



Study Data Tabulation Model Implementation Guide: Human Clinical Trials

Prepared by the
CDISC Submission Data Standards Team

Notes to Readers

- This is the implementation guide for Version 1.1 of the CDISC Study Data Tabulation Model, posted for comment.
- This Implementation Guide comprises version 3.1.1 of the CDISC Submission Data Standards and Domain Models.
- See CDISC notes and assumptions regarding use of --OCCUR variable in CM, SU, AE, MH domains in Section 6.

Revision History

Date	Version	Summary of Changes
2004-07-14	3.1	Released version reflecting all changes and corrections identified during comment periods.
2005-08-26	3.1.1 Final	Released version reflecting all changes and corrections identified during comment period.

Note: Please see [Appendix 10.7](#) for Representations and Warranties; Limitations of Liability, and Disclaimers.

CONTENTS

1	INTRODUCTION.....	5
1.1	PURPOSE.....	5
1.2	ORGANIZATION OF THIS DOCUMENT.....	6
1.3	RELATIONSHIP TO PRIOR CDISC DOCUMENTS	6
1.4	HOW TO READ THIS IMPLEMENTATION GUIDE	8
1.5	SUBMITTING COMMENTS	8
2	FUNDAMENTALS OF THE SDTM.....	9
2.1	OBSERVATIONS AND VARIABLES	9
2.2	DATASETS AND DOMAINS	10
2.3	SPECIAL-PURPOSE DOMAINS	11
2.4	THE GENERAL DOMAIN CLASSES	11
2.5	THE CDISC STANDARD DOMAIN MODELS.....	12
2.6	CREATING A NEW DOMAIN	13
2.6.1	How to Include Variables in New Domains (Steps 4 through 7).....	15
3	SUBMITTING DATA IN STANDARD FORMAT.....	16
3.1	STANDARD METADATA FOR DATASET CONTENTS AND ATTRIBUTES.....	16
3.2	USING THE CDISC DOMAIN MODELS IN REGULATORY SUBMISSIONS.....	17
3.2.1	CDISC Submission Dataset Definition Metadata	17
3.2.2	CDISC Submission Value-Level Metadata	19
3.2.3	Conformance.....	20
4	ASSUMPTIONS FOR DOMAIN MODELS	21
4.1	GENERAL ASSUMPTIONS FOR ALL DOMAINS	21
4.1.1	General Dataset Assumptions	21
4.1.2	General Variable Assumptions	22
4.1.3	Coding and Controlled Terminology Assumptions	23
4.1.4	Actual and Relative Time Assumptions	25
4.1.5	Other Assumptions.....	31
5	MODELS FOR SPECIAL PURPOSE DOMAINS.....	35
5.1.1	Demographics Domain Model — DM.....	35
5.1.1.1	Assumptions for Demographics (DM) domain model	37
5.1.2	Comments Domain Model — CO.....	38
5.1.2.1	Assumptions for Comments (CO) domain model.....	39
6	DOMAIN MODELS BASED ON THE GENERAL CLASSES.....	40
6.1	INTERVENTIONS.....	40
6.1.1	Concomitant Medications — CM.....	40
6.1.1.1	Assumptions for Concomitant Medications (CM) domain model	43
6.1.2	Exposure — EX	45
6.1.2.1	Assumptions for Exposure (EX) domain model	47
6.1.3	Substance Use - SU.....	49
6.1.3.1	Assumptions for Substance Use (SU) domain model	52
6.2	EVENTS	53
6.2.1	Adverse Events — AE	53
6.2.1.1	Assumptions for Adverse Events (AE) domain model.....	56
6.2.2	Disposition — DS	59
6.2.2.1	Assumptions for Disposition (DS) domain model	60
6.2.3	Medical History — MH	62
6.2.3.1	Assumptions for Medical History (MH) domain model	64

6.3	FINDINGS.....	66
6.3.1	ECG Test Results — EG.....	66
6.3.1.1	Assumptions for ECG (EG) domain model.....	69
6.3.2	Inclusion/Exclusion Exceptions — IE.....	70
6.3.2.1	Assumptions for Inclusion/Exclusion Exceptions (IE) domain model.....	71
6.3.3	Laboratory Test Results — LB.....	72
6.3.3.1	Assumptions for Laboratory Test Results (LB) domain model.....	75
6.3.4	Physical Examinations — PE.....	76
6.3.4.1	Assumptions for Physical Examinations (PE) domain model.....	78
6.3.5	Questionnaires — QS.....	79
6.3.5.1	Assumptions for Questionnaire (QS) domain model.....	81
6.3.6	Subject Characteristics — SC.....	82
6.3.6.1	Assumptions for Subject Characteristics (SC) domain model.....	83
6.3.7	Vital Signs — VS.....	84
7	TRIAL DESIGN DATASETS.....	87
7.1	INTRODUCTION.....	87
7.2	PLANNED ELEMENTS, ARMS, AND VISITS.....	88
7.2.1	Trial Elements.....	88
7.2.2	Trial Arms.....	89
7.2.3	Trial Visits.....	90
7.3	SUBJECT ELEMENTS AND VISITS.....	91
7.3.1	Subject Elements.....	91
7.3.2	Subject Visits.....	92
7.4	TRIAL ARMS.....	93
7.4.1	Identifying Trial Arms.....	93
7.4.2	Developing the Trial Arms Table.....	94
7.4.3	Distinguishing between Branches and Transitions.....	101
7.4.4	Trial Epoch Concept.....	101
7.4.5	Rules concept.....	106
7.4.6	Recap of Trial Arms Variables.....	106
7.4.7	Truncated Arms.....	107
7.5	TRIAL ELEMENTS.....	107
7.5.1	Identifying Trial Elements.....	109
7.5.2	Developing the Trial Elements Table.....	110
7.5.3	Recap of Trial Elements Variables.....	112
7.5.4	Distinguishing Elements from Epochs.....	112
7.6	TRIAL VISITS.....	113
7.6.1	Identifying Trial Visits.....	113
7.6.2	Developing the Trial Visits Table.....	113
7.6.3	Recap of Trial Visits Variables.....	114
7.7	SUBJECT ELEMENTS.....	115
7.7.1	Identifying Subject Elements.....	115
7.7.2	Unplanned Elements.....	115
7.7.3	Deriving SE End Date/Times.....	115
7.7.4	Recap of Subject Elements Variables.....	116
7.7.5	Using Subject Elements Data to Place Subject Data within an Element or Epoch.....	116
7.8	SUBJECT VISITS.....	116
7.8.1	Identifying Subject Visits.....	116
7.8.2	Recap of Subject Visits Variables.....	117
7.9	TRIAL INCLUSION/EXCLUSION CRITERIA.....	117
7.10	TRIAL SUMMARY INFORMATION.....	118
7.11	HOW TO MODEL THE DESIGN OF A CLINICAL TRIAL.....	119
8	REPRESENTING DATA RELATIONSHIPS.....	120
8.1	RELATING GROUPS OF RECORDS WITHIN A DOMAIN.....	121
8.1.1	--GRPID Example.....	121

