3 implements the family of power boundaries proposed by Wang and Tsiatis (1987) and further extended by Pampallona and Tsiatis (1994). Both papers contain extensive tabulations of the constants defining the boundaries and of expected sample numbers for numerous combinations of the various design parameters. East 3 also uses the spending function approach for generating stopping boundaries at the design stage. Tables of boundaries and inflation factors derived from published spending functions are available and have been published by Jennison and Turnbull (2000). We have verified that the numbers in these tables match corresponding numbers generated by East 3.
3. **Checks against Simulation.** The East 3 simulation module provides a further way to check some properties of the designs proposed by East 3 up to Monte Carlo accuracy. For any given set of boundaries, the several different quantities have been checked against the theoretical operating characteristics of any chosen design, such as type-I and type-II error probabilities, stopping probabilities and average sample number. Specifically we have verified the following through simulation:
 - (a) We have simulated studies with varying values for the effect size, ranging all the way from the null hypothesis up to the alternative hypothesis. In every case we have verified that the theoretical power obtained from the design module of East 3 matches with the power obtained by simulation.
 - (b) We have compared the exit probabilities, look by look, between the simulation results and the theoretical results obtained from the design details module of East 3. The exit probabilities match, up to Monte Carlo accuracy.

Y On validating the East Software

- (c) We have compared the average sample size obtained by simulation with the corresponding average sample size displayed on the design worksheet for H_0 , H_1 and $H_{1/2}$. The results match.
4. **Logical Checks.** Several logical checks have been implemented where the behavior of East 3 can either be predicted with certainty or where a high level of consistency is expected among varying but related situations. Some examples are given below:
- (a) East 3 has been extensively tested against published tables and commercial software for fixed-sample size designs. The fixed sample designs in East 3 are special cases of the group sequential designs for which East 3 was primarily developed.
 - (b) We have designed many studies with a variety of spending functions and with both equal and unequal spacings for the interim looks. We have then invoked the interim monitoring module in East 3 and implemented the monitoring schedule exactly as prescribed in the design stage. We have thereby verified through two independent computation procedures that the error spent, and stopping boundaries produced at the interim monitoring stage are identical to the corresponding values at the design stage.
 - (c) We have documented (in Appendix C) that the stopping boundaries used at the interim monitoring stage of a Wang-Tsiatis or Pampallona-Tsiatis design are derived from inverting ten-look, equally spaced stopping boundaries, generated at the design stage. The design and interim monitoring output have therefore been compared for 10-look designs that were actually monitored with 10 equally spaced looks. The results from these two independent methods of obtaining the output match.
 - (d) In the interim monitoring module, before the first look is performed, the conditional power chart corresponds to the usual power curve for fixed sample designs. This serves to validate that the power specified in the design module matches the initial estimate of conditional power.
 - (e) In the interim monitoring module, the suggested optimal look position before any data have been entered into the worksheet must correspond to the sample size requirements of a fixed sample design. We have verified that this theoretical requirement is satisfied.
 - (f) The General module can set up and allow monitoring of a group sequential design on the basis of the sample size requirement of the corresponding fixed sample design. Therefore, for any arbitrary group sequential design set up and monitored in either of the Normal, Binomial or Survival modules it is possible to replicate virtually all the output with the General module given the sample size requirement of its fixed sample counterpart.

- (g) A number of actual clinical applications published in the literature were replicated in East 3. The East 3 results were consistent with the published results. Many of these applications were used as case studies in the earlier East-2000 software.
- (h) The exit probabilities under either H_0 , H_1 or $H_{1/2}$ are displayed in the design details worksheet. We have verified that the sum of these exit probabilities, for any of the above hypotheses, is 1.
- (i) We have verified that the expected sample sizes under H_0 , H_1 or $H_{1/2}$, as displayed on the design worksheet, match with the corresponding expected samples sizes computed directly from the exit probabilities and cumulative accruals, displayed as design details in East 3.
- (j) We have verified that the cumulative alpha spent matches with the cumulative exit probabilities under H_0 from the design details portion of East 3.
- (k) We have verified that for studies with H_0 -only boundaries, the cumulative alpha spent at any intermediate look matches the cumulative exit probability under H_0 , up to that intermediate look.
- (l) We have verified that for 1-sided studies with H_1 -only boundaries, the cumulative beta spent at any intermediate look matches the cumulative exit probability under H_1 , up to that intermediate look.
- (m) We have verified that there is internal consistency between the final adjusted confidence intervals, computed by the Tsiatis, Rosner and Mehta (1989) stage-wise method, and the final adjusted p-value. That is, the final adjusted confidence interval excludes the parameter of interest if and only if the final stopping boundary is crossed, and the final adjusted p-value is less than alpha.
- (n) We have verified that there is internal consistency between the repeated confidence intervals of Jennison and Turnbull (1998) and the value of the final test statistic. That is, one extreme of the repeated confidence interval will coincide with zero for superiority trials (or with the non-inferiority margin for non-inferiority trials) if and only if the observed test statistic falls on a boundary value.
- (o) We have verified that there is internal consistency between the final adjusted p-value and the final cumulative alpha that was spent when the test statistic coincides with the stopping boundary. These two values are computed independently but logically they have to be equal.
- (p) We have verified that the maximum information obtained from the information based design module of East 3 corresponds to the maximum

Y On validating the East Software

sample size obtained from the normal or binomial design modules, for studies in which the effect size, power, type-1 error, stopping boundaries and spacing of looks is kept the same.

5. **Checks against Public Domain Software.** Public domain Fortran routines developed at the University of Wisconsin (see Reboussin et. al., 2002) can be freely downloaded from <http://www.landemets.com>. East 3 replicated the results produced by this software for adjusted p-values, confidence intervals and unbiased estimators following sequential monitoring. The stopping boundaries are evaluated differently in the two procedures and result in small differences. A detailed explanation for these differences is provided in Appendix F.

Y.2 Fixed-Sample Designs (FSD)

Y.2.1 Details

The statistical results computed by FSD have been subjected to rigorous and extensive quality-assurance testing for purposes of validation. A summary table displaying the methods used for each statistical procedure is given below. Each row of the table corresponds to a statistical procedure and the columns **C1-C6** correspond to comparison of **FSD** result with results using other software as indicated below.

- **C1 Column:** Comparison with nQuery 7.0.
- **C2 Column:** Comparison with SAS 9.1.
- **C3 Column:** Comparison with independent developed R programs and SAS macros.
- **C4 Column:** Comparison with PASS 2008.
- **C5 Column:** Comparison with StatXact8.
- **C6 Column:** Comparison with East 5.2.

In the following tables,

”1” indicates that the comparison was made for the test and results from **FSD** were comparable to the respective software.

”2” indicates that the comparison was made but the results did not match for reasons indicated at the bottom of the table.

”-” denotes that a comparable test in other software was not available or the comparison was not applicable.

Module : Design

| Sr. | Test Name | C1 | C2 | C3 | C4 | C5 | C6 |
|----------|--|----|----|----|----|----|----|
| 1 | Continuous: One Mean | | | | | | |
| | Single Mean: Z Test | - | - | 1 | - | - | 1 |
| | Single Mean: t Test | 1 | 1 | 1 | - | - | - |
| | Difference of Means for Paired Data: Superiority: Z Test | - | - | 1 | - | - | 1 |
| | Difference of Means for Paired Data: Superiority: t Test | 1 | - | 1 | - | - | - |
| | Difference of Means for Paired Data: Non-Inferiority: Z Test | - | - | 1 | - | - | - |
| | Difference of Means for Paired Data: Non-Inferiority: t Test | 1 | - | 1 | - | - | - |
| | Difference of Means for Paired Data: Equivalence: t Test | - | - | 1 | - | - | - |
| | Ratio of Means for Paired Data: Superiority: Z Test | - | - | 1 | - | - | - |
| | Ratio of Means for Paired Data: Superiority: t Test | - | - | 1 | - | - | - |
| | Ratio of Means for Paired Data: Non-Inferiority: Z Test | - | - | 1 | - | - | - |
| | Ratio of Means for Paired Data: Non-Inferiority: t Test | - | - | 1 | - | - | - |
| | Ratio of Means for Paired Data: Equivalence: t Test | - | - | 1 | - | - | - |

Y On validating the East Software

Module : Design

| Sr. | Test Name | C1 | C2 | C3 | C4 | C5 | C6 |
|-----|---|----|----|----|----|----|----|
| 2 | Continuous: Two Means | | | | | | |
| | Difference of Means for Independent Data: Superiority: Z Test | - | - | 1 | - | - | 1 |
| | Difference of Means for Independent Data: Superiority: t Test | 1 | 1 | 1 | - | - | - |
| | Difference of Means for Independent Data: Non-Inferiority: Z Test | - | - | 1 | - | - | 1 |
| | Difference of Means for Independent Data: Non-Inferiority: t Test | 1 | 1 | 1 | - | - | - |
| | Difference of Means for Independent Data: Equivalence: t Test | - | - | 1 | - | - | - |
| | Ratio of Means for Independent Data: Superiority: Z Test | - | - | 1 | - | - | - |
| | Ratio of Means for Independent Data: Superiority: t Test | 1 | - | 1 | - | - | - |
| | Ratio of Means for Independent Data: Non-Inferiority: Z Test | - | - | 1 | - | - | - |
| | Ratio of Means for Independent Data: Non-Inferiority: t Test | - | - | 1 | - | - | - |
| | Ratio of Means for Independent Data: Equivalence: t Test | 1 | - | 1 | - | - | - |
| | Wilcoxon Mann Whitney Test for Independent Data | 1 | - | 1 | - | - | 1 |
| | Difference of Means for Crossover Data: Superiority: t Test | 1 | - | 1 | - | - | - |
| | Difference of Means for Crossover Data: Non-Inferiority: t Test | - | - | 1 | - | - | - |
| | Difference of Means for Crossover Data: Equivalence: t Test | 1 | - | 1 | - | - | - |
| | Ratio of Means for Crossover Data: Superiority: t Test | - | - | 1 | - | - | - |
| | Ratio of Means for Crossover Data: Non-Inferiority: t Test | - | - | 1 | - | - | - |
| | Ratio of Means for Crossover Data: Equivalence: t Test | 1 | - | 1 | - | - | - |

Y On validating the East Software

Module : Design

| Sr. | Test Name | C1 | C2 | C3 | C4 | C5 | C6 |
|----------|---|----|----|----|----|----|----|
| 3 | Continuous: Many Means | | | | | | |
| | One Way ANOVA | 1 | - | 1 | - | - | - |
| | One Way Contrast | 1 | - | 1 | - | - | - |
| | One Way Repeated Measures (Constant Correlation) ANOVA | 1 | - | 1 | - | - | - |
| | One Way Repeated Measures Contrast (Constant Correlation) | 1 | - | 1 | - | - | - |
| | Two Way ANOVA | 1 | - | 1 | - | - | - |
| 4 | Continuous: Regression | | | | | | |
| | Linear Regression: Single Slope | 1 | - | 1 | - | - | - |
| | Linear Regression for Comparing Two Slopes | 1 | - | 1 | - | - | - |
| | Repeated Measures for Comparing Two Slopes | - | - | 1 | - | - | - |
| 5 | Discrete: Single Proportion | | | | | | |
| | Single Proportion (Asymptotic) | 1 | - | 1 | - | - | 1 |
| | Single Proportion (Exact) | 1 | - | 1 | - | - | 1 |
| | McNemars Test for Matched Pairs(*) | 2 | - | 1 | - | - | 1 |
| 6 | Discrete: Two Proportion | | | | | | |
| | Difference of Proportions: Superiority | 1 | - | 1 | - | - | 1 |
| | Difference of Proportions: Non-Inferiority | 1 | - | 1 | - | - | 1 |
| | Difference of Proportions: Equivalence | - | - | 1 | - | - | 1 |
| | Ratio of Proportions: Superiority | - | - | 1 | - | - | 1 |
| | Ratio of Proportions: Non-Inferiority (Wald test) | - | - | 1 | - | - | 1 |
| | Ratio of Proportions: Non-Inferiority (Score test) | - | - | 1 | 1 | - | - |
| | Odds Ratio of Proportions: Superiority(**) | 2 | - | 1 | - | - | 1 |
| | Odds Ratio of Proportions: Non-Inferiority | - | - | 1 | - | - | 1 |
| | Common Odds Ratio for Stratified 2x2 Table | 1 | - | 1 | - | - | - |
| | Fisher Exact Test | 1 | - | - | - | 1 | - |

Module : Design

| Sr. | Test Name | C1 | C2 | C3 | C4 | C5 | C6 |
|-----------|--|----|----|----|----|----|----|
| 7 | Discrete: Many Proportion | | | | | | |
| | Trend in R Ordered Proportions | 1 | - | 1 | - | - | - |
| | Chi-square Test for Rx2 Table | 1 | - | 1 | - | - | - |
| | Chi-square Test of Specified Proportions in C Categories | 1 | - | 1 | - | - | - |
| | Two-Group Chi-square Test Comparing Proportions in C Categories | 1 | - | 1 | - | - | - |
| | Chi-square Test of Comparing Proportions in RXC Table | 1 | - | 1 | - | - | - |
| | Wilcoxon Rank Sum Test for Ordered Categorical Data | 1 | - | - | - | - | 1 |
| 8 | Discrete: Regression | | | | | | |
| | Logistic Regression with Single Normal Covariate | - | - | 1 | 1 | - | - |
| | Logistic Regression with Single Normal Covariate Adjusted for other Covariates | - | - | 1 | 1 | - | - |
| 9 | Discrete: Agreement | | | | | | |
| | Cohen's Kappa(***) | 2 | - | 1 | - | - | - |
| | Cohen's Kappa (C Ratings) | - | - | 1 | - | - | - |
| 10 | Events: Survival | | | | | | |
| | Logrank Test: Superiority | - | - | - | - | - | 1 |
| | Logrank Test: Non-Inferiority | - | - | - | - | - | 1 |

Note

(*)The results for McNemar's Test for Matched Pairs from FSD do not match with those from nQuery as FSD uses Normal approximation while nQuery uses the Chi-square test.

(**) The formulation of the Odds Ratio of Proportions: Superiority Test is different in FSD and nQuery which results in the mismatch between their results.

(***) There is a difference in the results from FSD and nQuery for the Cohen's Kappa Test due to the difference in the techniques followed.

