Trial Design and Statistical Considerations in Rare Disease Clinical Trials

Yan Wang, Ph.D.
Statistical Team Leader, Division of Biometrics IV
Office of Biostatistics
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

Mitochondrial Symposium
September 6, 2019

Disclosure Statement

- No conflicts of interest
- Nothing to disclose
- This talk reflects the views of the author and should not be construed to represent FDA's views or policies
- In this talk "drug" refers to both drugs and biologics

Outline

- Design
- Endpoint
- Analysis
- Quality data

Randomized, double-blinded, and placebo controlled trial design is most commonly used

Most reliable design to determine effectiveness of a drug

- **Randomization**: unbiased assignment of patients to trial arms
- **Double-blinded:** assigned treatments are blinded to patients and investigators
- Minimize/eliminate potential biases caused by
  - Differences in prognostic patient characteristics (known/unknown)
  - Placebo effect, observer effect, and differences in standard of care
- Placebo control does not imply that the control group is untreated → all patients receive standard of care → limit ethical concern

Primary Endpoints

- Provide primary assessment of treatment effect
- Consist of multiple components in many rare disease trials
- **Composite endpoint**: components correspond to distinct events
  → *e.g. cardiac events, renal events, or death for Fabry disease*
- **Multi-component endpoint**: a within-patient combination of multiple components
  → *e.g. total Chorea score for 7 different parts of the body in patients with Huntington disease*
- **Multiple primary endpoints**: *selected in many rare disease trials due to genetic and clinical heterogeneity, and uncertainty of drug effect*

Multiple Primary Endpoints: Examples

- *Two primary endpoints are used in trials for late-onset Pompe disease, Hunter syndrome (MPS II), and MPS I*
  - *Distance walked during 6 minute walking test (6MWT)*
  - *Percent predicted forced vital capacity (FVC%)*

Hypothetical Trial

Change from baseline at 52 weeks	6MWT		FVC%	
	Placebo (N=24)	Drug (N=24)	Placebo (N=24)	Drug (N=24)
Mean (SD)	13 (60)	40 (76)	-0.1 (10)	3.5 (10)
Difference (95% CI) in Mean	**27** (-13, 67)		**3.6** (-2.3, 9.4)	
P-value				
Two sample t-test	**0.185**		**0.221**	

Challenge: *Many rare disease trials have low power to demonstrate statistically significant results due to small sample size or small treatment effect*

Analyses adjusted for prognostic variables can improve the power of significance tests and the precision of estimates of treatment effect

Hypothetical Trial

Change from baseline at 52 weeks	6MWT		FVC%	
	Placebo (N=24)	Drug (N=24)	Placebo (N=24)	Drug (N=24)
Mean (SD)	13 (60)	40 (76)	-0.1 (10)	3.5 (10)
Treatment Comparison				
Difference (95% CI*) in Mean	27 **(-11, 65)**		3.6 **(-2.1, 9.1)**	
P-value				
Two sample t-test	**0.185**		**0.221**	
ANCOVA*	**0.071**		**0.115**	

*Adjusted for baseline value.

P-values based on ANCOVA decreased by more than 40%

Global Tests for Multiple Endpoints

Hypothetical Trial

Change from baseline at 52 weeks	6MWT		FVC%	
	Placebo (N=24)	Drug (N=24)	Placebo (N=24)	Drug (N=24)
Mean (SD)	13 (60)	40 (76)	-0.1 (10)	3.5 (10)
Treatment Comparison				
Difference (95% CI*) in Mean	27 (-11, 65)		3.6 (-2.1, 9.1)	
P-value				
Two sample t-test	0.185		0.221	
ANCOVA*	0.071		0.115	
Global Test				
Rank-Sum-Test		**0.026**		
Combined-Test-Statistics		**0.010**		

*Adjusted for baseline value.

Indicating that **the drug has an effect on at least one endpoint**

Global Tests for Multiple Endpoints (2)

- **Rank-Sum-Test:** based on the sum of the ranks of data from two endpoints for each patient
  → Combines data at patient-level

- **Combined-test-statistics:** based on the two test statistics for treatment comparison for each endpoint
  → Combines test statistics at endpoint-level