8.2	RELATING PEER RECORDS IN SEPARATE DATASETS	122
8.2.1	RELREC Dataset	123
8.2.2	RELREC Dataset Examples.....	123
8.3	RELATING DATASETS	124
8.3.1	RELREC Dataset Relationship Example	124
8.4	RELATING NON-STANDARD VARIABLE VALUES TO A PARENT DOMAIN	125
8.4.1	SUPPQUAL Dataset	126
8.4.2	Submitting Supplemental Qualifiers in Separate Datasets.....	127
8.4.3	SUPPQUAL Examples	127
8.5	RELATING COMMENTS TO A PARENT DOMAIN	128
8.5.1	COMMENTS Example.....	128
8.6	HOW TO DETERMINE WHERE DATA BELONG IN THE SDTM.....	129
9	IMPLEMENTATION EXAMPLES	131
9.1	DEMOGRAPHICS EXAMPLE	131
9.2	INTERVENTIONS EXAMPLES.....	131
9.2.1	CM Example: Intermittent Use of Concomitant Medications.....	131
9.2.2	EX Examples:	133
9.2.3	SU Example:	135
9.3	EVENTS EXAMPLES	139
9.3.1	AE Example	139
9.3.2	DS Examples.....	140
9.3.3	MH Example:	143
9.4	FINDINGS EXAMPLES.....	144
9.4.1	EG Examples	144
9.4.2	IE Example.....	147
9.4.3	LB Examples.....	148
9.4.4	PE Example.....	151
9.4.5	QS Examples.....	153
9.4.6	SC Example	155
9.4.7	VS Example	156
9.5	TRIAL DESIGN EXAMPLES.....	158
9.5.1	Defining Epochs - a How-to Example Using the Example Crossover Study.....	158
9.5.2	Trial Summary Example	162
10	APPENDICES.....	163
10.1	CDISC SDS TEAM	163
10.2	GLOSSARY OF TERMS	164
10.3	STANDARDIZED AND RESERVED CODES	165
10.3.1	Reserved Domain Codes.....	165
10.3.2	Electrocardiogram Test Codes (for measured or calculated parameters)	167
10.3.3	Vital Signs Test Codes	167
10.3.4	Supplemental Qualifiers Name Codes	168
10.3.5	Trial Summary Codes	169
10.4	CDISC VARIABLE-NAMING FRAGMENTS	171
10.5	LESSONS LEARNED FROM THE PILOT.....	173
10.6	REVISION HISTORY	175
10.7	REPRESENTATIONS AND WARRANTIES; LIMITATIONS OF LIABILITY, AND DISCLAIMERS	183

1 Introduction

1.1 PURPOSE

This document describes the CDISC Version 3.1.1 (V3.1.1) Submission Data Standards (SDS), which have been prepared by the Submissions Data Standards team of the Clinical Data Interchange Standards Consortium (CDISC). Like its predecessors, V3.1.1 is intended to guide the organization, structure, and format of standard clinical trial tabulation datasets submitted to a regulatory authority such as the US Food and Drug Administration (FDA). V3.1.1 supersedes all prior versions of the CDISC Submission Data Standards.

This document should be used in close concert with the CDISC Study Data Tabulation Model (SDTM) available at <http://www.cdisc.org/models/sds/v3.1/index.html> and describes how to implement the SDTM for use with data tabulations submitted for human clinical trials. Version 1.1 of the SDTM, which should be read before this Implementation Guide (SDTMIG), describes a general conceptual model for representing clinical study data that is submitted to regulatory authorities. SDTMIG V3.1.1 provides specific domain models, assumptions, business rules, and examples for preparing standard datasets that are based on the SDTM.

Tabulation datasets, which are electronic listings of individual observations for a subject that comprise the essential data collected in a clinical trial, are one of four types of data currently submitted to the FDA along with patient profiles, listings, and analysis files. By submitting tabulations that conform to the standard structure, sponsors may benefit by no longer having to submit separate patient profiles or listings with a product marketing application. V3.1.1 is not currently intended to fully meet the needs supported by analysis datasets, which will continue to be submitted separately in addition to the tabulations. Since July 2004, the FDA has referenced use of the SDTM in the Study Data Specifications for the Electronic Common Technical Document, available at <http://www.fda.gov/cder/regulatory/ersr/Studydata-v1.1.pdf>.

The availability of standard submission data will provide many benefits to regulatory reviewers. Reviewers can be trained in the principles of standardized datasets and the use of standard software tools, and thus be able to work with the data more effectively with less preparation time. Another benefit of the standardized datasets is to support the FDA's efforts to develop a repository for all submitted trial data and a suite of standard review tools to access, manipulate, and view the tabulations. Use of these data standards is also expected to benefit industry by streamlining the flow of data from collection through submissions, and facilitating data interchange between partners and providers. Note that the SDTM represents an interchange standard, rather than a presentation format -- it is assumed that tabulation data will be transformed by software tools to better support viewing and analysis.

This document is intended for companies and individuals involved in the collection, preparation, and analysis of clinical data that will be submitted to regulatory authorities. Audiences are also advised to read the [CDISC Submission Metadata Model](http://www.cdisc.org/pdf/SubmissionMetadataModelV2.pdf) available at <http://www.cdisc.org/pdf/SubmissionMetadataModelV2.pdf> for additional historical background on how to provide metadata descriptions for submission data. The primary goal of the Metadata Model is to provide regulatory reviewers with a clear understanding of the datasets provided in a submission by communicating clear descriptions of the structure, purpose, attributes, and contents of each dataset and dataset variable. Guidance, specifications, and regulations for the application of this model will be provided separately by regulatory authorities. Audiences are advised to refer to these guidance documents for the most current recommendations for the submission of clinical data.