Y On validating the East Software

Module : Analysis

| Sr. | Test Name | C1 | C2 | C3 | C4 | C5 | C6 |
|-----|--|----|----|----|----|----|----|
| 1 | Continuous: One Mean | | | | | | |
| | Single Mean: Z Test | - | - | 1 | - | - | - |
| | Single Mean: t Test | - | 1 | - | - | - | - |
| | Difference of Means for Paired Data: Superiority: Z Test | - | - | 1 | - | - | - |
| | Difference of Means for Paired Data: Superiority: t Test | - | 1 | - | - | - | - |
| | Difference of Means for Paired Data: Non-Inferiority: Z Test | - | - | 1 | - | - | - |
| | Difference of Means for Paired Data: Non-Inferiority: t Test | - | 1 | - | - | - | - |
| | Difference of Means for Paired Data: Equivalence: t Test | - | 1 | - | - | - | - |
| | Ratio of Means for Paired Data: Superiority: Z Test | - | - | 1 | - | - | - |
| | Ratio of Means for Paired Data: Superiority: t Test | - | 1 | - | - | - | - |
| | Ratio of Means for Paired Data: Non-Inferiority: Z Test | - | - | 1 | - | - | - |
| | Ratio of Means for Paired Data: Non-Inferiority: t Test | - | 1 | - | - | - | - |
| | Ratio of Means for Paired Data: Equivalence: t Test | - | 1 | - | - | - | - |
| | Wilcoxon Signed Rank Test | - | 1 | 1 | - | - | - |

Module : Analysis

| Sr. | Test Name | C1 | C2 | C3 | C4 | C5 | C6 |
|----------|---|----|----|----|----|----|----|
| 2 | Continuous: Two Means | | | | | | |
| | Diff of Means for Independent Data: Superiority: Z | - | - | 1 | - | - | - |
| | Diff of Means for Independent Data: Superiority: t | - | 1 | - | - | - | - |
| | Diff of Means for Independent Data: NI: Z | - | - | 1 | - | - | - |
| | Diff of Means for Independent Data: NI: t | - | 1 | - | - | - | - |
| | Diff of Means for Independent Data: Equivalence: t | - | 1 | - | - | - | - |
| | Ratio of Means for Independent Data: Superiority: Z | - | - | 1 | - | - | - |
| | Ratio of Means for Independent Data: Superiority: t | - | 1 | - | - | - | - |
| | Ratio of Means for Independent Data: NI: Z | - | - | 1 | - | - | - |
| | Ratio of Means for Independent Data: NI: t | - | 1 | - | - | - | - |
| | Ratio of Means for Independent Data: Equivalence: t | - | 1 | - | - | - | - |
| | Wilcoxon Mann Whitney Test for Independent Data | - | 1 | 1 | - | - | - |
| | Diff of Means for Crossover Data: Superiority: t | - | 1 | - | - | - | - |
| | Diff of Means for Crossover Data: NI: t | - | 1 | - | - | - | - |
| | Diff of Means for Crossover Data: Equivalence: t | - | 1 | - | - | - | - |
| | Ratio of Means for Crossover Data: Superiority: t | - | 1 | - | - | - | - |
| | Ratio of Means for Crossover Data: NI: t | - | 1 | - | - | - | - |
| | Ratio of Means for Crossover Data: Equivalence: t | - | 1 | - | - | - | - |
| | Wilcoxon Mann Whitney Test: 2x2 Crossover | - | 1 | 1 | - | - | - |

Y On validating the East Software

Module : Analysis

| Sr. | Test Name | C1 | C2 | C3 | C4 | C5 | C6 |
|-------------------|--|----|----|----|----|----|----|
| 3 | Continuous: Many Means | | | | | | |
| | One way ANOVA | - | 1 | 1 | - | - | - |
| | One Way Repeated Measures (Constant Correlation) ANOVA | - | 1 | 1 | - | - | - |
| | Two Way ANOVA | - | 1 | 1 | - | - | - |
| 4 | Continuous: Regression | | | | | | |
| | Multiple Linear Regression | - | - | - | - | 1 | - |
| | Repeated Regression | - | 1 | - | - | - | - |
| | Linear Mixed Effects Model: Difference of Means (crossover data) | - | 1 | - | - | - | - |
| | Linear Mixed Effects Model: Ratio of Means (crossover data) | - | 1 | - | - | - | - |
| 5 | Discrete: Single Proportion | | | | | | |
| | Single Proportion (Asymptotic) | - | 1 | - | - | - | - |
| | Single Proportion (Exact) | - | - | - | - | 1 | - |
| | McNemars Test for Matched Pairs | - | - | - | - | 1 | - |
| 6 | Discrete: Two Proportion | | | | | | |
| | Difference of Proportions: Superiority | - | - | - | - | 1 | - |
| | Difference of Proportions: Non-Inferiority (Wald) | - | - | 1 | - | - | - |
| | Difference of Proportions: Non-Inferiority (Score) | - | - | - | - | 1 | - |
| | Difference of Proportions: Equivalence | - | - | - | - | 1 | - |
| | Ratio of Proportions: Superiority | - | - | - | - | 1 | - |
| | Ratio of Proportions: Non-Inf (Wald) | - | - | 1 | - | - | - |
| | Ratio of Proportions: Non-Inf (Score) | - | - | - | 1 | - | - |
| | Odds Ratio of Proportions: Superiority | - | - | - | - | 1 | - |
| | Odds Ratio of Proportions: Non-Inf (Wald) | - | - | 1 | - | - | - |
| | Odds Ratio of Proportions: Non-Inf (Score) | - | - | - | - | 1 | - |
| | Common Odds Ratio for Stratified 2x2 Table | - | - | - | - | 1 | - |
| Fisher Exact Test | - | - | - | - | 1 | - | |

Module : Analysis

| Sr. | Test Name | C1 | C2 | C3 | C4 | C5 | C6 |
|-----------|---|----|----|----|----|----|----|
| 7 | Discrete: Many Proportion | | | | | | |
| | Trend in R Ordered Proportions | - | - | - | - | 1 | - |
| | Chi-square Test for Rx2 Table | - | - | - | - | 1 | - |
| | Chi-square Test of Specified Proportions in C Categories | - | - | - | - | 1 | - |
| | Two-Group Chi-square Test Comparing Proportions in C Categories | - | - | - | - | 1 | - |
| | Chi-square Test of Comparing Proportions in RXC Table | - | - | - | - | 1 | - |
| | Wilcoxon Rank Sum Test for Ordered Categorical Data | - | - | - | - | 1 | - |
| 8 | Discrete: Regression | | | | | | |
| | Logistic Regression | - | - | - | - | 1 | - |
| | Probit Regression | - | - | - | - | 1 | - |
| | Clog Log Regression | - | - | - | - | 1 | - |
| 9 | Discrete: Agreement | | | | | | |
| | Cohen's Kappa | - | - | - | - | 1 | - |
| 10 | Events: Survival | | | | | | |
| | Logrank Test: Superiority | - | 1 | 1 | - | - | - |
| | Logrank Test: Non-Inferiority | - | 1 | 1 | - | - | - |

Module: Simulation

Results of Simulations in FSD are validated by checking the internal consistency. For example, the estimated probability of rejection from Simulations was compared with the analytical result obtained from FSD design procedure.

Module: Data Explorer

The Data Explorer tests' outputs have been compared with the corresponding Cytel Studio 8 results.

Y.2.2 FSD MC Procedures

The Multiple Comparison procedures implemented in FSD MCP have been validated extensively. Various methods were employed for the statistical validation of these

Y On validating the East Software

procedures. The following summary table states the methods used for validating each of the Multiple Comparison Procedures. Each row of the table corresponds to a procedure and the columns C1-C4 correspond to the validation method used as described below:

- **C1 Column:** Comparison with SAS 9.1
- **C2 Column:** Comparison with R 2.12.1 (Packages used: 'multxpert', 'mutoss')
- **C3 Column:** Comparison with PASS 2005
- **C4 Column:** Comparison with independently developed (in-house) R/SAS macros

In the following tables, '1' indicates that the comparison was made for the test and results from FSD MCP were comparable to the respective software; '2' indicates that the comparison was made but the results either matched partially or did not match for reasons indicated at the bottom of the table; '-' denotes that a comparable test in other software was not available or the comparison was not applicable.

Table Y.1: Module: Design

| Sr.# | MCP | C1 | C2 | C3 | C4 |
|------|---------------------------|----|----|----|----|
| 1 | Dunnett's single step (*) | 2 | - | 2 | 1 |
| 2 | Dunnett's step down | - | - | - | 1 |
| 3 | Dunnett's step up | - | - | - | 1 |
| 4 | Bonferroni | - | - | - | 1 |
| 5 | Sidak | - | - | - | 1 |
| 6 | Weighted Bonferroni | - | - | - | 1 |
| 7 | Holm's step down | - | - | - | 1 |
| 8 | Hochberg's step up | - | - | - | 1 |
| 9 | Hommel's step up | - | - | - | 1 |
| 10 | Fixed sequence | - | - | - | 1 |
| 11 | Fallback | - | - | - | 1 |

Note: (*) The critical value for Dunnett's single step was available with SAS and hence validated with it. This procedure was also compared with PASS. However PASS provides for 2-sided test and FSD MCP has 1-sided test. Hence the results were comparable in case of scenarios where the treatment means were either all greater than or all less than the control mean. Note that these tests are simulation based and hence cannot be matched exactly with PASS.

Table Y.2: Module: Analysis

| Sr.# | MCP | C1 | C2 | C3 | C4 |
|------|----------------------------|----|----|----|----|
| 1 | Dunnett's single step (**) | 2 | 1 | - | 1 |
| 2 | Dunnett's step down | - | 1 | - | 1 |
| 3 | Dunnett's step up | - | 1 | - | 1 |
| 4 | Bonferroni | 1 | 1 | - | - |
| 5 | Sidak | 1 | 1 | - | - |
| 6 | Weighted Bonferroni | - | 1 | - | - |
| 7 | Holm's step down | 1 | 1 | - | - |
| 8 | Hochberg's step up | 1 | 1 | - | - |
| 9 | Hommel's step up | 1 | 1 | - | - |
| 10 | Fixed sequence | - | 1 | - | - |
| 11 | Fallback | - | 1 | - | - |

(**) The critical value and Simultaneous CI was available with SAS and hence validated with it.

Z *List of East Beta Testers*

East 6.3

Dey Jyotirmoy Abbvie

Wang, Xin Abbvie

Dunbar, Martin Abbvie

Zeng, Jiewei Abbvie

Munasinghe, Wijith P Abbvie

Yodit,Seifu Allergan

Matcham, James Astrazeneca

Su, Hong-Lin Astrazeneca

Zhang, Charlie Biomarin

Wu, Xiaoling Celgene

Liu, Kejian Celgene

Isaacson, Jeff Clovis

Wang, Qiang Daiichi

Wang, Yibin Daiichi

Chen, Shuquan Daiichi

Lee, James Daiichi

Bekele, Neby Gilead

Li, Xiaoming Gilead

Zhang, Grace GSK

Anderson, Keaven Merck

Gause, Christine Merck

Tsai, Kuenhi Merck

Huang, Xiaobi (Shelby) Merck

Xu, Jialin Merck

Mehrotra, Devan Merck

Lu, Lin Nektar

Goldwasser, Meredith Novartis

Holmgren, Eric Oncomed

Ro, Sunhee Onyx

Liao, Olivia Onyx

Perevozskaya, Inna Pfizer

Alun Bedding Roche-Genentech

Lin, Jianchang Takeda

Liu, Patrick Takeda

Pickard, Mike Takeda

Wang, Ling Takeda

Liu, Yi Takeda

Z *List of East Beta Testers*

East Architect

Keaven Anderson Merck

Loic Darchy Sanofi

Yahya Daoud Baylor Health

Bing Gao Amgen

Brenda Gaydos Eli Lilly

Sally Hollis Astrazeneca

Xin Huang Pfizer Inc.

Chris Jennison University of Bath

Sheela Kolluri Pfizer Inc.

Chunming Mark Li Pfizer Inc.

Jianxin Lin Merck

Xun Lin Pfizer Inc.

Jiajun Liu Merck

John Loewy qPharmetra

Richard Manski Abbott Laboratories

Joel Miller Eli Lilly

May Mo Amgen

Yili Pritchett Abbott Laboratories

Bill Prucka Eli Lilly

Natasa Rajcic Pfizer Inc.

Brad Robertson Eli Lilly

Supriya Satwah Unilever

Yue Shentu Merck

Sam Suzuki Amgen

Enayet Talukder Pfizer Inc.

Jie Tang Pfizer Inc.

Qi Tang Abbott Laboratories

Bruce Turnbull Cornell University

Xuan Wang Baylor Health

Jim Ware Harvard University

Bin Yao Amgen

Tianhui Zhou Pfizer Inc.

Z *List of East Beta Testers*

East 4

Marilyn Agin Pfizer

Robert Chew Pfizer

Loic Darchy Sanofi-Aventis

Andrew Kramar CRLC Val d'Aurelle

Steve Lagakos Harvard School of Public Health

Xiaoming Li Merck

Devan Mehrotra Merck

Mike Smith Pfizer

Thomas Stiger Pfizer

Chau Thach Merck

East 3

Dan Anbar Millenium Biostatistics

Keaven Anderson Centocor

Linda Christie Dana-Farber Cancer Institute

George Cotsonis Emory University

Loic Darchy Sanofi-Synthelabo

Dave DeMets University of Wisconsin

Brenda Gaydos Eli Lilly and Company

Vicki Hertzberg Emory University

Joan Hilton UC San Francisco

Chris Jennison University of Bath

Kyungmann Kim University of Wisconsin

Andrew Kramar CRLC Val d'Aurelle

Peter Lachenbruch FDA CBER

Steve Lagakos Harvard School of Public Health

Robert Lagos New England Research Unit

Elizabeth Ludington Statistics Collaborative

Mike Lynn Emory University

Young Park Wyeth

Z *List of East Beta Testers*

Heather Ribaud Harvard School of Public Health

Wasima Rida FDA Center for Biologics Evaluation & Research

Larry Roi Aventis Pharmaceuticals

Roy Tamura Eli Lilly and Company

Butch Tsiatis North Carolina State University

Bruce Turnbull Cornell University

Sue-Jane Wang FDA –Center for Drug Evaluation Research

Wendy Wilson Aventis Pharmaceuticals

Ru-Fang Yeh UC San Francisco

Boguang Zhen FDA/CBER/OBE/DB

East-2000

Mirza Ali Otsuka America Pharmaceuticals

Dan Anbar Millenium Biostatistics

Nupun Andhivarthai TAP Holdings Inc.

Tad Archambault Virtu Stat Ltd.

Peter Armitage University of Oxford

Juergen Berger University of Hamburg

Harry Bushar FDA CDRH

David DeMets University of Wisconsin

Richard Hellmund Zeneca Pharmaceuticals

Jay Herson Applied Logic Associates

Irving Hwang Irving Consulting Group

Allen Izu Chiron Corporation

Chris Jennison University of Bath

Kyungmann Kim University of Wisconsin

Peter Lachenbruch FDA CBER

Steve Lagakos Harvard School of Public Health

Gordon Lan Pfizer Inc.

Gracie Lieberman Genentech Inc.

Z *List of East Beta Testers*

Scott Maxwell University of Notre Dame

Larry Muenz Independent Consultant

Theophile Niyonsenga University of Sherbrooke

Abdul Sankoh Genetics Institute

Greg Stoddard University of Utah

Judy Sy Genentech Inc.