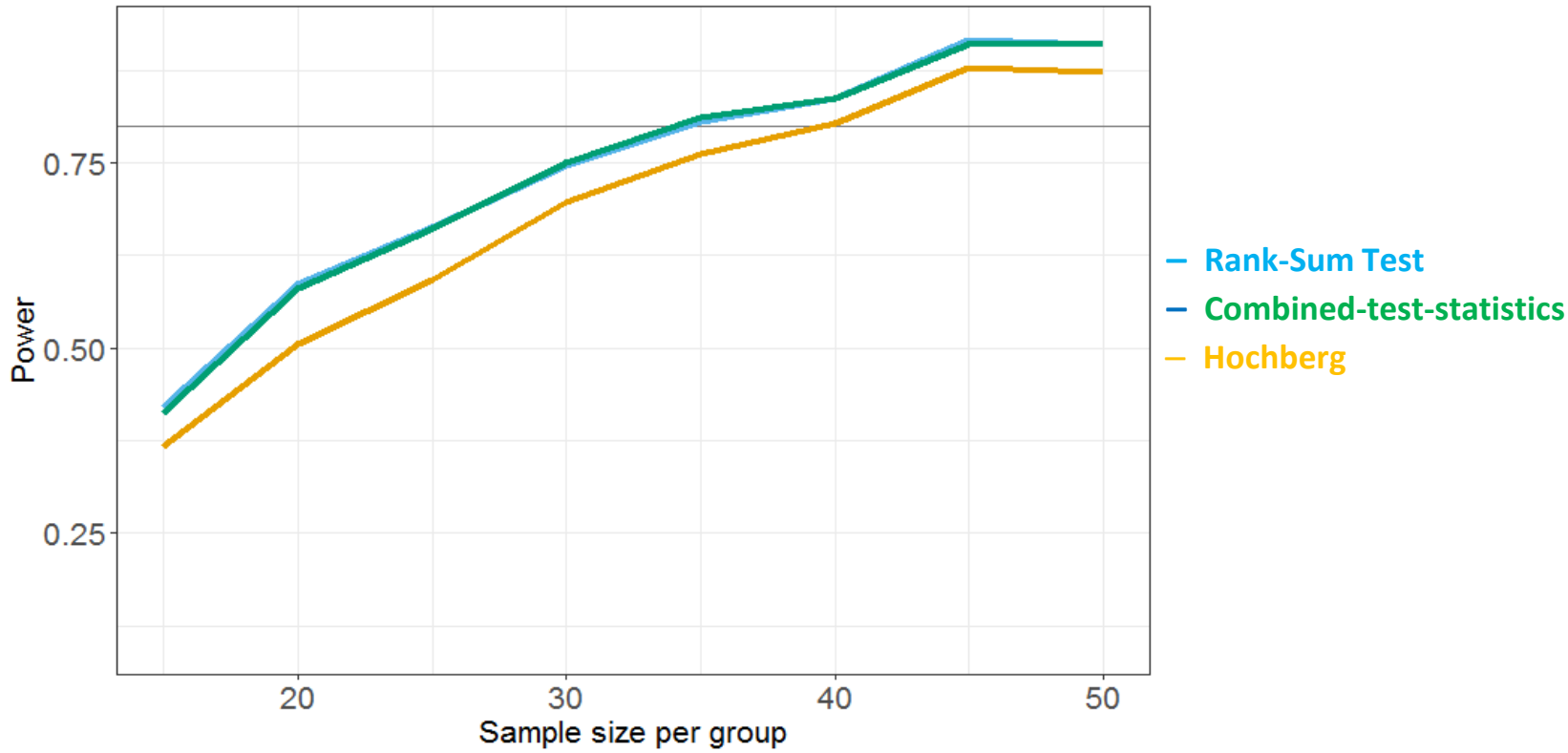
Global Tests: Interpretations

- Testing a global null hypothesis
  - **The drug has no effect on either endpoint**
- The p-value should be presented and interpreted with descriptive summary statistics for each endpoint
- When p-value < 0.05, reject the global null hypothesis and conclude that the drug has an effect on at least one endpoint
  - Justify whether the observed effect(s) are clinically meaningful
  - "p-value < 0.05" may not necessarily indicate an overall benefit if discordant effects are observed