1.2 ORGANIZATION OF THIS DOCUMENT

This document is organized into the following sections:

- [Section 1](#), INTRODUCTION, provides an overall introduction to the V3.1.1 models and describes changes from prior versions.
- [Section 2](#), FUNDAMENTALS OF THE SDTM, recaps the basic concepts of the SDTM, and describes how this implementation guide should be used in concert with the SDTM.
- [Section 3](#), SUBMITTING DATA IN STANDARD FORMAT, explains how to describe metadata for regulatory submissions, and how to assess conformance with the standards.
- [Section 4](#), ASSUMPTIONS FOR DOMAIN MODELS, describes basic concepts, business rules, and assumptions that should be taken into consideration before applying the domain models.
- [Section 5](#), MODELS FOR SPECIAL PURPOSE DOMAINS, describes the Demographics and Comments special-purpose domains.
- [Section 6](#), DOMAIN MODELS BASED ON THE GENERAL CLASSES, provides specific annotated metadata models based on the three general observation classes. These include the revised V3.1 models for the ten common domains previously modeled by CDISC in Version 2.0; the domains for Inclusion/Exclusion Exceptions, Subject Characteristics, and Substance Use introduced in Version 3.0; and the Questionnaires domain introduced with V 3.1.
- [Section 7](#), TRIAL DESIGN DATASETS, describes implementation issues related to the use of the Trial Design Model described in the SDTM.
- [Section 8](#), REPRESENTING DATA RELATIONSHIPS, describes how to represent relationships between separate domains, datasets, and/or records.
- [Section 9](#), IMPLEMENTATION EXAMPLES, provides specific examples based on actual clinical-trial data. Several of these examples were prepared as a result of the V3 Pilot Project conducted in the Summer of 2003.
- [Section 10](#), APPENDICES, provides additional background material and describes other supplemental material relevant to implementation.

1.3 RELATIONSHIP TO PRIOR CDISC DOCUMENTS

As stated above, this document, Version 3.1.1, together with the SDTM, represents the most recent version of the CDISC Submission Data Domain Models, previously known as V3.1. Since all updates to Version 3.1 are intended to be backward compatible with version 3.1, the term “V3.x” is used to refer to Version 3.1 and all subsequent versions.

The most significant changes since V3.1 in this version include:

- The implementation guide has been expanded with new variables for consistency with Version 1.1 of the SDTM
- Sponsors may now include any valid variable from the same SDTM general class in a domain model. Previously, domain models were restricted to only those qualifier variables specified in the Implementation Guide
- A new Trial Summary (TS) dataset has been added to the Trial Design Model to describe summary characteristics of the study

- A preferred alternative means of submitting supplemental qualifiers in separate datasets per domain has been defined in [Section 8](#).
- Substantial clarifications in [Section 8](#)
- A new section on warranties and disclaimers has been added as [Appendix 10.7](#)
- Numerous corrections have been applied to the text, assumptions, domain models and examples throughout.

A detailed list of changes between versions is provided in [Appendix 10.6](#).

Note that V3.1.1 continues to represent most data in a vertical structure. Standard horizontal listings of data, such as those described in the V2 horizontal representations of ECGs and Vitals by visit, will be produced by FDA standard review tools. A complete list of changes from V3.1 is included in the V3.1.1 Change Log (see [Section 10.6.3](#)).

V3.1 was the first fully implementation-ready version of the CDISC Submission Data Standards that was directly referenced by the FDA for use on clinical studies involving human drug products. However, future improvements and enhancements such as V3.1.1 will continue to be made as sponsors develop more experience submitting data in this format. Therefore, CDISC will be preparing regular updates to the implementation guide to provide corrections, clarifications, additional domain models, examples, business rules, and conventions for using the standard domain models. CDISC will produce further documentation on controlled terminology as separate publications beginning in 2005, so sponsors are encouraged to check the CDISC website (www.cdisc.org/standards/) frequently for additional information.

1.4 HOW TO READ THIS IMPLEMENTATION GUIDE

The SDS V3.1.1 Implementation Guide (SDTMIG) is best read online, so the reader can benefit from the many hyperlinks included to both internal and external references. The following guidelines may be helpful in reading this document:

1. First read the SDTM to gain a general understanding of the conceptual model for the Submission Data Standards.
2. Next, read Sections 1-3 of this document to review the key concepts for preparing domains and submitting data to regulatory authorities. Refer to the [Glossary in Section 10.2](#) as necessary.
3. Read the General Assumptions for all Domains [in Section 4](#).
4. Review [Sections 5 and 6](#), referring back to Assumptions as directed (hyperlinks are provided).
5. Review the Implementation examples for each domain [in Section 9](#) to gain an understanding of how to apply the domain models for specific types of data.
6. [Read Section 7](#) to understand the fundamentals of the Trial Design Model and consider how to apply the concepts for typical protocols. Since the Trial Design model includes many new concepts, readers may choose to defer this step until after they have mastered the essentials of creating standardized domains. New extensions to the trial design model will be published separately on the CDISC web site.
7. [Review Section 8](#) to learn the advanced concepts of how to express relationships between datasets, records and additional variables not specifically defined in the models.
8. Finally, review the additional appendices as appropriate.

1.5 SUBMITTING COMMENTS

Comments on this document can be submitted through the [CDISC Discussion Board](#).

2 Fundamentals of the SDTM

2.1 OBSERVATIONS AND VARIABLES

The V3.x submission data standards are based on the SDTM Version 1.1 document, which provides a general framework for organizing information collected during clinical trials that is to be submitted to the FDA. The SDTM is built around the concept of observations collected about subjects who participated in a clinical study. Each observation can be described by a series of variables, corresponding to a row in a dataset or table. Each variable can be classified according to its *Role*. A Role determines the type of information conveyed by the variable about each distinct observation and how it can be used. Variables can be classified into four major roles:

- *Identifier* variables, which identify the study, subject of the observation, the domain, and the sequence number of the record
- *Topic* variables, which specify the focus of the observation (such as the name of a lab test)
- *Timing* variables, which describe the timing of the observation (such as start date and end date)
- *Qualifier* variables, which include additional illustrative text, or numeric values that describe the results or additional traits of the observation (such as units or descriptive adjectives).