Peter Thall University of Texas

Bruce Turnbull Cornell University

Duolao Wang London School of Hygiene and Tropical Medicine

Richard Wu RPR Pharmaceuticals

Peter Zhang Otsuka America Pharmaceuticals

Huaqing Zhao The Children's Hospital of Philadelphia

References

- Abad-Santos F et al. (2005). Assessment of sex differences in pharmacokinetics and pharmacodynamics of almodipine in a bioequivalence study. *Pharmacological Research*, 51, 445-452.
- Agresti A (2002). *Categorical Data Analysis*. (2nd Ed). John Wiley & Sons, New York.
- Agresti A, Min Y. (2001). On small-sample confidence intervals for parameters in discrete distributions. *Biometrics* 57: 963-971.
- Andersen EB (1990). *The Statistical Analysis of Categorical Data*. Springer-Verlag, Berlin-Heidelberg.
- Anderson K (2002). Evaluating sponsor responsibilities for interim analysis with DMC's. Presented at the Clinical Trials Data Monitoring Committees meeting, Philadelphia Barnett International Conference Group, Philadelphia.
- Anderson S, Hauck WW (1990). Consideration of individual bioequivalence. *J. Pharmacokin. Biopharm*, 18, 259-273.
- Andrews DF and Herzberg AM (1985). *Data*. Springer-Verlag, New York.
- Armitage P (1955). Test for linear trend in proportions and frequencies. *Biometrics*, 11: 375-386
- Armitage P (1957). Restricted sequential procedures. *Biometrika*, 44, 9-56.
- Armitage P (1975). *Sequential Medical Trials*. Blackwell Scientific Publications, Oxford.
- Armitage P, McPherson CK and Rowe BC (1969). Repeated significance tests on accumulating data. *J. R. Statist. Soc. A*, 132, 232-44.
- Arvin AM, Kushner JH, Feldman S, et al. (1982). Human leukocyte interferon for the treatment of varicella in children with cancer. *New England Journal of*

References

- Medicine* 306:761-765.
- Ayer M, Brunk HD, Ewing GM, Reid WT, Silverman. An empirical distribution function for sampling with incomplete information (1955). *Ann. Math. Statistics*;26:641-647.
- Barlow RE, Bartholomew DJ, Bremner JM, Brunk HD. *Statistical Inference Under Order Restrictions*. New York: John Wiley; 1972.
- Barnard GA (1945). A new test for 2×2 tables. *Nature* 156:177.
- Belsey D, Kuh E and Welsch R (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*, Wiley New York.
- Benjamini Y, Hochberg Y (1997). Multiple hypothesis testing with weights. *Scandinavian Journal of Statistics*, 24, 407-418.
- Berger RL, Boos DD (1994). P values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association* 89:1012-1016.
- Beta Blocker Heart Attack Trial. (1981). Beta Blocker Heart Attack Trial” Design features. *Controlled Clinical Trials*, 2, 275-85.
- Beta-Blocker Heart Attack Trial (1982). A randomized trial of propranolol in patients with acute myocardial infarction. *JAMA*, 247, 1707-14.
- Bickel PJ, Klaasen CAJ, Ritov Y and Wellner JA (1993). *Efficient and adaptive estimation for semiparametric models*. John Hopkins University press, Baltimore.
- Block DA, Kraemer HC (1989). 2x2 kappa coefficients: measures of agreement or association. *Biometrics*. 45: 269-287
- Blyth C, Still H (1983). Binomial confidence intervals. *Journal of American Statistical Association*, 78:108-116.
- Bofinger E (1987). Step-down procedures for comparison with a control. *Australian Journal of Statistics*, 29, 348-364.
- Bonferroni CE (1935). Il calcolo delle assicurazioni su gruppi di teste. *In Studi in onore*

del Professore Salvatore Ortu Carboni. Rome, Italy, 13-60.

- Bonferroni CE (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* 8, 3-62.
- Bowerman B, O'Connel R, Dickey D (1986). *Linear Statistical Models, an Applied Approach*. Duxbury Press, Belmont, California.
- Brannath W, Posch M, and Bauer P (2002). Recursive Combination Tests. *JASA*, 97, 236-244.
- Brannath W, Mehta CR, Posch M (2009). Exact Confidence Bounds Following Adaptive Group Sequential Tests. *Biometrics*, 65(2), 539-546.
- Breslow NE, Day NE (1980). The Analysis of Case-Control Studies. *IARC Scientific Publications No. 32* Lyon, France.
- Breslow NE, Day NE (1987). *The Design and Analysis of Cohort Studies*. IARC Scientific Publication N0.82, Lyon, France.
- Bristol DR (1993a). Probabilities and sample sizes for the two one-sided tests procedure, *Communications in Statistics - Theory and Methods*, A22(7), 1953-1961.
- Bristol DR (1993b). Planning Survival Studies To Compare A Treatment To An Active Control. *Journal of Biopharmaceutical Statistics*, 3(2), 153-158.
- Burgess IF, Brown CM, and Lee PN (2005). Treatment of head louse infestation with 4% dimeticone lotion: randomised controlled equivalence trial. *BMJ* 330:1423.
- Cantor AB (1996). Sample size calculation for Cohen's kappa. *Psychological Methods*. 1(2): 150-153.
- Casagrande JT, Pike MC, and Smith PG (1978). An improved approximate formula for comparing two binomial distributions. *Biometrics*, 34, 483-86.
- Casella G (1986). Refining binomial confidence intervals. *Canadian Journal of Statistics* 14:113-129.

References

- Chambers J, Cleveland W, Kleiner B, Tukey P. (1983). *Graphical Methods for Data Analysis*. Wadsworth.
- Chan ISF (1998). Exact tests of equivalence and efficacy with a non-zero lower bound for comparative studies. *Statistics in Medicine* 17:1403-1413.
- Chan ISF, Zhang Z (1999). Test based exact confidence intervals for the difference of two binomial proportions. *Biometrics* 55:1201-1209
- Chen JYH, DeMets DL, Lan KKG.(2004). Increasing the sample size when the unblinded interim result is promising. *Statistics in Medicine*, 23(7), 1023-1038.
- Chernick MR, Liu CY (2002). The saw-toothed behavior of power versus sample size and software solutions: single binomial proportion using exact methods. *The American Statistician*, 56: 149-155.
- Clopper CJ, Pearson E. (1934). The use of confidence or fiducial limits illustrated in the case of binomial. *Biometrika* 26:404-413.
- Chow SC and Liu JP (1992). *Design and Analysis of Bioavailability and Bioequivalence Studies*. Marcel Dekker, New York.
- Chow S, Shao J, Wang H (2003). *Sample Size Calculations in Clinical Research*. Taylor and Francis, New York.
- Cleveland, W (1993). *Visualizing Data*. Hobart Press.
- Cleveland, W (1985). *Elements of Graphing Data*. Wadsworth.
- Cochran WG, Cox GM (1957). *Experimental Designs*. Second Edition, New York: John Wiley & Sons, Inc.
- Coe PR, Tamhane AC (1993). Exact Repeated Confidence Intervals for Bernoulli Parameters in a Group Sequential Clinical Trial. *Controlled Clinical Trials* 14, 19-29.
- Cohen J (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20: 37-46.
- Cole JW, Grizzle JE (1966). Applications of Multivariate Analysis of Variance to

- Repeated Measures Experiments. *Biometrics*, 22, 810:828.
- Collett D (1994). *Modelling Survival Data in Medical Research*. Chapman & Hall, London.
- Collett, D. (2002). *Modeling Binary Data, 2nd ed.* Boca Raton, FL: CRC Press.
- Collett, D and Jemain, AA (1985). Residuals, outliers and influential observations in regression analysis. *Sains Malaysiana*, 14, 493-511
- Conover WJ (1980). *Practical Nonparametric Statistics*, 2nd edition. John Wiley & Sons, New York.
- Cook D, Weisberg S (1982). *Residuals and Influence in Regression*. Chapman and Hall, London.
- Cook RD (1979). Influential observations in linear regression. *JASA*, 74, 169:174.
- Cook TD, DeMets DL. (2008). *Introduction to statistical methods for clinical trials*. Chapman and Hall. 7: 296.
- Cook RD and Weisberg S (1982). *Residuals and Influence in Regression*. Chapman & Hall, London.
- Corcoran C D, Mehta CR and Senchaudhuri P (2000) : Power Comparisons for Tests of Trend in Dose Response Studies. *Statistics in Medicine*, 19, 3037-3050.
- Cox DR and Snell EJ (1989). *Analysis of Binary Data. 2nd Edition*. Chapman and Hall, London.
- CRASH Trial Collaborators (2004). Effect of intravenous corticosteroids on death within 14 days in 10008 adults with clinically significant head injury. *Lancet*, 364, 1321-28.
- Crowder M and Hand D (1990). *Analysis of repeated measures*, Chapman & Hall/CRC.
- Cui L, Hung HMJ, and Wang S (1999). Modification of sample size in group sequential clinical trials. *Biometrics*, 55, 853-857.
- Davidson MH et al (1999). Weight control and risk factor reduction in obese subjects

References

- treated for 2 years with Orlistat (1999). *JAMA*, 281, 235-42.
- DeMets DL and Gail MH (1985). Use of logrank tests and group sequential methods at fixed calendar times. *Biometrics*, 41, 1039-44.
- DeMets DL, Hardy R, Friedman LM and Lan KKG (1984). Statistical aspects of early termination in the Beta-Blocker Heart Attack Trial. *Controlled Clinical Trials*, 5, 362-72.
- DeMets DL and Lan KKG (1995). The alpha spending function approach to interim data analyses. In: *Recent advances in clinical trial design and analysis*, Thall PF Ed. Kluwer Academic Publishers, Boston.
- DeMets DL and Ware JH (1980). Group sequential methods for clinical trials with a one-sided hypothesis. *Biometrika*, 67, 651-60.
- Diggle PJ (1988). An Approach to the Analysis of Repeated Measurements. *Biometrics*, 44, 959: 971.
- Diletti, E., Hauschke, D. and Steinijans, VW (1991). Sample Size Determination for Bioequivalence Assessment by Means of Confidence Intervals. *International Journal of Clinical Pharmacology, Therapy and Toxicology*, 29, 1, 1-8.
- Dixon WJ, Massey FJ (1983). *Introduction to Statistical Analysis. fourth Edition*, McGraw-Hill, 14.
- Dmitrienko A, Tamhane AC, Bretz F (2010). *Multiple Testing Problems in Pharmaceutical Statistics*. Chapman & Hall.
- Draper NR, Smith H (1966). *Applied Regression Analysis*. New York: John Wiley & Sons, Inc.
- Draper CC, Voller A, Carpenter RG (1972). The Epidemiologic Interpretation of Serologic Data in Malaria. *American Journal of Tropical Medicine and Hygiene*, 21, 696-703.
- Duffy SW (1984). Asymptotic and exact power for the McNemar test and its analogue with R controls per case. *Biometrics* 40:1005-1015.
- Dunnnett CW (1980). Pairwise Multiple Comparisons in the Homogeneous Variance, Unequal Sample Size Case. *Journal of the American Statistical Association*,

75, 789-795.

Dunnett CW (1985). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 50, 1096-1121.

Dunnett CW and Gent M (1977). Significance testing to establish equivalence between treatments, with special reference to data in the form of 2 x 2 tables. *Biometrics*, 33, 593-602.

Dunnett CW, Tamhane AC (1991). Step-down multiple tests for comparing treatments with a control in unbalanced one-way layouts. *Statistics in Medicine*, 10,939-947.

Dunnett CW, Tamhane AC (1992). A step-up multiple test procedure. *Journal of the American Statistical Association*, 87, 162-170

Dunnett CW, Tamhane AC (1992). Comparisons between a new drug and active and placebo controls in efficacy clinical trial. *Statistics in Medicine*, 11, 1057-1063.

Dunnett CW, Tamhane AC (1995). Step-up multiple testing of parameters with unequally correlated estimates. *Biometrics*, 51, 217-227.

Dupont, WD. and Plummer, WD., Jr. (1998). Power and Sample Size Calculations for Studies Involving Linear Regression. *Controlled Clinical Trials*, 19, 589-601.

Du Toit, Steyn, Stumpf (1986). *Graphical Exploratory Data Analysis*. Springer-Verlag.

Egret Siz (1997). *Sample size and power for nonlinear regression models. Version 1. Reference manual*. Cytel Software Corporation, Cambridge, MA.

Elashoff, JD. (2005) nQuery Advisor Version 6.0. Statistical Solution Ltd., Los-Angeles,CA.

Everitt BS (1995). The Analysis of Repeated Measures: A Practical Review with Examples. *The Statistician*, 44, 113:135.

Facey KM (1992). A sequential procedure for a Phase II efficacy trial in hypercholesterolemia. *Controlled Clinical Trials*, 13, 122-133.

References

- Fairbanks K and Madsen R (1982). P values for tests using a repeated significance test design. *Biometrika*, 69, 69-74.
- Farrington CP and Manning G (1990). Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine*, 9, 1447-1454.
- FDA (2002). Bioequivalence Guidance, Guidance for Industry No. 35, Oct. 9, 2002.
- FDA (CDER and CBER) (2010). Draft Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics, February 2010.
- Feng S, Liang Q, Kinser R, Newland K and Guilbaud R (2006). Testing equivalence between two laboratories or two methods using paired-sample analysis and interval hypothesis testing. *Analytical and Bioanalytical Chemistry*, 385(5), 975-981.
- Fienberg SE (1980). *The Analysis of Cross-classified Categorical Data. 2nd Edition*. M.I.T. Press, Cambridge, MA.
- Firth, D (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27-38.
- Fisher, RA (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh.
- Fisher R (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7 (2), 179-188.
- Fitzmaurice GM, Laird NM, Ware JH (2004). *Applied Longitudinal Analysis*. John Wiley & Sons, New York.
- Flack VF, Afifi AA, Lachenbruch PA (1988). Sample Size Determinations for the two rater kappa statistic. *Psychometrika*, 53(3): 321-325.
- Fleiss JL. (1981). *Statistical Methods for Rates and Proportions*. John Wiley & Sons, New York.
- Fleiss JL, Cohen J (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement* 33: 613-619.

- Fleiss, JL, Levin, B and Pike, MC (2003). *Statistical Methods for Rates and Proportions*. John Wiley & Sons, New York.
- Fleiss JL, Tytun A, Ury SHK (1980). A simple approximation for calculating sample sizes for comparing independent proportions. *Biometrics* 36: 343-346
- Fleming, TR (1982). One-Sample Multiple Testing Procedure for Phase II Clinical Trials. *Biometrics*, 38, 143-151.
- Fleming, TR (2008). Current issues in non-inferiority trials. *Stat Med*, 27(3), 317-32.
- Freedman LS (1982). Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine*, 1: 121-129.
- Freirich et.al. (1963) The effect of 6-Mercaptopurine on the duration of Steroid induced remissions in acute leukemia: A model for evaluation of other potentially useful therapy, *Blood*, 1963 21 699:716
- Friede Tim and Schmidli Heinz (2010). Blinded sample size re-estimation with count data: Methods and applications in multiple sclerosis. *Statistics in Medicine*, 29 1145-1156
- Gallo P, Chuang-Stein C, Draglin V, Gaydos B, Krams M, Pinheiro J (2006). Adaptive designs in clinical drug development – an executive summary of the PhRMA Working Group. *J. Biopharm Statist.*, 16, 275-83.
- Gao P, Ware J, Mehta C (2008). Sample size re-estimation for adaptive sequential design in clinical trials. *J. Biopharm. Statist.*, 18(6), 1184-96.
- Gao P, Liu, L, Mehta C (2013). Exact Inference for adaptive group sequential designs. *Statistics in Medicine*, 32(23):3991-4005
- Gao P, Liu, L, Mehta C (2014). Adaptive Sequential Testing for Multiple Comparisons. *Journal of Biopharmaceutical Statistics*, 24: 1035-1058
- Gart JJ, Nam J (1988). Approximate interval estimation of the ratio of binomial parameters: A review and corrections for skewness. *Biometrics* 44: 323-338.
- Goodman SN, Zahurak ML, and Piantadosi S (1995). Some practical improvements in the continual reassessment method for phase I studies. *Statistics in Medicine*, 14:1149-1161.