Similar issue for composite endpoints and multi-component endpoints

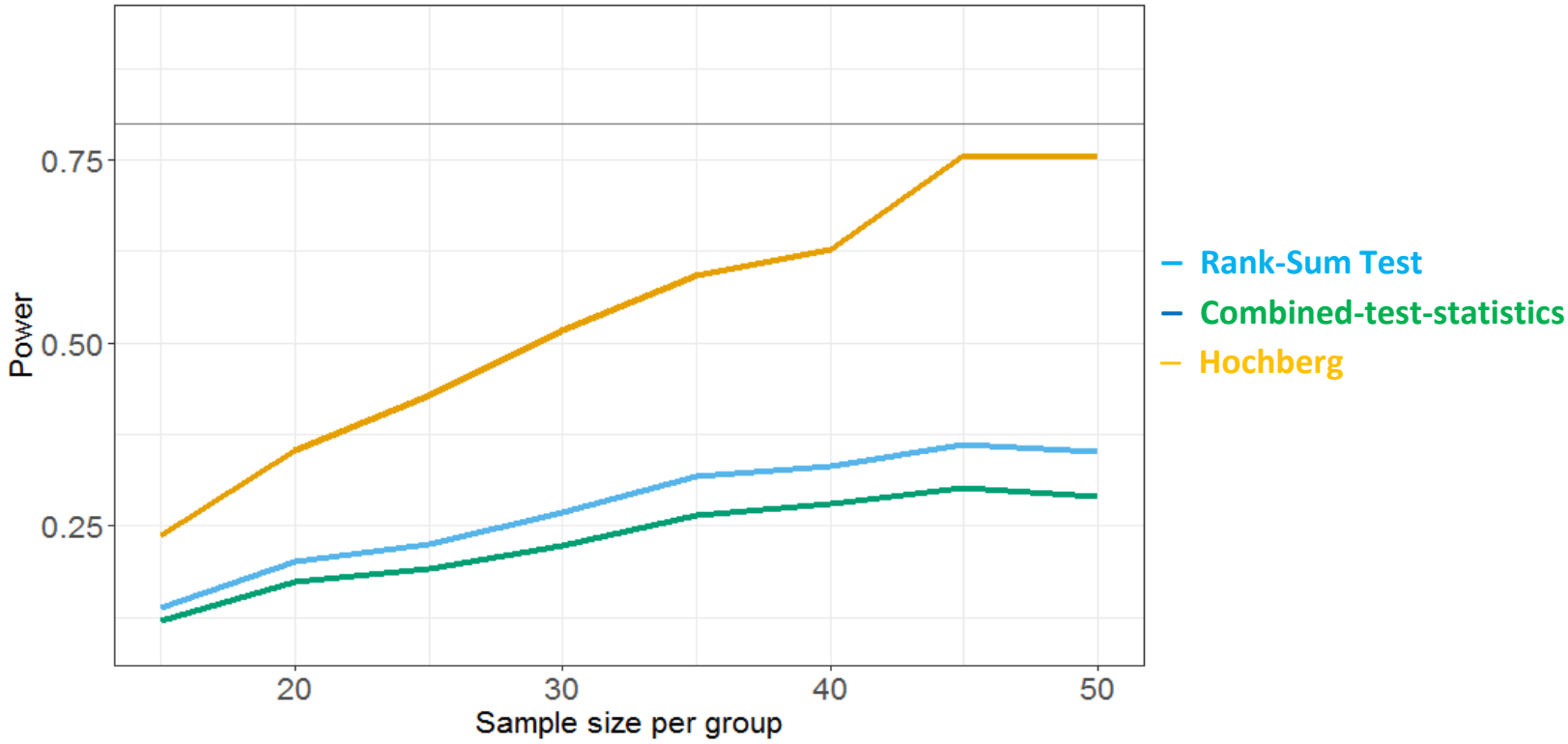
Simulation Study #1

Global Tests: can be more powerful when a drug has an effect on both endpoints



Simulation Study #2

Global Tests: are less powerful when a drug has an effect only on one endpoint



Quality Data: Essential to Success of Small Sized Trials

- Reduce noise → reduce variability of outcome measurements → increase statistical power

Example Mean difference in FVC% = 5% and N = 40 per arm (α=0.05)
10% Variability ↓ from 10 to 9, **16%** power ↑ from 60% to 70%

- Detailed plans should be developed to
  - Standardize methods and procedures for outcome assessments
  - Minimize dropouts and missing data
  - Train and remind study sites to encourage patients to complete the study even after they stop study treatment early

Conclusions

To overcome significant challenges in designing and conducting adequate and well-controlled rare disease trials, we support ***innovative trial designs and analyses*** provided they are well thought through, justified, and able to

*"**distinguish the effect of a drug from other influences**, such as spontaneous change in the course of the disease, placebo effect, or biased observation."*[1]

[1] *21 CFR 314.126*

References

1) ICH E10: Choice of Control Group and Related Issues in Clinical Trials (https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf)
2) FDA Draft Guidance for Industry Rare Diseases: Common Issues in Drug Development (https://www.fda.gov/media/119757/download)
3) FDA Draft Guidance for Industry Rare Diseases: Natural History Studies for Drug Development (https://www.fda.gov/media/122425/download)
4) FDA Draft Guidance for Industry: Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biologics with Continuous Outcomes, April 2019 (https://www.fda.gov/media/123801/download)
5) O'Brien, P. C.: "Procedures for comparing samples with multiple endpoints". *Biometrics, Vol. 40, No. 4: 1079-1087, 1984*
6) Ristl R. et al.: "Methods for the analysis of multiple endpoints in small populations: A review". *Journal of Biopharmaceutical Statistics, 29:1, 1-29, 2018*
7) Morris M., Lee W., and Wang Y.: "Evaluation of Testing Methods for Multiple Endpoint in Small Sized Trials: Application to Rare Diseases" . *CDER ORISE research project, 2019*