A fifth type of variable role, *Rule*, was introduced with V3.1 to express an algorithm or executable method to define start, end, or looping conditions in the Trial Design model.

The set of Qualifier variables can be further categorized into five sub-classes:

- *Grouping Qualifiers* are used to group together a collection of observations within the same domain. Examples include --CAT and --SCAT.
- *Result Qualifiers* describe the specific results associated with the topic variable for a finding. It is the answer to the question raised by the topic variable. Examples include --ORRES, --STRESC, and --STRESN. Many of the values in the DM domain are also classified as Result Qualifiers.
- *Synonym Qualifiers* specify an alternative name for a particular variable in an observation. Examples include --MODIFY and --DECOD, which are equivalent terms for a --TRT or --TERM topic variable, --TEST and --LOINC which are equivalent terms for a --TESTCD.
- *Record Qualifiers* define additional attributes of the observation record as a whole (rather than describing a particular variable within a record). Examples include --REASND, AESLIFE, and all other SAE flag variables in the AE domain; and --BLFL, --POS and --LOC, --SPEC, --LOT, --NAM.
- *Variable Qualifiers* are used to further modify or describe a specific variable within an observation and is only meaningful in the context of the variable they qualify. Examples include --ORRESU, --ORNRHI, and --ORNRLO, all of which are variable qualifiers of --ORRES, and --DOSU and --DOSFRM, all of which are variable qualifiers of --DOSE.

For example, in the observation, 'Subject 101 had mild nausea starting on Study Day 6,' the Topic variable value is the term for the adverse event, 'NAUSEA'. The Identifier variable is the subject identifier, '101'. The Timing variable is the study day of the start of the event, which captures the information, 'starting on Study Day 6', while an example of a Record Qualifier is the severity, the value for which is 'MILD'. Additional Timing and Qualifier variables could be included to provide the necessary detail to adequately describe an observation.

2.2 DATASETS AND DOMAINS

Observations are normally collected for all subjects in a series of domains. A domain is defined as a collection of logically-related observations with a topic-specific commonality about the subjects in the trial. The logic of the relationship may relate to the scientific subject matter of the data, or to its role in the trial. Typically, each domain is represented by a dataset, but it is possible to have information relevant to the same topicality spread among multiple datasets (see [Section 8.6](#)). Each dataset is distinguished by a unique, two-character DOMAIN code that should be used consistently throughout the submission. This DOMAIN code is used in the dataset name, the value of the DOMAIN variable within that dataset, and as a prefix for most variable names in the dataset.

The dataset structure for observations is a flat file representing a table with one or more rows and columns. Normally, one dataset is submitted for each domain. Each row of the dataset represents a single observation and each column represents one of the variables. Each dataset or table is accompanied by metadata definitions that provide information about the variables used in the dataset. The metadata are described in a data definition document named 'Define' that is submitted along with the data to regulatory authorities. (See the Case Report Tabulation Data Definition Specification (define.xml), available on the CDISC website, for information about an xml representation of the data definition document.) The [CDISC Submission Metadata Model](#) uses seven distinct metadata attributes to be defined for each dataset variable in the metadata definition document:

- The *Variable Name* (limited to 8-characters for compatibility with the SAS Transport format)
- A descriptive *Variable Label*, using up to 40 characters, which should be unique for each variable in the dataset
- The data *Type* (e.g., whether the variable value is a character or numeric)
- The set of controlled terminology for the value or the presentation format of the variable (*Controlled Terms* or *Format*)
- The *Origin* or source of each variable
- The *Role* of the variable, which determines how the variable is used in the dataset. For the V3.x domain models, roles are used to represent the categories of variables as Identifier, Topic, Timing, or the five types of Qualifiers. Since these roles are predefined for all domains that follow the general classes, they do not need to be specified by sponsors in their Define data definition document. Actual submission metadata may use additional role designations, and more than one role may be assigned per variable to meet different needs.
- *Comments* or other relevant information about the variable or its data.

Data stored in dataset variables include both raw (as originally collected) and derived values (e.g., converted into standard units, or computed on the basis of multiple values, such as an average). In the SDTM only the name, label, and type are listed with a set of CDISC guidelines that provide a general description for each variable used by a general observation class. The Domain models included in [Sections 5 and 6](#) of this document provide additional information about Controlled Terms or Format and Origin, as well as notes on proper usage. Comments are included as necessary according to the needs of individual studies.

The presence of an asterisk (*) in the 'Controlled Terms or Format' column indicates that a discrete set of values (controlled terminology) is expected to be made available for this variable. This set of values may be sponsor-defined in cases where standard vocabularies have not yet been defined (represented by a single *) or from an external published source such as MedDRA (represented by **). The CDISC controlled terminology group will be publishing additional guidance on use of controlled terminology separately in the future.

2.3 SPECIAL-PURPOSE DOMAINS

The CDISC V3.x Submission Data Domain Models include special-purpose domains with a specific structure and cannot be extended with any additional qualifier or timing variables other than those specified. Two of these – Demographics and Comments -- are described in [Section 5](#) of this document. Demographics includes a set of standard variables that describe each subject in a clinical study; Comments describes a fixed structure for recording free-text comments on a subject, or comments related to records or groups of records in other domains.

Additional fixed structure, non-extensible special-purpose domains are discussed in the Trial Design model described in [Section 7](#) and in [Section 8](#), which discusses record relationships.

2.4 THE GENERAL DOMAIN CLASSES

Most observations collected during the study (other than those represented in special purpose domains) should be divided among three general observation classes: Interventions, Events, or Findings:

- The *Interventions* class captures investigational treatments, therapeutic treatments, and surgical procedures that are intentionally administered to the subject (with some actual or expected physiological effect) either as specified by the study protocol (e.g., “exposure”), coincident with the study assessment period (e.g., “concomitant medications”), or other substances self-administered by the subject (such as alcohol, tobacco, or caffeine)
- The *Events* class captures occurrences or incidents independent of planned study evaluations occurring during the trial (e.g., 'adverse events' or 'disposition') or prior to the trial (e.g., 'medical history').
- The *Findings* class captures the observations resulting from planned evaluations to address specific questions such as observations made during a physical examination, laboratory tests, ECG testing, and sets of individual questions listed on questionnaires.