References

- Graubard BI, Korn EL (1987). Choice of column scores for testing independence in ordered $2 \times K$ contingency tables. *Biometrics*, 43: 471-476.
- Greenland S. (1991). On the logical justification of conditional tests for two-by-two contingency tables. *The American Statistician*. 45, 248:251.
- Greenwood, Jr. (1926). The Natural Duration of Cancer. *Reports of Public Health and Related Subjects*, Vol. 33, HMSO, London.
- Gu Kangxia, Ng Hon Keung Tony, Tang Man Lai, and Schucany William R. (2008). Testing the Ratio of Two Poisson Rates. *Biometrical Journal*, 50 (2008) 2, 283-298.
- Hajek P, Taylor TZ, and Mills P (2002). Brief intervention during hospital admission to help patients to give up smoking after myocardial infarction and bypass surgery: randomised controlled trial. *BMJ*, 324(7329), 87-89.
- Hauck WW, Preston PE and Bois FY (1997). A group sequential approach to crossover trials for average bioequivalence. *J. Biopharm. Statist.* 7, 87-96.
- Hauschke D, Steinijs VW, Diletti E, Burke M (1992). Sample size determination for bioequivalence assessment using a multiplicative method. *J. Pharmacokin. Biopharm.*, 20, 559-563.
- Hauschke D, Kieser M, Diletti E and Burke M (1998). Sample Size Determination for Proving Equivalence Based on the Ratio of Two Means for Normally Distributed Data. *Statistics in Medicine*, 18, 93-105.
- Haybittle JL (1971). Repeated assessment of results in clinical trials of cancer treatment. *Brit.J.Radiology*, 44, 793-797.
- Helen Brown, Robin Prescott (2006). *Applied Mixed Models in Medicine*. John Wiley & Sons, Ltd.
- Heinze G and Schemper M (2002). A solution to the problem of separation in regression. *Statistics in Medicine*, 21, 2409-2419.
- Hochberg Y (1988). A sharper Bonferroni procedure for multiple significance testing. *Biometrika*, 75, 800-802.
- Hochberg Y, Tamhane AC (1987). *Multiple Comparison Procedures*, Wiley, New York

- Hocking RR (1985). *The Analysis of Linear Models*. Belmont, CA: Brooks/Cole Publishing Co.
- Hodges JL, Lehmann EL (1963). Estimates of location based on rank tests. *The Annals of Mathematical Statistics*, 34: 598-611.
- Hollander M, Wolfe DA (1999). *Nonparametric Statistical Methods. second Ed.* John Wiley and Sons, New York.
- Holm S (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65-70.
- Hommel G (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75, 383-386.
- Hommel G (1989). A comparison of two modified Bonferroni procedures. *Biometrika*, 76, 624-625.
- Hosmer DW and Lemeshow S (2000). *Applied Logistic Regression*. Second Edition. Wiley, New York.
- Hsieh FY (1989). Sample size tables for logistic regression. *Statistics in Medicine*. 8: 795-802.
- Hutto C, Parks WP and Lai S (1991). A hospital based prospective study of perinatal infection with HIV-1. *J. Pdiatr.*, 118, 347-53.
- Hwang IK, Shih WJ, and DeCani JS (1990). Group sequential designs using a family of type I error probability spending functions. *Statistics in Medicine*, 9, 1439-1445.
- Iezzi R, Cotroneo A R, Giammarino A, Spigonardo F and Storto M L (2011). Low-dose multidetector-row CT-angiography of abdominal aortic aneurysm after endovascular repair. *European Journal of Radiology*, 79, 21-28.
- Jennison C and Turnbull BW (1989). Interim analyses: the repeated confidence interval approach (with discussion). *J.Roy.Statist.Soc.B*, 51, 305-361.
- Jennison C and Turnbull BW (1997). Group sequential analysis incorporating covariate information. *JASA*, 92, 1330-41.

References

- Jennison C and Turnbull BW (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman and Hall/CRC, London.
- Jennison, C and Turnbull, BW (2003). Mid-course sample size modification in clinical trial. *Statistics in Medicine*, 22, 971-993.
- Jennison, C and Turnbull, BW (2006). Adaptive and nonadaptive group sequential tests. *Biometrika*, 93(1), 1-21.
- Ji Y, Liu P, Li Y, and Bekele N (2010). A modified toxicity probability interval method for dose finding trials. *Clinical trials*, 7:653-656.
- Johnson and Wichern (1998) *Applied Multivariate Statistical Analysis.4th Edition*, Prentice Hall
- Jones B and Kenward MG (2003). *Design and analysis of cross-over trials*. Chapman and Hall/CRC, New York.
- Kalbfleisch JD and Prentice RL (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New Jersey.
- Kangxia Gu, Hon Keung Tony Ng, Man Lai Tang and William R. Schucany (2008). Testing the Ratio of Two Poisson Rates. *Biometrical Journal* 50 (2008) 2, 283-298
- Kapur A, Malik IS, et al (2005). The coronary artery revascularisation in diabetes (CARDia) trial: Background, aims, and design. *Am Heart J*, 149, 13-19.
- Keene Oliver N., Jones Mark R. K., Lane Peter W., and Anderson Julie (2007). Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. *Pharmaceutical statistics* 6, 89-97
- Kemeny N, Reichman B, Geller N and Hollander P (1988). Implementation of the group sequential methodology in a randomized trial in metastatic colorectal carcinoma. *Am J Clin Oncol*, 11, 66-72.
- Kendall MG, Stuart A (1979). *The Advanced Theory of Statistics*, 4th edition. Macmillan Publishing Co. Inc., New York.
- Keselman HJ, Algina J, Kowalchuk RK, Wolfinger RD (1998). A Comparison of Two

- Approaches for Selecting Covariance Structures in the Analysis of Repeated Measures. *Communications in Statistics-Computation and Simulation*, 27(3), 591:604.
- Kim K (1989). Point estimation following group sequential tests. *Biometrics*, 45, 613-17.
- Kim K and DeMets DL (1987). Confidence intervals following group sequential test in clinical trials. *Biometrics*, 43, 857-64.
- Kim K and Tsiatis AA (1990). Study duration for clinical trials with survival response and early stopping rule. *Biometrics*, 46, 81-92.
- Kimura D and Zenger (1997). Standardizing sablefish (*Anoplopoma fimbria*) long-line survey abundance indices by modeling the log-ratio of paired comparative fishing cpues. *ICES Journal of Marine Science*, 54, 48-59.
- Kolassa J (1995). A comparison of size and power calculations for the Wilcoxon statistic for ordered categorical data. *Statistics in Medicine*, 14: 1577-1581.
- Kontula KT, Anderson LC, Paavonen T, Myllyla L, Teerenhovi L, and Vuopio P (1980). Glucocorticoid receptors and glucocorticoid sensitivity of human leukemic cells *Int.J.Cancer*, 26:177-183.
- Kontula KT, Paavonen T, Vuopio P, and Anderson LC (1982). Glucocorticoid receptors in hairy-cell leukemia. *Int.J.Cancer*, 30:423-426.
- Krall JM, Uthoff VA, and Harley JB (1975). A Step-up Procedure for Selecting Variables Associated with Survival. *Biometrics*, 31: 49-57.
- Kreyszig E (1970). *Introductory Mathematical Statistics*. John Wiley & Sons, Inc., New York.
- Kutner M, Nachtsheim C and Neter J (2004). *Applied Linear Regression Models, 4th Edition*, IRWIN, Chicago.
- Laarman GJ, Suttorp MJ, Dirksen MT, Loek van Heerebeek, Kiemeneij F, Slagboom T, Ron van der Wieken, Tijssen JGP, Rensing BJ, and Patterson M (2006). Paclitaxel-Eluting versus Uncoated Stents in Primary Percutaneous Coronary Intervention. *NEJM*, 355, 1105-1113.

References

- Lachin JM (1977). Sample size determinations for rxc comparative trials. *Biometrics*, 33: 315-324.
- Lachin, JM (1981). Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials*, 2, 93-113.
- Lan, K., Hu, P., and Proschan, M. (2009). A conditional power approach to the evaluation of predictive power. *Statistics in Biopharmaceutical Research*, 1, 131-136.
- Lan KKG and DeMets DL (1983). Discrete sequential boundaries for clinical trials. *Biometrika*, 70, 659-663.
- Lan KKG and Wittes J (1988). The B-value: A tool for monitoring data. *Biometrics*, 44, 579-85.
- Lan KKG and Zucker D (1993). Sequential monitoring of clinical trials: the role of information and Brownian motion. *Stats. in Med.*, 12, 753-65.
- Landis JR, Koch GG (1977). The Measurement of interrater agreement for categorical data. *Biometrics*, 33: 159-174 .
- Laster L and Johnson M (2003). Non-inferiority trials: the ‘at least as good as’ criterion. *Stats. in Med.*, 22, 187-200.
- Lee ET (1992). *Statistical Methods for Survival Data Analysis*. John Wiley & Sons, New York.
- Lehmacher W and Wassmer G (1999). Adaptive sample size calculations in group sequential trials. *Biometrics*, 55, 1286-1290.
- Lehmann, EL (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco.
- Liebetrau, AM (1983). *Measures of association*, Sage Publications.
- Li G, Shih WJ, Xie T and Lu J (2002). A sample size adjustment procedure for clinical trials based on conditional power. *Biostatistics*, 3,2, 277-287.
- Li Lingling, Evans Scott, Uno Hajime, Wei L.J (2009) Predicted Interval Plots (PIPS): A Graphical Tool for Data Monitoring of Clinical Trials. *Statistics in*

Biopharmaceutical Research Vol 1, No.4 348-355

- Little RJA (1989). Testing the equality of two independent binomial proportions. *The American Statistician*, 43, 283-288
- Machin, D and Campbell, MJ (1987). *Statistical Tables for Design of Clinical Trials*, Blackwell Scientific Publications, Oxford.
- MacLaren, N.M (1989). The Generation of Multiple Independent Sequences of Pseudorandom Numbers. *Applied Statistics*, 38:351-359.
- Maindonald JH (1984). *Statistical Computation*. John Wiley & Sons, New York.
- Makuch RW, Parks WP (1988). Response of serum antigen level to AZT for the treatment of AIDS. *AIDS Research and Human Retroviruses* 4: 305-316.
- Mander AP and Sweeting MJ (2015). A Product of Independent beta Probabilities dose Escalation (PIPE) design for dual-agent Phase I trials. *Statistics in Medicine*, 34, 1261-1276.
- Mantel (1966) Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemother Rep.*, 50(3):163-70.
- Mantel N, Haenszel W (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 22:719-748.
- Marcus R (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63, 655-660.
- Martinez-Martin P, Valldeoriola F, Molinuevo JL, Nobbe FA, Rumia J, and Tolosa E (2000). Pallidotomy and quality of life in patients with parkinson's disease: An early study. *Movement Disorders*, 15(1), 65-70.
- Maurer W, Hothorn LA, Lehmacher W (1995). Multiple comparisons in drug clinical trials and preclinical assays: a priori ordered hypotheses. *Biometrie in der Chemisch-in-Pharmazeutischen Industrie*. 6, Vollman, J. (editor). Fischer-Verlag, Stuttgart, 3-18.
- Mehta, CR (2004). A Note on Standardizing $\hat{\delta}$ for Designs with Binomial Endpoints. *In House Technical Report*.

References

- Mehta CR, Liu L (2016). An objective re-evaluation of adaptive sample size re-estimation: commentary on 'Twenty-five years of confirmatory adaptive designs'. *Statistics in Medicine*, 35(3), 350-358
- Mehta CR, Patel NR (1986). A hybrid algorithm for Fisher's exact test on unordered $r \times c$ contingency tables. *Communications in Statistics*, 15:387-403.
- Mehta CR and Tsiatis AA (2001). Flexible sample size considerations using information-based interim monitoring. *Drug Information Journal*, 35, 1095-1112.
- Mehta CR, Bauer P, Posch M, Brannath W (2007). Repeated confidence intervals for adaptive group sequential trials. *Statistics in Medicine*, 26 (30), 5422-5433.
- Mehta CR and Pocock SJ (2011). Adaptive Increase in Sample Size when Interim Results are Promising: A Practical Guide with Examples. *Statistics in Medicine* 30(28): 3267-84
- Miettinen OS (1986). On the matched pairs design in the case of all- or- none responses. *Biometrics*, 24: 339-352.
- Miettinen, OS. and Nurminen, M (1985). Comparative Analysis of Two Rates. *Statistics in Medicine*, 4, 213-226.
- Miller AJ (1990). *Subset selection in Regression*. Chapman & Hall, London.
- Montgomery D, Peck E and Vining G (2001). *Introduction to Linear Regression Analysis, 3rd Edition*, Wiley, New York.
- Moseley JB, O'Malley K, Petersen NJ, et al. (2002). A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *New England Journal of Medicine*, 347, 81-8.
- Muller, H-H and Schafer, H (2001). Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential approaches. *Biometrics*, 57, 886-891.
- Naik UD (1975). Some selection rules for comparing p processes with a standard. *Communications in Statistics*. Series A. 4, 519-535.
- Nam J (1987). A simple approximation for calculating sample sizes for detecting linear

- trend in proportions. *Biometrics*, 43, 701-705.
- Neuenschwander B, Branson M, and Gsponer T (2008). Clinical aspects of the bayesian approach to phase I cancer trials. *Statistics in Medicine*, 27:2420-2439.
- Neuenschwander B, Matano A, Tang Z, Roychoudhury S, Wandel S and Bailey S (2015). A Bayesian Industry Approach to Phase I Combination Trials in Oncology. In: *Statistical Methods in Drug Combination Studies*, Zhao W, Yang H Ed. Chapman Hall/CRC Press, Boca Raton.
- Noether GE (1987). Sample size determination for some common nonparametric tests. *J. American Statistical Assoc.*, 82, 645-647.
- nQuery Advisor (2005). Software for Sample Size and Power estimation. Statistical Solutions, Saugus.
- O'Brien RG, Muller KE (1993). *Applied Analysis of Variance in Behavioral Science*. Marcel Dekker, New York. 8: 297-344
- O'Brien PC, Fleming TR (1979). A multiple testing procedure for clinical trials. *Biometrics*, 35, 549-56.
- O'Hagan A, Stevens JW, Campbell M J (2005). Assurance in clinical trial design. *Pharmaceutical Statistics*, 4(3), 187-201.
- Oliver N, Keene, Mark R. K. Jones, Peter W. Lane, Julie Anderson (2007). Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. *Pharmaceutical statistics*, 6, 89-97
- O'Quigley J, Pepe M, and Fisher L (1990). Continual reassessment method: A practical design for phase I clinical trials in cancer. *Biometrics*, 46:33-48.
- Overall JE, Doyle SR (1994). Estimating sample sizes for repeated measures designs. *Controlled Clinical Trials*, 15: 100-123.
- Owen, DB (1965). A Special Case of Bivariate Non-Central t-Distribution. *Biometrika*, 52, 3, 437-446.
- Pampallona S and Tsiatis AA (1994). Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favor of the

References

- null hypothesis. *J. Statist. Planning and Inference*, 42, 19-35.
- Pampallona S, Tsiatis AA and Kim K (1995). *Spending functions for type I and type II error probabilities of group sequential trials*. Technical report, Dept. of Biostatistics, Harvard School of Public Health, Boston.
- Pampallona S, Tsiatis AA and Kim K (2001). Interim monitoring of group sequential trials using spending functions for the type I and type II error probabilities. *Drug Information Journal*, 35, 1113-1121.
- Parker M, Puddey IB, Beilin LJ and Vandongen R. (1990). A 2-way factorial study of alcohol and salt restriction in treated hypertensive men. *Hypertension*, 16, 398-406.
- Patel HI (1983). Use of baseline measurements in the two-period cross-over design. *Communications in Statistics-Theory and Methods*, 12, 2693-712.
- Patterson S , Jones B (2006). *Bioequivalence and Statistics in Clinical Pharmacology*. Chapman & Hall/CRC, Taylor & Francis Group.
- Pearson et.al.(2003) Treatment effects of methylphenidate on cognitive functioning in children with mental retardation and ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43, 677-685.
- Phillips KE (1990). Power of the two one-sided tests procedure in bioequivalence. *Journal of Pharmacokinetics and Biopharmaceutics* 18.
- Pitt B, Zannad F, Remme WJ, Cody R, Castaigne A, Perez A, Palensky J, Wittes J (1999). The effects of spironolactone on morbidity and mortality in patients with severe heart failure. *New England Journal of Medicine*, 341, 10, 709-717.
- Pocock SJ (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64, 191-99.
- Posch M and Bauer P (1999). adaptive two stage designs and the conditional error function. *Biometrical Journal*, 41, 689-696.
- Posch Martin, Koenig Franz, Branson Michael, Brannath Warner, Dunger-Baldauf Cornelia and Bauer Peter (2005). Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Statistics in Medicine*;

24: 3697-3714.

Pregibon D (1981). Logistic Regression Diagnostics. *Ann. Statist.*, 9: 705-724.

Pritchett Y, Jemai Y, Chang Y, et al. (2011). The use of group sequential, information-based sample size re-estimation in the design of the PRIMO study of chronic kidney disease. *Clinical Trials*, 8(2), 165-174.

Proschan, MA and Hunsberger, SA (1995). Designed extension of studies based on conditional power. *Biometrics*, 51, 1315-1324.

Purich E (1980). *Bioavailability/Bioequivalency Regulation: An FDA Perspective in Drug Absorption and Disposition* (Ed. K.S. Albert), American Statistical Association and Academy of Pharmaceutical Sciences, Washington, D.C., 115-137.

Rabbee N, Coull BA, Mehta CR, Patel NR, Senchaudhuri P (2003). Power and sample size for ordered categorical data. *Statistical Methods in Medical Research*, 12, 73-84.

Reboussin DM, DeMets DL, Kim K, and Lan KKG (2002). Programs for computing group sequential boundaries using the Lan-DeMets method. SDAC, Dept.of Biostat. and Med.Informatics, University of Wisconsin Medical School.

Rencher AC (1995). *Methods of Multivariate Analysis*. John Wiley & Sons, New York

Ribas A, Hauschild A, Kefford R, Punt C, Haanen J, Marmol M, Garbe C, Gomez-Navarro J, Pavlov D and Marshall M (2008). Phase III, open-label, randomized, comparative study of tremelimumab (CP-675,206) and chemotherapy (temozolomide [TMZ] or dacarbazine [DTIC]) in patients with advanced melanoma. *Journal of Clinical Oncology*, 26(suppl), 485s, abstr LBA9011.

Robins J, Breslow N, Greenland S (1986). Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics* 42:311-323.

Rothman M, Li N, Chen G, Chi G, Temple R, Tsou HH (2003). Design and analysis of non-inferiority mortality trials in oncology. *Stats. in Med.*, 22, 239-264.

Ryan TP (1997). *Modern Regression Methods*. John Wiley & Sons, New York.

References

- Sabin T, Matcham J, Bray S, Copas A, and Parmar MKB. (2014). A quantitative process for enhancing end of Phase 2 decisions. *Statistics in Biopharmaceutical Research* 6:67-77.
- Santner TJ and Duffy DE (1989). *The Statistical Analysis of Discrete Data*. Springer-Verlag, New York.
- Santner TJ, Snell MK (1980). Small-sample confidence intervals for $p_1 - p_2$ and p_1/p_2 in 2×2 contingency tables. *Journal of the American Statistical Association* 75:386-394.
- Santner TJ, Yamagami S. (1993). Invariant small sample confidence intervals for the difference of two success probabilities. *Communications in Statistics, Part B – Simulation and Computation* 22:33-59.
- Sarkar, S. (1998). Some probability inequalities for ordered MTP2 random variables: a proof of Simes conjecture. *Annals of Statistics* 26, 494-504.
- Sarkar, S., and Chang, C. K. (1997). Simes' method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association*, 92, 1601-1608.
- Scharfstein DO and Tsiatis AA (1998). The use of simulation and bootstrap in information-based group sequential studies. *Stats. in Med.*, 17, 75-87.
- Scharfstein DO, Tsiatis AA, and Robins JM (1997). Semiparametric efficiency and its implication on the design and analysis of group-sequential studies. *JASA*, 92, 1342-50.
- Schoenfeld DA (1981). The asymptotic properties of comparative tests for comparing survival distributions. *Biometrika*, 68, 316-9.
- Schoenfeld DA (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*, 39, 499-503.
- Schoenfeld DA and Richter, JR (1982). Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. *Biometrics*, 38, 163-70.
- Schuirmann DJ (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J.*

Pharmacokin. Biopharm., 15, 657-680.

Schultz JR, Nichol FR, Elfring GL, and Weed SD (1973). Multiple stage procedure for drug screening. *Biometrika*, 29, 293-300.

Scott Patterson, Byron Jones (2006). *Interdisciplinary statistics Bioequivalence and Statistics in Clinical Pharmacology*. Chapman and Hall/CRC

Self SG, Mauritsen RH.(1988). Power and Sample size calculations for generalized linear models, *Biometrics*, 44, 79-86.

Self SG, Mauritsen RH, Ohara J (1992). Power calculation for likelihood ratio tests in generalized linear models. *Biometrics*, 48, 31-39.

Shein-Chung Chow, Jen-pei Liu (1998). *Design and Analysis of Clinical Trials Concepts and Methodologies*. Wiley series in Probability and Statistics.

Shen Y and Fisher L (1999). Statistical inference for self-designing clinical trials with one-sided hypothesis. *Biometrics*, 55, 190-197.

Sheskin DJ (2004). *Handbook of Parametric and Nonparametric Statistical Procedures* (3rd ed.). Chapman Hall/CRC Press, Boca Raton, FL.

Sidak Z (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62, 626-633.

Sidik K (2003). Exact unconditional tests for testing non-inferiority in matched-pairs design. *Statistics in Medicine* 22:265-278.

Siegel S, Castellan NJ (1988). *Nonparametric statistics for the behavioral sciences*. 2nd edition. McGraw-Hill, New York.

Simon R (1989). Optimal Two-Stage Designs for Phase II Clinical Trials. *Controlled Clinical Trials* 10:1-10.

Simon R, Rubinstein L, Arbuck SG, Christian MC, Freidlin B and Collins J (1997). Accelerated Titration Designs for Phase I Clinical Trials in Oncology. *Journal of the National Cancer Institute*, 89, 1138-1147.

Snapinn SM, Small RD (1986). Tests of significance using regression models for ordered categorical data. *Biometrics*, 42:583-592.

References

- Snedecor GW, Cochran WG (1989), *Statistical Methods*. 8th Edition, Iowa State University Press, Ames, IA.
- SPAF III Writing Committee for the Stroke Prevention in Atrial Fibrillation Investigators (1998). Patients With Nonvalvular Atrial Fibrillation at Low Risk of Stroke During Treatment With Aspirin: Stroke Prevention in Atrial Fibrillation III Study. *JAMA*, 1998;279:1273-1277.
- Spaulding C, Henry P, Teiger E, Beatt K, Bramucci E, Carrie D, Slama MS, Merkely B, Erglis A, Margheri M, Varenne O, Cebrian A, Stoll HP, DB Snead DB, Bode C (2006). Sirolimus-Eluting versus Uncoated Stents in Acute Myocardial Infarction. *NEJM*, 355, 1083-1104.
- Sprent P (1993). *Applied Nonparametric Statistical Methods*. 2nd edition. Chapman and Hall, New York.
- Steinijans VW, Hauck WW, Diletti E, Hauschke D, and Anderson S (1992). Effect of changing the bioequivalence range from (0.80, 1.20) to (0.80, 1.25) on the power and sample size. *Int J Clin Pharmacol Ther Toxicol*, 30, 571-575.
- Storer BE (1989). Design and analysis of phase I clinical trials. *Biometrics*, 45:925-937.
- Stroke Prevention in Atrial Fibrillation Investigators (1996). Adjusted-dose warfarin versus low-intensity, fixed-dose warfarin plus aspirin for high-risk patients with atrial fibrillation: Stroke Prevention in Atrial Fibrillation III randomised clinical trial. *Lancet*, 348, 633-38.
- Stout W, Marden J, Travers K. (1999). *Statistics: Making Sense of Data*. Mobius Communications.
- Suissa S, Shuster J (1985). Exact unconditional sample sizes for the 2×2 binomial trial. *Journal of Royal Statistical Society Series A* 148:317-327.
- Suissa S, Shuster J (1991). The 2×2 matched-pairs trial: Exact unconditional design and analysis. *Biometrics* 47:361-372.
- Sweeting M, Mander A, Sabin T. (2013). berm : Bayesian Continual Reassessment Method Designs for Phase I Dose-Finding Trials. *Journal of Statistical Software* 54(13): 1-26.

- Tarone , RE. (1985). On heterogeneity tests based on efficient scores. *Biometrika* 72(1): 91-95.
- Thomas RG, Conlon M (1992). Sample size determination based on Fisher’s exact test for use in 2x2 comparative trials with low event rates. *Controlled Clinical Trials*. 13: 134-147.
- Tim Friede and Heinz Schmidli (2010). Blinded sample size reestimation with count data: Methods and applications in multiple sclerosis. *Statistics in Medicine*. 2010, 29 1145–1156
- Tsiatis AA (1981). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika*, 68, 311-15.
- Tsiatis AA (1982). Group sequential methods for survival analysis with staggered entry. In: *Survival analysis* (eds. Crowley J and Johnson RA), Hayward, California: Institute of Mathematical Statistics, 257-68.
- Tsiatis AA and Mehta CR (2003). On the inefficiency of the adaptive design for monitoring clinical trials. *Biometrika*, 90, 367-378.
- Tsiatis AA, Rosner GL and Mehta CR (1984). Exact confidence intervals following a group sequential test. *Biometrics*, 40, 797-03.
- Tukey JW (1977). *Exploratory data analysis*. Addison-Wesley Publishing, Reading, MA.
- Upton GJG (1992). Fisher’s exact test. *J. R. Statist. Soc. Ser. A*, 155, 395-402.
- Van de Werf F (2006). Drug-Eluting Stents in Acute Myocardial Infarction. *NEJM*, 355, 1169-1170.
- Venzon DJ and Moolgavkar SH (1988). A method for computing Profile-Likelihood-Based Confidence Intervals. *Applied Statistics*, 37; 1, 87-94.
- Volberding PA, Lagakos SW, et. al. (1990). Zidovudine in asymptomatic human immunodeficiency virus infection. *New England Journal of Medicine*, 322:14, 941-949.
- Wald A, Wolfowitz J (1940). On a test whether two samples are from the same

References

- population. *Ann Math Stat* 11:147-162.
- Walter SD (1976). The estimation and interpretation of attributable risk in health research *Biometrics* 32:829-849.
- Wang SJ, Hung HMJ, Tsong Y, Cui L (2001). Group sequential strategies for superiority and non-inferiority hypotheses in active controlled clinical trials. *Statistics in Medicine*, 20, 1903-1912.
- Wang SK and Tsatis AA (1987). Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics*, 43, 193-99.
- Weight control and risk factor reduction in obese subjects treated for 2 years with Orlistat (1999). *JAMA*, 281, 235-42.
- Werner M, Tolls R, Hultin J, Mellecker J (1985). Sex and age dependence of serum calcium, inorganic phosphotrus, total protein, and albumin in a large ambulatory population. *Fifth International Congress on Automation, Advances in Automated Analysis*, 3, 59-65;
- Werner, M., Tolls, R. E., Hultin, J. V., and Mellecker, J(1970) Influence of sex and age on the normal range of eleven serum constituents. *Z. Klin. Chem. Klin. Biochem.* 8, 105-115 (1970).
- Westfall PH and Krishen A (2001). Optimally weighted, fixed sequence, and gatekeeping multiple testing procedures. *Journal of Statistical Planning and Inference*, 99, 25-40.
- Wiens B (2003). A fixed-sequence Bonferroni procedure for testing multiple endpoints. *Pharmaceutical Statistics*, 2, 211-215.
- Wiens B, Dmitrienko A (2005). The fallback procedure for evaluating a single family of hypotheses. *Journal of Biopharmaceutical Statistics*, 15, 929-942.
- Wilcoxon, F (1945). Individual Comparisons by Ranking Methods. *Biometrics*, 1, 80-83.

Index

A

- Accrual Model, 1614, 1627, 1639, 1652
- accrual time, 2310
- accrual versus study duration chart, 2308
- acute coronary syndromes, 1041
- adaptive design, 1035
- adaptive group sequential design, 1044
- adaptive re-design, 1226
- adaptive simulation settings, 1063
- adding a futility boundary, 1046
- adjusted inference, 2328
 - adjusted confidence interval, 1415, 2328, 2330
 - adjusted p-value, 427, 2328–2329
 - adjusted point estimate, 2330
 - inner wedge, 2331
 - no adjusted inference, 2332
 - ordering the sample space, 2328
- admissible design, 774
- Agreement, 649
- allocation ratio
 - equal, 117
 - unequal, 203
- alpha spending function, 1471, 2315
- Alzheimer’s disease clinical trial, 242
- Analysis-Descriptive Statistics, 1827
- analysis, 1810
- Analysis of Variance
 - two-way, 1841
- analysis
 - Case Data Editor, 1810
 - crossover, 1919, 1923, 1934, 1939, 1950, 1954
 - equivalence, 1946, 1950, 1954
 - non-inferiority, 1929, 1934, 1939
 - noninferiority, 1901
 - one-sided test, 1890
 - paired data, 1901
 - ratio of means, 1903, 1910, 1915, 1929, 1946
 - superiority, 1890, 1919, 1923
 - Wilcoxon Signed Rank Test, 1898
- binary endpoint, 2060
- ANOVA, 235, 1982
 - one-way, 2547
 - one way, 232, 1976
 - repeated constant correlation, 235, 1982
 - two-way, 2549
 - two way, 237, 1985
 - two-way, 1841
- arbitrary error probability, 156
- Area, 1861
- ASN-average sample number, 2305
- assurance, 174, 981
- asymmetric two-sided boundaries, 1481, 1484
- attained significance level, 726
- auto-hide, 11
- availability of adaptive features, 1022

B

- Backward Image Confidence Interval, 1221
 - example, 1272
- balanced randomization, 395
- Bar Plot, 1855
- Barnard’s unconditional test, 2570
- Bayesian, 174, 981, 2151
- Bayesian predictive power, 177
- benefits of adaptive designs, 1028

Index

- benefits of simulating adaptive design, 1167
 - Berger-Boos correction, 2589
 - beta spending function, 1471, 1484, 2294, 2315
 - binomial design, 2298
 - pooled variance, 395, 427
 - pooled vs unpooled, 394, 2298
 - unpooled variance, 395, 427
 - binomial distribution, 394
 - binomial endpoint, 474–475, 751–752, 1359
 - Binomial Endpoint Analysis, 2069
 - Non-Inferiority Trial, 2088
 - Superiority Trial, 2069
 - Binomial Endpoint
 - Create Multiple Designs, 356
 - Fixed Sample Design, 350
 - Group Sequential Design, 352
 - Interim Monitoring, 359
 - Simulations, 357
 - binomial endpoint
 - unknown baseline, 1401,
 - binomial study design, 1401
 - information based, 1401
 - Binomial Superiority Regression, 644
 - binomial
 - equivalence exact test, 767
 - exact, 714
 - chi square statistic, 395
 - Simon’s two-stage design, 774
 - superiority
 - exact, 714,
 - one-sample, 714
 - two-sample equivalence test, 767
 - two-sample exact tests, 736
 - two-sample tests, 736
 - bivariate log-normal, 121, 136
 - bivariate normal, 113, 121, 136
 - bivariate t, 129, 212
 - Blyth-Still-Casella intervals, 2565
 - Bonferroni procedure, 252, 577, 2031, 2181
 - survival, 2243
 - boundaries, 2286, 2293, 2295
 - asymmetric, 1481, 1484
 - early rejection of H₀, 2293
 - early rejection of H₀ or H₁, 2294–2295
 - early rejection of H₁ only, 2297
 - Generalized Haybittle-Peto, 2286
 - Haybittle-Peto, 2286
 - Pampallona-Tsiatis, 2288
 - spending function, 2293, 2295
 - Wang-Tsiatis, 2287
 - boundary chart, 147
 - p-value scale, 148
 - boundary crossing probability, 2352
 - boundary scale, 148
 - boundary scales
 - conditional power scale, 484
 - boundary shape parameter, 2348
 - O’Brien-Fleming, 145
 - Pocock, 149
 - Box Plots, 1863
 - Bubble Plots, 1864
 - BWCI, 1221
 - example, 1272
 - by-passing test statistic calculator, 2332
- C**
- calendar time, 1402, 2280
 - CAPTURE clinical trial, 23, 395, 1623
 - case data editor, 55
 - CDL method, 1021, 1160
 - acute coronary syndromes example, 1171
 - alpha preservation, 1177
 - binomial endpoint example, 1171
 - comparison of group and adaptive sequential designs, 1176