In most cases, the identification of the general class appropriate to a specific collection of data by topicality is straightforward. Often the Findings general class is the best choice for general observational data collected as measurements or responses to questions. In cases when the topicality may not be as clear, the choice of class may be based more on the scientific intent of the protocol or analysis plan or the data structure. [Appendix 10.6](#) of this document proposes additional guidelines for choosing the appropriate general class and proposes a modeling approach where a findings structure should be used to represent data that may relate to interventions or events records.

All datasets based on any of the general observation classes share a set of common Identifier variables and Timing variables, which are described in the SDTM. Three general rules apply when determining which variables to include in a domain:

- The same set of Identifier variables applies to all domains based on the general observation classes. An optional identifier can be used wherever appropriate.
- Any valid Timing variable is permissible for use in any submission dataset (such as to describe studies with more precise time points such as a Pharmacokinetics trial), but it should be used consistently where applicable for all domains.
- Any additional Qualifier variables from the same general class may be added to a domain model.

Assumptions for use of the Domain are described in [Section 4](#) of this document.

2.5 THE CDISC STANDARD DOMAIN MODELS

The following standard domains with their respective DOMAIN codes have been defined or referenced by CDISC in this document (domain models marked with an asterisk (*) are not included in this document and will be posted separately for comment):

Special-Purpose Domains:

- Demographics - DM
- Comments - CO

Interventions:

- Concomitant Medications - CM
- Exposure - EX
- Substance Use - SU

Events:

- Adverse Events - AE
- Disposition - DS
- Medical History - MH
- *Protocol Deviations - DV

Findings:

- *Drug Accountability - DA
- ECG Tests - EG
- Inclusion/Exclusion Exceptions - IE
- Laboratory Tests - LB
- *Microbiology Specimens - MB
- Questionnaires - QS
- *Microbiology Susceptibility - MS
- Physical Examinations - PE
- *Pharmacokinetics Concentrations - PC
- Subject Characteristics - SC
- *Pharmacokinetics Parameters - PP
- Vital Signs - VS

Trial Design Domains:

- Trial Elements - TE
- Trial Arms - TA
- Trial Visits - TV
- Subject Elements - SE
- Subject Visits - SV
- Trial Inclusion/Exclusion Criteria - TI
- Trial Summary - TS

Special-Purpose Relationship Datasets (defined in [Section 8](#)):

- Supplemental Qualifiers - SUPQUAL
- Relate Records - RELREC

Note, a sponsor would only submit the data domains that are actually collected. Decisions on what data to collect should be based on the scientific objectives of the study. A list of standard domain codes for other commonly used domains is provided in [Appendix 10.3.1](#). In the future, additional domain models based on the three general observation classes, and additional standard domain codes will be published by CDISC, and sponsors are encouraged to check the CDISC web site periodically for updates.

When preparing submissions based on the domain models, sponsors should not add any variables other than additional relevant timing variables and qualifiers from the same general class to the V3.x models, since non-standard variables could compromise the FDA's abilities to populate the data repository and use standard tools. A sponsor is free to drop certain variables from the domain model, and the corresponding descriptions from the data definition document, but new variables (other than those that are from the same general class) must not be added, and existing variables should not be renamed, or modified for novel usage.

When evaluating how to represent data according to the V3.x model, sponsors should keep in mind that the V3/V3.x paradigm is very different from that of previous versions. Data that historically would have been modeled horizontally into one record is now often converted into multiple records in V3.x. If additional variables remain that cannot be mapped to the variables defined for a standard domain model, Sponsors should use the Supplemental Qualifiers dataset described in [Section 8.4](#) of this document. Note that any collected data in an analysis dataset must also appear in a tabulation dataset. If a sponsor wishes to propose that new variables be added to the Version 3.x models, they should provide the rationale and description of the added variable(s) to the CDISC SDS Team so the impact on V3.x models can be assessed. Specific examples are also requested to illustrate the context for the new variables. This information can be provided to the CDISC SDS Team by posting it on the [CDISC Public Discussion Forum](#).

2.6 CREATING A NEW DOMAIN

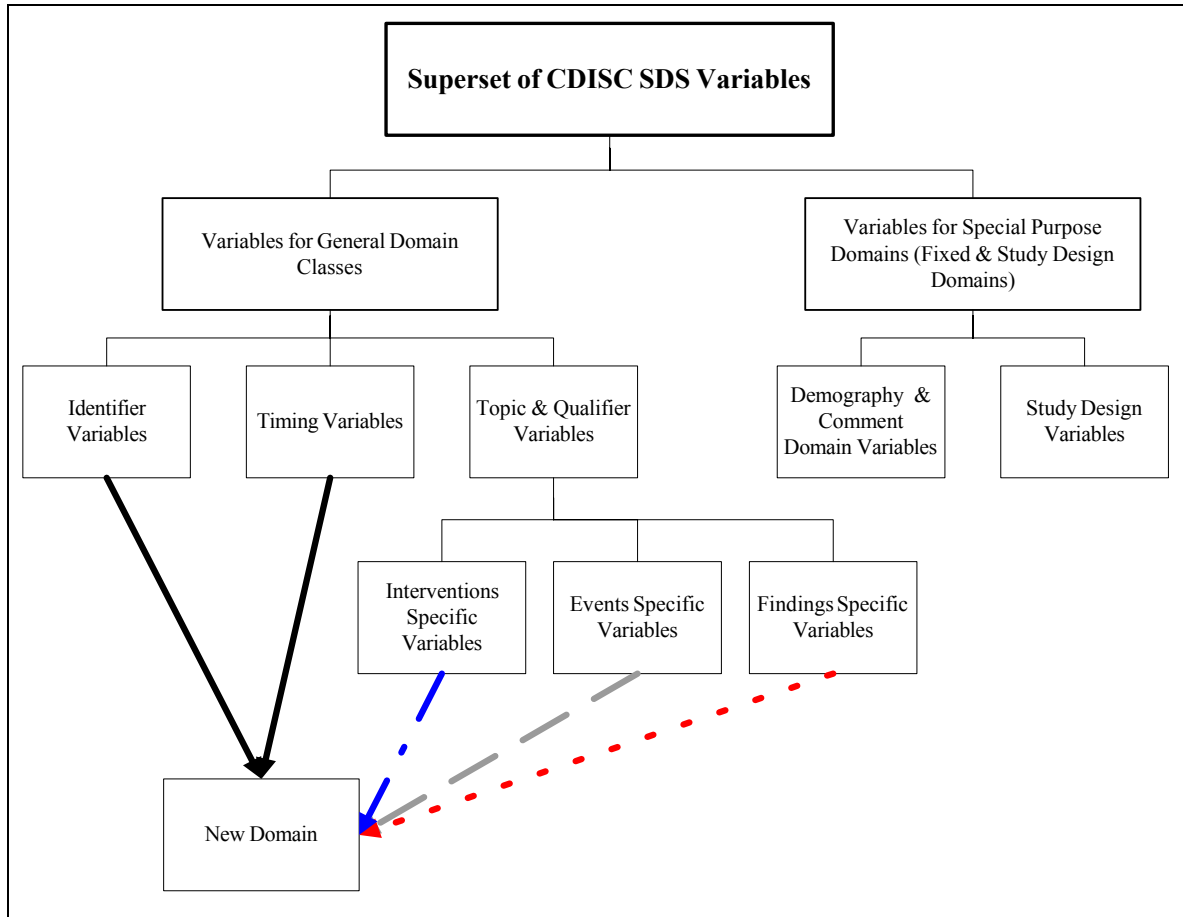
This section describes the overall process of how to create a new CDISC SDTM domain. New domains should only be created if a published standard domain does not already exist for the data being modeled. When a new domain is created, one of the most important decisions is the choice of the proper general class. Deciding on the number of separate domains to use within a general class should be made based on the specific requirements of the study. See [Section 9](#) of this document for specific examples of the application of the CDISC SDTM for Interventions, Events and Findings general domain models. When creating a new domain, sponsors should follow the predefined steps listed below:

1. Ensure that there is a definite need to create a new domain. For example, the Findings vertical structure can often accommodate many types of information that may have previously spanned separate domains in a sponsor's internal database. In V3.x, such information can often be represented as separate records in the same submissions domain. Mechanisms for incorporating new Qualifier variables not described in the general observation classes and for defining relationships between separate datasets or records are described in [Section 8 of this document](#).
2. Once the need for a new domain is determined, verify that there are no existing domain models by reviewing the current models posted on the [CDISC website](#).
3. Choose the general observation class (Interventions, Events, or Findings) that best fits the data as follows:
 - a. Identify the topic of the observation and determine which of the three general observation classes it most closely resembles. If the new domain shares both the same topicality and general observation class as an existing domain in the submission, the existing domain should be used.
 - b. Look for other existing domain models that may serve as a relevant prototype (most domains will follow the Findings model).
 - c. Determine if the chosen general observation class has most of the required and expected qualifiers for the new domain.
 - d. Select the variables that apply to the new domain once you have selected the general observation class.
4. Select and include the identifier variables (USUBJID, STUDYID, and --SEQ).
5. Include the topic variable from the identified general observation class (e.g., --TESTCD for Findings).

6. Select and include the relevant Qualifier variables from the identified general observation class (e.g., Findings) omitting any extraneous variables from the general model that do not apply.
7. Select and include the applicable timing variables from the SDTM.
8. Check with the CDISC website for a previously identified two-character domain identifier or abbreviation. If one has not been assigned by CDISC, then the sponsor may select the unique two-character domain code to be used consistently throughout the submission.
9. Apply the two-character domain code to the appropriate variables in the domain.
 - a. Replace all variable prefixes shown in the models with two hyphens '--' with the domain code.
10. Set the order of variables consistent with the order of the domain model most similar to the new domain.
11. Adjust the labels of the variables only as appropriate to properly convey the meaning in the context of the data being submitted in the newly-created domain. Use title case for all labels (title case means to capitalize the first letter of every word except for articles, prepositions, and conjunctions).
12. Compare for consistency with CDISC General Assumptions and other variables used in similar domain examples to ensure new variables are not being created when appropriate standard variables already exist.
13. Label the dataset within the metadata definition document as follows: <the domain name>, <model type: Findings, Events, Interventions>, CDISC SDTM Version <Version Number e.g., 3.1.1>, <dataset structure (e.g., 'one record per subject' or 'one record per event') >.
14. Submit a request to add specific new variables necessary to represent additional information in the general observation classes through the CDISC Public Discussion Forum.

As noted in Step 8 of the above process, each domain is distinguished by a unique, two-character DOMAIN code that should be used consistently throughout the submission. The two-character identifier should be controlled for consistency so that the FDA tools for receiving data can recognize it. The DOMAIN code should also be included in the name of the corresponding dataset files (e.g., ae.xpt) and also used as a prefix for variables to distinguish them in the event that merges are performed between datasets using SAS. CDISC will maintain a separate list of standard domain codes on the CDISC website and sponsors may post additional suggested codes to add to this list through the CDISC Public Discussion Forum.

2.6.1 How to Include Variables in New Domains (Steps 4 through 7)



As illustrated in the figure above, the CDISC SDTM has an inclusive superset or a predefined list of variables. The SDTM variables are subdivided into two general variable classes: those for the 'general observation classes' and those for the 'special domains' (those with fixed structures and the trial design domains). For more information regarding these classifications, please see [sections 2.3 and 2.4](#). From a data modeling perspective, once it has been determined that a new domain is needed (i.e., there is no published equivalent domain), the sponsor determines the domain's general observation class (either as an Intervention, Event, or Finding) as described in steps 1, 2 and 3 above.

Next the sponsor must go through a process of selecting the variables to include in the new domain (Steps 4 through 7). All new domains will contain variables from the Identifier (e.g., subject of the observation, the domain) and Timing (e.g., visit designators, start and end dates) variable classes. The 'Identifier Variables' and 'Timing Variables' boxes being connected to the 'New Domain' box via a bolded solid black line illustrate this. The new domain, however, can only inherit the Topic and Qualifier variables from the specific general observation class that it was determined to fit. Therefore, if a new domain has been classified as a Findings domain, it cannot contain Qualifier variables that were defined for the Intervention or Events domain classes and vice versa. This principle is illustrated in the graphic by the 'Intervention', 'Events', and 'Findings' boxes being connected to the 'New Domain' box via dashed lines. In summary, all new domains include Identifier and Timing variables. However, the choice of Topic and Qualifier variables that can be included is dependent upon which general observation class the domain has been modeled as 'Interventions', 'Events', or 'Findings'.