- group sequential design example, 1172
- normal endpoint example, 1162
- preservation of type-1 error, 1177
- Schizophrenia example, 1162
- simulation parameters, 1164
- chart of accrual versus study duration, 2308
- chart of conditional power, 2322
- chart of post-hoc power, 2321
- chart of stopping boundaries, 147
- Chen DeMets and Lan method, 1021, 1160
- chi-square statistic, 394–395, 2300
- CHIRAND, 70
- choice of variance, 2310
- CHW method, 1021, 1055
 - acute coronary syndromes example, 1093
 - IM, 1089, 1152
 - preservation of type-1 error, 1058
 - repeated confidence intervals, 1058
 - adaptive design example, 1077
 - adding a futility boundary, 1105
 - binomial endpoint example, 1093
 - calculation of repeated p-values, 1060
 - comparison of adaptive design to fixed sample design, 1087
 - conditional power, 1061
 - fixed sample study example, 1076
 - interim monitoring sheet, 1089, 1152
 - normal endpoint example, 1074
 - operating characteristics of adaptive group sequential design, 1100
 - repeated p-values, 1058
 - schizophrenia example, 1074
 - simulation results by zone, 1085
 - statistical theory, 1056
- CHW simulation Input, 1081, 1099
- CHW Simulation
 - assumptions in East, 1062
- CHW Statistic, 1058
- class of spending functions, 2347
- Classification table, 2170–2171, 2179, 2638
- coefficient of variation, 110, 137, 199–200, 217, 228, 1915, 1929
- Cohen’s Kappa, 649
- Collinearity diagnostics, 2552
- column functions, 69
- combined efficacy and futility, 1444
- comparing survival curves, 2219
- comparison of adaptive and non-adaptive group sequential designs, 1103
- comparison of CHW
 - CDL and conventional Wald, 1213
- Comparison of Designs, 25
- comparison of multiple comparison procedures - Analysis, 2207
- comparison of
 - fixed sample and adaptive design, 1039
 - group sequential and fixed sample design, 1034
- Completers Prediction Plot, 1622, 1631
- computation of boundaries, 2352
- computing boundaries for the exact group sequential test, 2608
- computing conditional power for a pre specified sample size, 1361
- computing conditional power for specified number of events, 1368

Index

- computing conditional power for
 - specified sample size, 1354
- computing number of events (overall), 1386
- computing number of events for specified conditional power, 1370
- computing sample size for desired conditional power, 1357
- computing the required sample size increase, 1038
- Conditional exact test, 2139
- conditional power, 196, 424, 484, 2321–2322, 2324
- conditional power at ideal next look position, 2321
- conditional power calculator, 1361,
- conditional power calculator, 1107, 1350
 - in simulation, 1371
 - interim monitoring, 1351, 1366
- conditional power chart, 2322
- conditional power for decision making, 1350,
- conditional power, 1061
 - boundary scale, 484
 - ideal next look position, 2321
 - informal use of, 1441
 - one sided tests, 2322
 - stopping for futility, 2321
 - target, 1082
 - two sided, 1061
 - two sided tests, 2322
- conditional rejection probabilities, 1222
- confidence interval, 422, 424, 427, 1415
- Confidence interval
 - Clopper-Pearson method, 2565
 - equivalence test, 2576
- confidence interval
 - adjusted, 422, 424, 427, 1415
- Confidence Intervals, 2185, 2188, 2191, 2194, 2197, 2199, 2203, 2206
- Plot, 2185, 2188, 2191, 2194, 2197, 2199, 2203, 2206
- conjunctive power, 248, 251, 253–255, 259–262, 586, 590, 592, 594–595, 597, 599, 1011, 1014, 1017
 - survival, 1005, 1007, 1009
- conservative futility boundaries, 1453
- conservative spending function, 1471,
- Correlations, 1843
 - Kendall’s Tau, 1843
 - Pearson’s Correlation, 1843
 - Spearman’s Rho, 1843
- count data, 790
- CP calculator, 1350
- Cramer’s V, 2598
- create Variable, 56
- Cross Tabulation, 1832
- crossover data editor, 61
- crossover design, 206, 208, 221, 227, 1918, 1933, 1950, 1953, 1965
 - example, 207–208, 223, 228
 - ratio of means, 208, 227, 1953
 - simulation, 230
 - single look, 223, 229
- Crossover Plots, 1879
- crossover
 - analysis, 1919, 1923, 1934, 1939, 1950, 1954
 - equivalence, 1950, 1954
 - example, 1919, 1923, 1934, 1939, 1950, 1954
 - non-inferiority, 1934, 1939
 - superiority, 1919, 1923
- Cui Hung and Wang method, 1021, 1055

- cumulative information, 1413,–1414
- Cumulative Plot, 1870, 1872
 - left, 1870
 - Right, 1872
- CV, 200
- D**
- Data Exploration Plots, 1854
- Data Set
 - SMALLT, 2111
 - SWINE, 1841
 - Root, 1852
 - WERNER, 1848
- Data Sets
 - VACCINE, 2097
- Data sets
 - BD, 2081
 - Gupta, 2127
 - OESOPHAGEAL, 2081
 - VARI, 1870
- data transformation, 63
- data
 - Cancer, 2220, 2222, 2224, 2231, 2233
- dataset
 - CrossoverCaseData, 1919
 - Iris, 1941
 - Methylphenidate, 1890
 - Olestra, 1901
 - pkfood, 1934, 1950
- decision making, 1350
- deleting a design, 117
- deleting observations, 1943
- departure from design, 2318
- Descriptive Statistics
 - central tendency, 1827
 - coefficient of variation, 1827
 - count, 1827
 - dispersion, 1827
 - geometric mean, 1827
 - harmonic mean, 1827
 - kurtosis, 1827
 - maximum, 1827
 - mean, 1827
 - median, 1827
 - minimum, 1827
 - mode, 1827
 - skewness, 1827
 - standard deviation, 1827
 - standard error of mean, 1827
 - sum, 1827
 - variance, 1827
- design menu, 23
- design
 - many means, 232, 1976
 - sample size rounding, 16, 76, 347, 711, 787, 823, 1024, 1390, 1803
 - information based, 1394
 - Poisson, 1416, 1423
- designing a study given limited data, 1029
- designing the primary trial, 1226
- difference of means for crossover data, 171
- difference of proportions exact tests, 736
- disjunctive power, 248, 251, 253–255, 259–262, 586, 589, 591, 594–595, 597, 599, 1011, 1014, 1017
 - survival, 1005, 1007, 1009
- Dose-Finding Hypertension Trial, 2025
- dose response curve, 582
 - linear, 244
 - logistic, 582
- drawback of single group sequential design, 1043
- drift parameter, 141, 2275–2278, 2282–2283
- Dropout Prediction Plot, 1622, 1632

Index

- Dropout Predictions Plot, 1726, 1736, 1754, 1762, 1768
- dropout rates, 872
 - specifying the parameters, 994
- dropouts, 872, 994
- Dropouts Prediction Plot, 1644, 1657
- Dropouts Predictions Plot, 1747, 1779, 1786
- Dunnett procedure, 241–242
- Duration-Accrual chart, 871
- E**
- early rejection H1, 2297
- early rejection of H0, 2293
- early rejection of H0 or H1, 2294–2295
- early stopping, 157, 394, 412
- early stopping for efficacy, 1439
- early stopping for futility, 1484, 2294
- early stopping
 - for benefit or futility, 412
 - for futility, 1456
- east workbook, 28
- edit data, 58
- edit simulation, 135, 226
- editing simulation, 120
- effect size, 1393, 1425, 1929
 - equivalence, 1945
- efficacy boundaries, 1444, 1471
- efficacy trial, 23, 395
- efficiency considerations for CDL, 1213
- efficient estimators, 2272
- Egret Siz, 1424
- Eliminating nuisance parameters, 2570
- Enrollment Plan, 1615, 1627, 1639, 1652
- Enrollment Prediction Plot, 1620, 1631, 1643, 1656, 1686, 1693, 1700, 1714
- enrollment range, 2308
- Enrollment/Events Prediction, 1609, 1675
 - At Design Stage, 1609
 - At Interim Monitoring Stage, 1658
- Enrollments Prediction Plot, 1746, 1753, 1778, 1785
- equivalence, 128, 211, 1907
- Equivalence example
 - binomial difference, 2106, 2108
- equivalence limits, 128, 211, 221
- equivalence testing of two binomials
 - power of, 2635
- equivalence testing of two independent binomials
 - power of, 2617
- equivalence
 - analysis, 1946, 1950, 1954
 - crossover, 1950, 1954
 - crossover design, 221, 227, 1950, 1953
 - example, 130, 137, 212, 217, 223, 228, 1908, 1941, 1946, 1950, 1954
 - normal, 211
 - paired data, 128, 136, 1907, 1910
 - power calculation, 129, 212
 - ratio of means, 136, 217, 227, 1910, 1953
 - simulation, 134, 139, 216, 220, 230
 - single look, 213, 218, 223, 229
 - test of hypothesis, 128, 211, 221
- error spending chart, 425
- error spending function
 - interim monitoring, 2313
- error spending functions, 413
- evaluating the BWCI and RCI Methods
 - by Simulation, 1272
- Events Prediction Plot, 1643, 1656, 1725, 1735, 1754, 1761, 1767
- Exact Conditional Test, 2613
- Exact confidence interval
 - difference of binomials, 2576
 - ratio of binomials, 2581, 2586
- exact power, 2617–2618

- Exact test, 2139
 - exact test
 - integer sample size, 709
 - exact unconditional power, 2617
 - exact
 - time limit option, 16, 76, 347, 711, 787, 823, 1024, 1390, 1803
 - example odds Ratio of proportions, 2103, Example, 2069
 - example
 - bioequivalence, 217
 - crossover, 1919, 1923, 1934, 1939, 1950, 1954
 - crossover design, 228
 - equivalence, 130, 137, 212, 217, 228, 1908, 1941, 1946, 1950, 1954
 - kappa, 2218
 - non-inferiority, 114, 122, 200, 1929, 1934, 1939
 - noninferiority, 1901
 - Odds ratio of proportions, 2078
 - one-sided test, 91, 1891
 - paired, 102, 114, 130, 1892, 1898, 1908
 - paired data, 122, 137
 - ratio of means, 122, 137, 200, 217, 228, 1915, 1923, 1929, 1939, 1946, 1954
 - repeated measure regression, 2006
 - superiority, 91, 102, 1890, 1892, 1915, 1919, 1923
 - Wilcoxon Signed Rank Test, 1898
 - acute coronary syndromes, 1041
 - Example
 - Alcohol and oesophageal cancer, 2081
 - Animal Toxicology Example 1, 2122
 - binomial test, 2060
 - Clinical Trial Data, 2069
 - McNemar test, 2065
 - example
 - negative symptoms Schizophrenia, 1030
 - Odds ratio of proportions, 2103
 - Example
 - Oral lesions data, 2127
 - pilot study for a new drug, 2060
 - example
 - two binomials, 1394
 - Example
 - Voters Preference, 2065
 - exit probabilities, 1426–1427
 - exit probability, 2304
 - expected events, 2305
 - expected information, 1394, 2304–2305
 - expected number of events, 2305
 - Expected response, 2179
 - expected sample size, 2304–2305
 - expected stopping time, 2304
 - binomial, 2304
 - normal, 2304
 - survival, 2305
 - exponential failure, 2308
 - expression builder, 64
 - Extended CDL method, 1160
 - acute coronary syndromes example, 1204
 - cut-off points for some typical two stage designs, 1197
 - necessity of CDL criteria, 1208
 - preservation of type-1 error, 1201, 1207
 - Schizophrenia example, 1198
 - extension of CDL method, 1191
- F**
- failure-time trials, 820
 - failure rate, 2308

Index

- Fallback, 2204
 - Fallback Procedure, 2203
 - fallback procedure, 260, 596, 1015, 2046, 2258
 - Fallback
 - Proportion of Alpha, 2204
 - Test Sequence, 2204
 - favorable zone, 1037
 - FDA data, 2122
 - animal toxicology, 2122
 - FDA Guidance on Adaptive Designs, 1020
 - filter designs, 27
 - filter variable, 60
 - Final look, 1414
 - first interim analysis, 1428
 - first interim look, 421
 - Fisher's exact test, 2069, 2618
 - 2 X 2 table, 2069
 - power of, 2617
 - Fisher exact test, 2597
 - Fisher information, 2271–2272
 - fixed accrual, 821
 - fixed follow-up designs, 2310
 - fixed sample design, 142, 1031
 - Fixed Sequence Procedure, 2200
 - fixed sequence procedure, 593
 - fixed sequence testing, 258, 2044
 - survival, 1012, 2256
 - fixed study duration, 821
 - flexibility of adaptive approach, 1053
 - flexible clinical trial, 1027
 - flexible interim monitoring, 2314
 - flexible monitoring, 2314
 - flexible stopping boundaries, 1460
 - follow-up, 2308, 2310
 - fixed, 2310
 - variable, 2308
 - follow up time, 2310
 - four parameter logistic curve, 581
 - Frequency Distribution, 1829
 - futility boundaries, 481, 1471, 2294, 2296–2297
 - overruling, 2296
 - futility boundary, 774, 1484
 - non-binding, 1486
 - overruling, 1486
 - futility stopping, 1441
 - futility stopping boundaries, 1444
 - conservative, 1453
 - FWER, 240, 577, 999, 2024, 2180
 - survival, 2240
- G**
- G vs I designs, 2303
 - gamma spending function, 2291
 - general design module, 2302
 - general design
 - Poisson data, 1426–1427
 - general designs, 1423
 - general distribution, 2332
 - interim monitoring, 2332
 - Generalized Haybittle-Peto boundaries, 2286
 - geometric mean, 137
 - getting started with East 6, 7
 - Analysis Menu, 21
 - Data Editor menu, 19
 - Design menu, 23
 - Home menu, 12
 - Interim Monitoring, 49
 - user interface, 12
 - workflow, 8
 - global power, 248, 253, 586–587, 590, 592, 1005, 1011
 - Gm spending function, 2291
 - Goodman-Kruskal Gamma, 1844
 - group sequential, 94, 103, 188
 - group sequential design, 145, 476, 1031, 1042, 1439, 2271
 - group sequential
 - paired data, 103

- superiority, 103
- guaranteed alpha error, 2315
- Gupta data set, 2127
- H**
- H0-H1 boundaries, 2294–2295
 - two-sided, 2296
- H0-only boundaries, 2293
- H1-only boundaries, 2297
- Hat matrix, 2554
- Haybittle-Peto boundaries, 1462, 1466, 2286
- help panel, 11
- hiding help panel, 11
- Histogram, 1874
- Hochberg's Step Up, 2194
 - Analysis, 2194
- Hocheberg procedure, 256, 592, 1010, 2039, 2194
 - survival, 2251
- Holm's Step Down, 2191
 - Analysis, 2191
- Holm step-down procedure, 254, 591, 2038, 2191, 2250
 - survival, 1007
- Hommel's Step Up, 2197
 - Analysis, 2197
- Hommel procedure, 256, 592, 1010, 2039, 2194, 2251
- Homogeneity test
 - examples, 2081
- Horizontal Bar Plot, 1857
- Horizontal Stacked Bar Plot, 1859
- hypergeometric variance, 2280
- hypothesis H1/2, 2304
- Hypothesis test
 - multiparameter, 2139
- hypothesis testing, 1222
- I**
- I vs G designs, 2303
- IF() function, 67
- impact of futility boundary, 1112
- implementing the adaptive changes
 - through a secondary trial, 1238
- incremental Wald statistics, 1056
- independent increments, 2352
- inflation factor, 1401, 1423, 2301
- Influence statistics, 1847
- Influential groups, 2179
- informal use of conditional power, 1441
- information based design, 1027, 1394–1395, 1401, 1409, 1428, 2302
- information based inference, 1394
- information based monitoring, 1401, 1428, 2333
 - sample size re-estimation, 2333
- information based
 - Poisson, 1416
 - stroke study, 1418
- information fraction, 1402, 1404–1406, 1413, 1431, 2274, 2320
- information measures, 2304
- information vs sample size, 2303
- inner wedge, 2296
- inner wedge boundaries, 2289
- inner wedge stopping boundaries, 2331
- input multiple values, 116
- interim monitoring, 491–492, 504–505, 519–520, 531–532, 727, 947, 1401, 1412, 1419, 1428, 2313
 - first look, 99, 107, 196
 - non-inferiority, 195
 - paired data, 106
 - second look, 108
 - superiority, 106
- error spending function, 2313

Index

- flexible, 2314
 - information based, 1394, 2333
 - Lan-DeMets, 2314
 - preserving alpha, 2314
 - theory, 2313
 - introduction to survival endpoint, 820
 - invoking CDL simulation, 1163
 - invoking CHW simulation, 1081, 1099
- J**
- Jennison and Turnbull, 1220, 2271
 - Jennison Turnbull theorem, 2284, 2303
- K**
- Kalbfleisch and Prentice, 900
 - Kendall's Coefficient of Concordance, 1844
 - Kendall's Tau, 2551, 1843
 - key advantage of adaptive plan, 1040
- L**
- Lan-DeMets, 97, 103, 190
 - Lan-DeMets spending function, 400–401, 1418, 1471, 2291, 2314
 - O'Brien-Fleming flavor, 400, 2291
 - Pocock flavor, 401, 2291
 - Lan-DeMets
 - extension to preserving beta, 2315
 - interim monitoring, 2314
 - last look, 2317
 - optimal placement, 2319–2320
 - recomputation of boundary, 2318
 - LD(OF) spending function, 2291
 - LD(PK) spending function, 2291
 - Left Cumulative Plot, 1870
 - Lehmacher and Wassmer, 1055
 - Likelihood ratio test, 2139
 - limitation of group sequential design, 1097
 - limitations of CDL method, 1160
 - linear mixed model ratio of means
 - crossover, 2023
 - log rank test, 2220, 2222, 2224, 2231, 2233
 - log transformation, 200, 217, 228, 1915, 1923, 1929, 1938, 1953
 - log transformed data, 200, 217, 228, 1953
 - logistic dose response curve, 581
 - Logistic Regression, 644
 - logistical issues, 1054
 - lognormal data, 110, 121, 136, 208, 217, 227, 1903, 1910, 1915, 1929, 1953
 - logrank score statistic, 2280
 - logrank statistic, 2280
 - long-term mortality, 820
 - loss of efficiency using CHW method, 1160
 - lower stopping boundary, 2288, 2294
- M**
- making adaptive changes to the primary trial, 1234
 - MAMS Designs, 285
 - Continuous Endpoint, 285
 - Many Proportions, 549, 2111
 - marginal power, 248, 251, 586
 - maximum events, 2282–2283, 2305
 - maximum failures, 2305
 - maximum information, 1394, 1412, 2278, 2282–2283, 2289, 2305
 - Maximum likelihood, 2138
 - Maximum likelihood estimates, 2580
 - Maximum likelihood
 - non-convergence, 2138
 - maximum sample size, 1412, 2278, 2289, 2305
 - maximum study duration, 2306
 - maximum usable sample size, 1082

- McNemar's, 729
- McNemar's conditional
 - exact test, 729
- McNemar's test, 2611
 - power, 2611
- McNemar's Test
 - power, 2612
- McNemar's
 - McNemar's conditional test, 714
- McNemar probabilities, 731
- McNemar test, 2566
- Measures of agreement, 2216
 - Cohen's Kappa, 2216
 - weighted Kappa, 2216
- median unbiased estimate, 2330
- menus in East 6, 12
- midcourse change in desired power, 1029
- minimax design, 774
- minimum usable sample size, 1082
- missing value code, 70
- missing values, 57
- MLE non-convergence, 2138
- Model Terms item, 2133, 2143
- monitoring for general distributions, 2332
- monitoring the primary trial, 1228
- Monte Carlo accuracy, 164
- motivation for adaptive sample size
 - changes, 1027
- multiple comparison
 - Bonferroni procedure (weighted), 577, 588
 - Dunnett procedure (step-down), 242, 250, 2028
 - Dunnett procedure, 241
 - fallback procedure, 596
 - p-value based procedures, 251, 2030
 - Sidak procedure, 590
- Muller and Schafer method, 1022, 1221
- Muller and Schafer Method: Interim Monitoring, 1327
- multi-arm trials, 2272
- Multicollinearity Criterion, 2553
- Multinomial distribution, 2111
- Multinomial probabilities, 2111
- multiple comparison procedures
 - survival, 1011
- multiple comparison survival
 - Hocheberg procedure
 - survival, 2251
- multiple comparison
 - Bonferroni procedure (weighted), 252, 2031, 2181
 - survival, 2243
 - Bonferroni procedure, 252, 2031, 2181
 - survival, 2243
 - Dunnett procedure (step-up), 250, 2029
 - Dunnett procedure, 242
 - fallback, 260, 1015, 2046
 - survival, 2258
 - fixed sequence, 258, 2044
 - fixed sequence procedure, 593
 - fixed sequence
 - survival, 1012, 2256
 - Hocheberg procedure, 256, 592, 2039, 2194
 - survival, 1010
 - Holm step-down procedure, 254, 2038, 2191, 2250
 - survival, 1007
 - Hommel procedure, 256, 592, 2039, 2194
 - survival, 1010, 2251
 - parametric procedures, 241
 - Sidak procedure, 252, 2031, 2033, 2181
 - survival, 2243, 2246
 - weighted Bonferroni, 2035
 - survival, 2247
- Multiple Designs, 25

Index

- Multiple Discrete Endpoints, 601
- Multiple Endpoints, 265
- Multiple Linear Regression, 1847, 2552
- multiple look, 94, 103, 188
 - paired data, 103
 - superiority, 103
- multiple values in input field, 116
- Multivariate Analysis of Variance, 1851
- Multivariate statistics, 1851
- Mutiple discrete endpoints gatekeeping, 601
- Mutliple endpoints gatekeeping, 265
- N**
- navigator panel, 8
- necessity of CDL constraint, 1170, 1180
- negative symptoms Schizophrenia, 1030
- new crossover data, 61
- no early stopping, 1436
- nominal critical point, 424
- non-binding futlity boundaries, 2296–2297
- non-binding futlity boundary, 1105, 1486
- non-central t, 129, 212
- Non-convergence of maximum likelihood, 2138
- non-inferiority, 113, 185, 474, 751, 1901, 1926
- non-inferiority and survival, 2283
- non-inferiority boundaries, 1471
- non-inferiority margin, 113, 1901
- non-inferiority
 - analysis, 1929, 1934, 1939
 - crossover, 1934, 1939
 - crossover design, 206, 208, 1918, 1933, 1965
 - exact, 752
 - example, 114, 122, 185, 200, 207–208, 1929, 1934, 1939
 - group sequential, 188
 - interim monitoring, 195
 - multiple look, 188
 - one sample, 1926
 - paired data, 113, 121, 1901, 1903
 - ratio of means, 121, 208, 1903, 1929
 - simulation, 119, 126, 193, 205
 - single look, 114, 186
 - t test, 118, 198, 204
 - test of hypothesis, 113, 1901
- binomial, 474, 751–752
- normal, 185, 1926
- power of, 2617
- Non-inferiority
 - Simulation, 884
- Non-Proportional Hazards, 884, 887, 894
 - group sequential, 894
 - Single-Look, 887
- Noninferiority test example
 - binomial difference, 2088, 2093, 2097, 2100
 - binomial ratio, 2097
- noninferiority
 - analysis, 1901
 - example, 1901
- Noninferiority
 - example, 2103
- noninferiority survival curves, 2228
- Normal Endpoint, 79
- normal endpoint, 91, 185, 211, 1890, 1913, 1926
- Normal Endpoint
 - Create Multiple Designs, 84
 - Fixed Sample Design, 79
 - Group Sequential Design, 81
 - Interim Monitoring, 88
 - Simulations, 86
- normal response, 91, 141, 185, 211, 1890, 1913, 1926
- NORMRAND, 70

- nuisance parameter, 2276, 2278
 Nuisance parameter elimination, 2573, 2579, 2585, 2588–2589
 nuisance parameters, 2303
- O**
 O’Brien-Fleming, 94, 97, 103, 190
 O’Brien-Fleming boundaries, 2291
 objections against CHW statistics, 1160
 observational study, 332
 observational vs experimental, 332
 odds Ratio of proportions, 2078
 Odds ratio of proportions
 example, 2078
 example, 2103
 Odds ratio output, 2136
 oncology trial, 714
 single arm, 714
 ONCOX trial, 1647, 1666
 one-sided test, 1447
 analysis, 1890
 example, 91, 1891
 one sample, 91, 1890
 exact, 714
 simulation, 98
 superiority, 91, 1890, 2060
 t test, 100
 one sided Pampallona-Tsiatis boundaries, 2288
 one sided
 simulation, 98
 one way ANOVA, 232, 1976
 operating characteristics of adaptive design, 1039
 operating characteristics of adaptive group sequential design, 1044
 operational issues, 1054
 optimal design, 774
 optimal placement of last look, 2317, 2319
 Options
 odds ratio, 2136
 ordering the sample space, 2328
 adjusted inference, 2328
 stage-wise, 2329
 Orlistat Trial, 1609
 overruling a futility boundary, 1486
 overruling futility boundaries, 2296
- P**
 p-value boundaries, 1462, 1466
 p-value scale, 148
 p-value
 adjusted, 427
 paired data, 101, 110, 113, 121, 128, 136, 1901, 1903, 1907, 1910
 analysis, 1901
 equivalence, 128, 1907
 example, 122, 137
 group sequential, 103
 interim monitoring, 106
 multiple look, 103
 non-inferiority, 113, 1901
 ratio of means, 110, 121, 136, 1903, 1910
 simulation, 105, 119, 126, 134, 139
 superiority, 101
 t test, 109, 118
 test of hypothesis, 101, 113, 128, 1901
 paired sample
 single look, 102, 114
 paired
 example, 102, 114, 130, 1892, 1898, 1908
 Pampallona-Tsiatis boundaries, 1469–1470, 2288
 one sided, 2288

Index

- Pancreatic cancer Trial, 1289
- parameter estimation, 1223
- parameters for group sequential design, 1033
- Parkinson's disease example, 1266
- patient accrual, 2271
- patient follow-up time, 2280
- Pearson's Contingency Coefficient, 2598
- Pearson's Correlation, 1843
- Pearson's Product-Moment Correlation, 1844
- Pearson chi-square test statistic, 394, 2300
- Pearson residual, 2179
- penultimate look, 2317
 - optimal placement, 2317
- PET, 2615
- phase III study, 1020
- PhaseII-III Designs
 - Theory, 2394
- PI, 70
- Pie Plot, 1860
- PIP, 1575
 - binomial data, 1592
 - continuous outcome data, 1603
 - normal data, 1603
 - survival data, 1575
 - Time to Event Data, 1575
- planned number of looks, 143, 478
- plot
 - conditional power, 196
 - general (user defined), 126, 139
 - power vs delta, 118, 134
 - power vs sample size, 118, 125, 133, 204, 215
 - power vs treatment effect, 118, 134
 - sample size vs sd of log ratio, 126, 139
- Plot
 - left cumulative, 1870
 - Right cumulative, 1872,
- Plots, 1854
 - horizontal stacked bar, 1859
 - PP Normal, 1867
 - QQ Normal, 1868
 - Simple bar, 1855
 - area, 1861
 - bar, 1855
 - box, 1863
 - bubble, 1864
 - data exploration, 1854
 - histogram, 1874
 - horizontal bar, 1857
 - pie, 1860
 - predictive interval, 1575
 - scatter, 1865
 - stacked bar, 1856
 - stem and leaf, 1875
 - step function, 1877
- Pocock boundaries, 2291
- Poisson, 1416
 - poisson endpoint, 790
 - Poisson endpoint, 1416, 1423
 - poisson
 - one sample, 790
- Poisson
 - information based, 1416
 - risk ratio, 1416
- pooled binomial design, 2279
- pooled designs, 427
- pooled estimate, 394, 2298
- pooled variance, 395, 427, 2298
 - binomial design, 395, 427
- pooled vs unpooled, 395, 427
- pooled
 - binomial, 394, 2298
- Post-fit file
 - classification table and, 2170, 2179
 - DEVTOX data, 2170
 - Seropos data, 2179
- post-hoc power, 2315, 2317–2318
- post-hoc power chart, 2321

- post-hoc power
 - calculations, 2318
 - chart, 2321
 - power, 2617
 - power and expected sample size
 - conditioned on promising zone, 1045
 - power and sample size for the exact fixed sample test, 2605
 - power and sample size for the exact group sequential test, 2607
 - power boundaries, 1469–1470, 2348
 - Wang-Tsiatis, 152
 - Pampallona-Tsiatis, 1469–1470
 - Wang-Tsiatis, 400–401, 1469–1470
 - power boundary, 191
 - power chart, 154
 - Power of McNemar’s test, 2612
 - power of sequential procedure, 2318
 - power of the exact fixed sample test, 2606
 - power spending function, 2292
 - power vs sample size, 204
 - power vs sample size chart, 725, 796
 - power
 - conjunctive, 248, 251, 253–255, 259–262, 586, 590, 592, 594–595, 597, 599, 1011, 1014, 1017
 - survival, 1005, 1007, 1009
 - disjunctive, 248, 251, 253–255, 259–262, 586, 589, 591, 594–595, 597, 599, 1005, 1011, 1014, 1017
 - survival, 1007, 1009
 - global, 248, 253, 586–587, 590, 592, 1005, 1011
 - marginal, 248, 251, 586
 - binomial equivalence test, 2617, 2635
 - comparing two binomials, 2617
 - conditional, 2321
 - departure from design, 2318
 - non-inferiority, 2617
 - two binomials, 2617
 - unconditional, 2617
 - PP Normal Plot, 1867
 - pre-specified weights, 1056
 - Predictive Interval Plots, 1575
 - predictive interval plots, 1575
 - binomial data, 1592
 - continuous outcome data, 1603
 - normal data, 1603
 - survival data, 1575
 - Time to Event Data, 1575
 - predictive power, 177, 2324
 - Pregibon delta beta, 2179
 - preserve type-1 error, 2315
 - preserving alpha, 2314
 - probability of early termination, 2615
 - probability of success, 174, 981
 - Proc IM
 - normal endpoint example, 1789
 - olistat example, 1789
 - process time, 1402
 - Profile Likelihood Based Confidence Intervals, 2157
 - promising zone, 1037, 1083
 - Proportional hazard, 2219
- Q**
- QQ Normal Plot, 1868
- R**
- RALES trial, 1633, 1658, 1755, 1768
 - random number generation, 70
 - random numbers, 70
 - randomization, 395
 - randomization fraction, 395