3 Submitting Data in Standard Format

3.1 STANDARD METADATA FOR DATASET CONTENTS AND ATTRIBUTES

The V3.x CDISC SDS domain models provide a standard depiction of some of the most commonly used data domains, using the metadata attributes originally described in the [CDISC Submission Metadata Model](#). The descriptive metadata attributes that should be included in a submission dataset definition file as applied in the domain models are:

- The CDISC-standard variable name (standardized for all submissions, even though sponsors may be using other variable names internally in their operational database)
- The CDISC-standard variable label
- Data type of the variable (character or numeric, to conform to the data types recognized by SAS)
- Controlled terms and formats used by the sponsor (the CDISC domain models use a single asterisk (*) or a double asterisk (**)) to indicate when controlled terminology applies)
- The origin or source of the data (e.g., CRF page, derived)
- The role of the variable in the dataset corresponding to the role in the SDTM (Since these roles are predefined for all standard domains that follow the general classes, they do not need to be specified by sponsors in their Define data definition document.)

In addition to these metadata attributes, the CDISC domain models include two other columns to assist sponsors in preparing their datasets — one column for notes relevant to the use of each variable, and one to indicate how a variable is classified as a CDISC Core Variable. The concept of core variable is used both as a measure of compliance and to provide general guidance to sponsors. CDISC Core Variables fall into three categories:

- A *Required* variable is any variable that is basic to the identification of a data record (i.e., essential key variables and a topic variable) or is necessary to make the record meaningful. Required variables should always be included in the dataset and cannot be null for any record.
- An *Expected* variable is any variable necessary to make a record useful in the context of a specific domain. Columns for Expected variables are assumed to be present in each submitted dataset even if some values are null.
- A *Permissible* variable should be used in a domain as appropriate when collected or derived. All Timing variables (including those not explicitly included in a domain model) and any Qualifier variable specified in a domain model are permissible for use in that domain. Null values are allowed, but the Sponsor can decide whether a Permissible variable should be included as a column when all values for that variable are null.

The Core Variable column provides guidance on which variables are normally required, expected or permissible to include in an actual dataset based on a model (see [Section 4.1.1.5](#)). However, any decisions regarding the specific content that is to be included for any submission should always be discussed in advance with the regulatory agency.

The domain models illustrate how to apply the SDTM when creating a specific domain dataset. In particular, these models illustrate the selection of a subset of the variables offered in one of the general observation classes along with applicable timing variables. The models also show how a standard variable from a general observation class should be adjusted to meet the specific content needs of a particular domain, including making the label more meaningful, specifying controlled terminology, and creating domain-specific notes and examples. Thus the domain models demonstrate not only how to apply the model for the most common domains, but also gives insight on how to apply general model concepts to other domains not yet defined by CDISC.

3.2 USING THE CDISC DOMAIN MODELS IN REGULATORY SUBMISSIONS

New users will find the V3.x CDISC domain models useful as a template for preparing submission files; however, users should be aware that the set of descriptive attributes shown in the models is not precisely the same as that required in the dataset definition document that must accompany a submission.

The specific differences are as follows:

- The column reserved for sponsor comments in the data definition document is not included in the CDISC domain models, since this needs to be created and populated by the sponsor for each dataset to explain any sponsor-specific rules and definitions
- The CDISC Notes column should not be submitted to the agency
- The CDISC Core column should not be submitted to the agency
- The CDISC References column should not be submitted to the agency.

The last three columns in the models that are not to be submitted are shaded so they can easily be distinguished. In addition to the variable notes, the CDISC models include a set of general assumptions applicable to all domains that describe many of the critical concepts behind the models (see [Section 4](#)). A set of specific assumptions applicable only to a particular domain is listed after the data definitions for that domain model. In summary, the main body of the Define data definition document submitted by a sponsor should display for each data domain the 'Variable Name', 'Variable Label', 'Type', 'Controlled Terms or Formats', 'Origin' and 'Comments' ('Role' is optional for standard domains).

The dataset definition document that accompanies a submission should also describe each dataset that is included in the submission and describe the logical key structure of each dataset. While most studies will include DM and a set of safety domains based on the three general observation classes (typically including EX, CM, AE, DS, MH, IE, LB, and VS), the actual choice of which data to submit will depend on the protocol and the needs of the regulatory reviewer. Dataset definition metadata should include dataset filenames, descriptions, locations, structures, purpose, keys, and comments as described in the [CDISC Submission Metadata Model](#) and shown below in [Section 3.2.1](#). Note that key variables should refer to the natural keys that determine the order of records within the dataset. The physical key structure that uniquely identifies each record in an SDTM dataset based on a general class is actually based on the STUDYID, USUBJID and --SEQ variable in each domain.

3.2.1 CDISC Submission Dataset Definition Metadata

Dataset	Description	Location	Structure	Purpose	Key Variables
DM	Demographics	dm.xpt	One record per subject	Tabulation	STUDYID, USUBJID
CO	Comments	co.xpt	One record per comment per subject	Tabulation	STUDYID, USUBJID, COSEQ
CM	Concomitant Medications	cm.xpt	One record per medication intervention episode per subject	Tabulation	STUDYID, USUBJID, CMTRT, CMSTDTCT

CDISC SDTM Implementation Guide (Version 3.1.1)