Index

- randomization
 - balanced, 395
 - unbalanced, 143, 395
 - range of acceptable sample sizes, 2308
 - range of interim outcomes for a sample size increase, 1037
 - ratio of means
 - analysis, 1903, 1910, 1915, 1929, 1946
 - crossover design, 208, 227, 1953
 - equivalence, 220
 - example, 122, 137, 200, 208, 217, 228, 1915, 1923, 1929, 1939, 1946, 1954
 - non-inferiority, 1929
 - paired data, 110, 121, 136, 1903, 1910
 - simulation, 126, 139, 205, 230
 - single look, 111, 218, 229
 - superiority, 1915
 - t test, 204
 - test of hypotheses, 110, 121, 136, 199, 217, 228, 1903, 1910, 1915, 1929, 1954
 - ratio of proportions exact, 743
 - ratio of proportions exact tests, 736
 - RCI, 1058–1059
 - RCI method, 1272
 - re-estimating sample size for a desired power, 1364
 - recoding a categorical variable, 68
 - recoding a continuous variable, 69
 - recompute boundary, 2317
 - reconstructing a combined trial from the primary and secondary trials, 1243
 - recursive integration, 2352
 - recursive integration algorithm, 1223
 - reducing the sponsor's risk, 1029
 - refractory unstable angina, 23, 395
 - regression, 332
 - sample size, 332
 - repeated confidence interval, 1059, 2324
 - Jennison and Turnbull, 196
 - Tsiatis- Rosner and Mehta, 198
 - repeated confidence intervals
 - proof of coverage, 2325
 - repeated measure regression
 - example, 2006
 - repeated measures, 338
 - sample size, 338
 - repeated p-value, 1059
 - repeated significance testing, 2328
 - adjusted confidence interval, 2328
 - adjusted p-value, 2328
 - repeated significance tests, 2271
 - rescuing an underpowered on-going study, 1028
 - Residuals, 1847
 - Results window, 2177
 - scrolling, 2177
 - rho family, 192
 - Rho spending function, 2292
 - rho spending function, 2292
 - Right Cumulative Plot, 1872
 - risk ratio, 1416
 - Poisson, 1416
 - risk set, 2280
 - ROC Curve, 2141, 2171, 2179
 - ROC Curve vs Classification Table, 2151
 - RPV, 1059
 - rule for sample size adaptation, 1083
- S**
- safety boundaries, 1471
 - Sakoda contingency coefficient, 2598
 - sample-size computation, 2617
 - sample size, 400, 406, 409, 413, 2278
 - sample size adjustment, 1021

- sample size calculation, 2607
- sample size for survival studies, 2310
- sample size ranges, 2310
- sample size re-estimation, 1394
- sample size vs information, 2303,
- sample size, 1027
 - regression, 332
 - repeated measures, 338
 - single slope, 332
 - two slopes, 336
 - maximum usable, 1082
 - minimum usable, 1082
 - re-estimation, 1027
- sawtooth chart, 725
 - poisson, 796
- Scatter Plots, 1865
- Scharfstein Tsiatis Robins theorem, 2284, 2303
- Schurman's method for log normal data, 136, 217, 1910
- Schurman's TOST procedure, 128, 211, 222
- score statistic, 2280
- Scores test, 2139
- Searching for Nuisance Parameters, 2582, 2587
 - Restricted Range, 2589
- second interim analysis, 1432
- second interim look, 423
- SELECTIF() function, 67
- selecting the criteria for an adaptive
 - sample size increase, 1036
- semiparametric information, 2272
- Sensitivity, 2141
- sequential design, 145
- Seropositivity example, 2172
- shape parameter, 2289
- show table, 134
- Sidak procedure, 252, 590, 2031, 2181
 - survival, 2243
- significance level, 2315
- Simon's design, 774
- Simple Bar Plot, 1855
- simulating preservation of type-1 error, 1168
- simulation, 163, 489, 504, 518, 2345
- simulation tool, 489, 504, 518
- simulation
 - crossover design, 225, 230
 - equivalence, 134, 139, 216, 220, 225, 230
 - non-inferiority, 119, 126, 193, 205
 - one sample, 98
 - one sided, 98
 - paired data, 105, 119, 126, 134, 139
 - ratio of means, 126, 139, 205, 220, 230
 - superiority, 98, 105
 - enhanced, 2345
 - Muller and Schafer method, 1245
- single-look, 92, 186
- single binomial proportion, 2605
 - exact design, 2605
- single look design, 23, 395, 476, 1436
- single look
 - crossover design, 223, 229
 - equivalence, 213, 218, 223, 229
 - non-inferiority, 114
 - paired sample, 102, 114
 - ratio of means, 111, 218, 229
 - superiority, 102, 111
- single mean, 91
- single slope
 - sample size, 332
- SMALLT Data Set, 2111
- Spearman's Product-Moment Correlation, 1844
- Spearman's Rho, 1843
- special functions, 70
- special protocol assessment (SPA), 1053

Index

- Specificity, 2141
- spending function, 97, 2291, 2348
- spending function boundaries, 157, 1471, 2293, 2295
- spending function
 - Lan-DeMets, 97, 190, 103
 - rho family, 192
 - alpha, 1471
 - beta, 1471, 1484
 - gamma, 2291
 - Gm, 2291
 - Lan-DeMets, 400–401, 1418, 1471, 2291
 - LD(OF), 2291
 - LD(PK), 2291
 - power, 2292
 - recompute boundary, 2317
 - rho, 2292
- spending functions
 - interpolated, 156
- SQEND, 70
- SQNO, 70
- Stacked Bar Plot, 1856
- stage-wise ordering, 2329
- standardized difference, 1458
- statistical method:normal and binomial, 1056
- Stem and Leaf Plots, 1875
- step-down Dunnett procedure, 242, 250, 2028
- step-up Dunnett's procedure, 2029
- step-up Dunnett procedure, 250
- Step Function Plots, 1877
- stop early to reject, 409, 424
- stopping boundaries, 96, 147, 189, 1471
 - flexible, 1460
 - for early rejection, 409
 - inner wedge, 2289
 - meet at last look, 2288, 2294
 - one sided H0 or H1, 2288
 - one sided Pampallona-Tsiatis, 2288
 - preserve alpha, 2289, 2294
 - preserve beta, 2289, 2294
 - two sided H0 or H1, 2289
 - two sided Pampallona-Tsiatis, 2289
 - upper and lower, 2288
- stopping boundary at last look, 2318
- stopping for futility
 - conditional power, 2321
- stopping probabilities, 95, 104, 1426–1427, 2304
- Stratified Simulation, 900
- stroke prevention study, 1424
- stroke study
 - information based, 1418
- study design, 333, 336
- study duration versus accrual chart, 2308
- Subject Profile Plot, 1882
- subsetting a dataset, 1943
- Summary Measures, 1827–1828
 - central tendency, 1827
 - coefficient of variation, 1827
 - count, 1827
 - dispersion, 1827
 - geometric mean, 1827
 - harmonic mean, 1827
 - kurtosis, 1827
 - maximum, 1827
 - mean, 1827
 - median, 1827
 - minimum, 1827
 - mode, 1827
 - skewness, 1827
 - standard deviation, 1827
 - standard error of mean, 1827
 - sum, 1827
 - variance, 1827
- summary of extended CDL method, 1197
- Superiority
 - Design, 865, 966
- superiority, 91
- Superiority, 865, 966

- superiority, 1890, 1913, 2060
 - superiority exact tests, 736
 - superiority trial, 23, 395
 - superiority
 - analysis, 1890, 1919, 1923
 - binomial, 2060
 - crossover, 1919, 1923
 - Superiority
 - Drop-Outs, 971
 - superiority
 - example, 91, 102, 1890, 1894, 1915, 1919, 1923
 - Superiority
 - Fixed Accrual Duration, 966
 - Fixed Study Duration, 966
 - Given Accrual Duration and Study Duration, 966
 - superiority
 - group sequential, 94, 103
 - interim monitoring, 99, 106
 - multiple look, 93–94, 103
 - normal, 91, 1890, 1913
 - one sample, 91, 1890, 1913, 2060
 - paired data, 101, 110
 - ratio of means, 110, 1915, 1929
 - simulation, 98, 105
 - single look, 92, 102, 111
 - t test, 100, 109, 112
 - test of hypothesis, 91, 101, 1890
 - Superiority
 - Drop-Outs, 872
 - Non-Constant accrual, 874, 972
 - Piecewise Constant Hazard, 876
 - Simulation, 877, 882, 972–973, 1119, 1183
 - Simulation with fixed study duration, 882–883, 973–974
 - Variable accrual, 874, 972
 - survival, 1014–1015, 1017, 2305
 - survival and non-inferiority, 2283
 - survival endpoint, 820
 - Survival Endpoint, 826
 - Compare Multiple Designs, 844
 - Fixed Sample Design, 826
 - Group Sequential Design, 830
 - Interim Monitoring, 855
 - R Integration, 859
 - Simulations, 852
 - survival endpoint: Lung Cancer Trial, 1112
 - Survival Endpoint: Pancreatic cancer Trial, 1289
 - Survival Simulation, 877, 946, 972, 992, 1119, 1183
 - survival simulations, 821
 - survival studies, 2308
 - enrollment range, 2308
 - sample size, 2310
 - sample size range, 2308
 - survival
 - choice of variance, 2310
 - expected stopping time, 2305
 - single look, 1436
 - survival:statistical method, 1071
- T**
- t-test, 1837
 - paired samples, 1838
 - independent samples, 1837
 - t test, 198, 204
 - non-inferiority, 118, 198, 204
 - one sample, 100
 - paired data, 109, 118
 - ratio of means, 204
 - superiority, 100, 109, 112
 - target conditional power, 1082
 - ten-look boundaries, 2292
 - inverted, 2292
 - test for mean
 - one sample, 91, 1890
 - test of hypotheses

Index

- ratio of means, [110](#), [199](#), [217](#), [228](#),
[1915](#), [1929](#), [1954](#)
- test of hypothesis
 - non-inferiority, [113](#), [1901](#)
 - paired data, [101](#), [113](#), [1901](#)
 - superiority, [101](#)
- test statistic, [400](#), [411](#), [413](#), [422](#), [424](#)
- test statistic calculator, [2332](#)
 - by-passing, [2332](#)
- the problem of overruns, [1034](#)
- theory, [2274](#), [2614](#)
- theory of interim monitoring, [2313](#)
- theory
 - McNemar's conditional exact test,
[2611](#)
 - binomial response designs, [2276](#)
 - exact test, [2611](#)
 - normal response design, [2274](#)
 - paired binomial, [2611](#)
 - Simon's design, [2614](#)
 - Simon's minimax design, [2614](#)
 - Simon's test, [2614](#)
 - Simon's two-stage optimal design,
[2614](#)
 - time-to-event response designs,
[2280](#)
 - two binomials, [2617](#)
- third interim look, [425](#)
- time-to-event trials, [820](#)
- time, [2280](#)
- time points, [155](#)
- time to event outcomes, [2219](#), [2228](#)
- time
 - calendar, [2280](#)
 - failure, [2280](#)
 - patient follow-up, [2280](#)
- TOST procedure, [128](#), [211](#), [222](#)
- transform variable, [63](#)
- Treatment by Period Plot, [1884](#)
- trial design, [23](#), [395](#), [1416](#)
- trial simulation, [489](#), [504](#), [518](#)
- triangular continuation region, [2288](#),
[2294](#)
- Tschuprov contingency coefficient, [2598](#)
- Tsiatis and Mehta, [1220](#)
- Tutorial
 - HIV data, [2132](#)
- two-sided H0-H1 boundaries, [2296](#)
- two-sided test, [1445](#)
- two-sided tests
 - asymmetric, [1481](#), [1484](#)
- Two-stage multi-arm design
 - Continuous Endpoint, [309](#)
 - Discrete Endpoint, [621](#)
- Two-stage treatment selection design
 - Continuous Endpoint, [309](#)
 - Discrete Endpoint, [621](#)
- two binomials, [1394](#)
- Two binomials
 - equivalence test, [2574](#)
 - exact one-sided p-value, [2578](#),
[2587](#)
 - non-inferiority test, [2571](#), [2583](#)
 - unconditional exact test, [2570](#)
- two binomials
 - equivalence testing, [2617](#), [2635](#)
- Two binomials
 - Fisher's exact test, [2597](#)
- two binomials
 - Fisher's exact test, [2617](#)
 - power, [2617](#)
 - unknown baseline, [1394](#)
- two independent binomials, [2617](#)
 - equivalence testing, [2617](#)
 - Fisher's exact test, [2617](#)
 - power, [2617](#)
- two one sided tests, [211](#), [222](#)
- Two Ordered Multinomials - Wilcoxon-Mann-Whitney,
[2117](#)
- two sample exact test, [751](#)
- two sample

non-inferiority, 1926
 superiority, 1913
 two sided Pampallona-Tsiatis boundaries,
 2289
 two slopes
 sample size, 336
 two way ANOVA, 237, 1985
 type-1 error preserved, 2315

U

unbalanced data, 2029
 unbalanced randomization, 143, 395
 unconditional type-1 error, 1261
 underlying theory for extension of CDL
 method, 1192
 underpowered studies, 1027
 unequally spaced analysis, 155
 unfavorable zone, 1037
 Uniform random numbers, 70
 unknown binomial rate, 1401
 unknown variance, 1409
 unplanned analyses, 2314
 unpooled estimate, 394, 2298
 unpooled variance, 395, 427
 binomial design, 395, 427
 unpooled vs pooled, 395, 427
 unpooled
 binomial, 394, 2298
 unweighted Wald statistic, 1058
 upper and lower stopping boundaries,
 422, 424
 upper limit to the sample size increase,
 1038
 upper stopping boundary, 2288, 2294

V

VACCINE data set, 2097
 VARI data set, 1870
 variable, 57
 variable follow-up designs, 2307–2308
 variable transform, 63

variable types, 57

variable

binary, 57
 categorical, 57
 integer, 57
 numeric, 57
 string, 57

variance, 2310

variance in survival studies, 2310

variance

null or alternative, 2310
 pooled, 578
 unpooled, 578

W

Wald statistic, 1058

 weighted, 1058

Wald test, 1847, 2139

Wang-Tsiatis boundaries, 1469–1470,
 2287

Wang-Tsiatis power boundaries, 152,
 400–401

Wang-Tsiatis power boundary, 191

Weighted Bonferroni, 2188

weighted Bonferroni procedure, 252, 588,
 2031, 2181

 survival, 2243

Weighted Bonferroni

 Analysis, 2188

weighted Wald statistic, 1058

weights: pre-specified or actual, 1160

Wilcoxon-Mann-Whitney test, 179, 1956
 example, 180

Wilcoxon-Mann-Whitney

 Two Ordered Multinomials, 2117

Wilcoxon scores, 2599

Wilcoxon Signed Rank Test
 analysis, 1898

Index

example, [1898](#)

Z

zoom, [37](#)