Dataset	Description	Location	Structure	Purpose	Key Variables
EX	Exposure	ex.xpt	One record per constant dosing interval per subject	Tabulation	STUDYID, USUBJID, EXTRT, EXSTDTC
SU	Substance Use	su.xpt	One record per substance type per visit per subject	Tabulation	STUDYID, USUBJID, VISITNUM, SUTRT
AE	Adverse Events	ae.xpt	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AETERM, AESTDTC
DS	Disposition	ds.xpt	One record per disposition status or protocol milestone per subject	Tabulation	STUDYID, USUBJID, DSSTDTC
MH	Medical History	mh.xpt	One record per medical history event per subject	Tabulation	STUDYID, USUBJID, MHTERM
EG	ECG	eg.xpt	One record per ECG observation per time point per visit per subject	Tabulation	STUDYID, USUBJID, EGTESTCD, VISITNUM, TPTNUM, EGSEQ
IE	Inclusion/Exclusion Exceptions	ie.xpt	One record per Inclusion/Exclusion criteria exception per subject	Tabulation	STUDYID, USUBJID, IETESTCD
LB	Laboratory Tests	lb.xpt	One record per lab test per time point per visit per subject	Tabulation	STUDYID, USUBJID, LBTESTCD, VISITNUM, TPTNUM
PE	Physical Exam	pe.xpt	One record per body system per visit per subject	Tabulation	STUDYID, USUBJID, VISITNUM, PETESTCD,
QS	Questionnaires	qs.xpt	One record per question per time point per visit per subject	Tabulation	STUDYID, USUBJID, QSTESTCD, VISITNUM, TPTNUM, QSSEQ
SC	Subject Characteristics	sc.xpt	One record per characteristic per subject	Tabulation	STUDYID, USUBJID, SCTESTCD
VS	Vital Signs	vs.xpt	One record per vital sign measurement per time point per visit per subject	Tabulation	STUDYID, USUBJID, VSTESTCD, VISITNUM, TPTNUM
DV*	Protocol Deviations	dv.xpt	One record per protocol deviation per subject	Tabulation	STUDYID, USUBJID, DVSEQ
MB*	Microbiology	mb.xpt	One record per specimen test or susceptibility test	Tabulation	STUDYID, USUBJID, MBTESTCD, MBSEQ

Dataset	Description	Location	Structure	Purpose	Key Variables
MS*	Microbiology Susceptibility	ms.xpt	One record per susceptibility observations.	Tabulation	STUDYID, USUBJID, MSTESTCD, MSSEQ
DA*	Drug Accountability	da.xpt	One record per accountability observation	Tabulation	STUDYID, USUBJID, DATESTCD, DADTC, DASEQ
PC*	PK Concentrations	pc.xpt	One record per concentration per analyte	Tabulation	STUDYID, USUBJID, PCTESTCD, VISITNUM, TPTNUM
PP*	PK Parameters	pp.xpt	One record per PK parameter per analyte	Tabulation	STUDYID, USUBJID, PPTESTCD, VISITNUM
TE	Trial Elements	te.xpt	One record per element	Tabulation	STUDYID, ETCD
TA	Trial Arms	ta.xpt	One record per planned element per arm	Tabulation	STUDYID, ARMCD, ETCD
TV	Trial Visits	tv.xpt	One record per planned visit per arm	Tabulation	STUDYID, VISITNUM, ARMCD
SE	Subject Elements	se.xpt	One record per actual element per subject	Tabulation	STUDYID, USUBJID, ETCD
SV	Subject Visits	sv.xpt	One record per subject per actual visit	Tabulation	STUDYID, USUBJID, VISITNUM
TI	Trial Inclusion/ Exclusion Criteria	ti.xpt	One record per I/E criterion	Tabulation	STUDYID, IETESTCD
TS	Trial Summary	ts.xpt	One record per trial summary parameter.	Tabulation	STUDYID, TSPARMCD, TSSEQ
RELREC	Related Records	relrec.xpt	One record per relationship	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, RELID
SUPQUAL	Supplemental Qualifiers	supqual.xpt	One record per qualifier value	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM

*Domain models to be published separately for comment

3.2.2 CDISC Submission Value-Level Metadata

In general, the CDISC Version 3.x data models are more closely related to normalized relational data models in a vertical structure. This structure requires every row to have a primary key since each row of a

dataset represents a single observation and each column represents one of the variables contained within the dataset.

Since the V3.x data structures are essentially static, much of the information that may have been communicated in columns in a more horizontal structure will now involve adding new records instead. The model also includes variables that contain values that can be used to convert the dataset from the 'vertical' format to a 'horizontal' representation (or more 'denormalized'). Therefore, for some domains there is a need to provide record-level metadata (referred to as value-level or hierarchical metadata). For example, the Vital Signs data domain could contain subject records related to diastolic and systolic blood pressure, height, weight, and body mass index (BMI). If these data were submitted in compliance with the CDISC standards, it would be provided in the more normalized structure of one row per vital signs measurement. This means that there could be five records per subject (one for each parameter), and the parameter names would be stored in the Test Code/Name variables, and the parameter value in a result variable. Since the unique Test Code/Names have different attributes (i.e., different origins, roles, and definitions) there is a need to provide metadata for this information.

The hierarchical metadata could be provided as a separate section of the Define data definition document. The table should be similar in structure to the CDISC Submission Metadata Model described previously. This information, which historically has been submitted as a pdf document named 'Define.pdf' may also be submitted in an XML format. For details on the CDISC specification for submitting the Define data definition document in XML, see <http://www.cdisc.org/models/sds/v3.1/index.html>.

3.2.3 Conformance

Conformance with the CDISC Domain Models is indicated by:

- Following the complete metadata structure for data domains and variables
- Following CDISC domain models wherever applicable
- Including all collected and relevant derived data in one of the standard domains, special-purpose datasets or general-observation-class structures
- Including all Required and Expected variables as columns in a domain
- Using CDISC-specified standard domain names and prefixes
- Using CDISC-specified standard variable names
- Using CDISC-specified variable labels for all standard domains (as described in the SDTMIG)
- Using CDISC-specified data types for all variables
- Following CDISC-specified controlled terminology and format guidelines for variables, when provided
- Ensuring that each record in a dataset includes a set of keys and a topic variable
- Conforming to all business rules described in the CDISC notes and general and domain-specific assumptions.

When creating new domains, sponsors should always review the CDISC-standard labels from comparable domains and adjust only if necessary to properly convey the meaning of the submitted data. Sponsors should also supply the correct origins, roles (optional for standard domains), and appropriate comments, plus any additional controlled terminology or format information required for the FDA reviewer to properly interpret the data. Since most regulatory submissions involve data that has been collected over many years, CDISC recognizes that full conformance with the SDS model may not be immediately achievable, but will improve over time.

4 Assumptions for Domain Models

4.1 GENERAL ASSUMPTIONS FOR ALL DOMAINS

4.1.1 General Dataset Assumptions

- 4.1.1.1 Review the Study Data Tabulation Model as well as this Implementation Guide before attempting to apply any of the individual domain models. The CDISC [Submission Metadata Model](#) may also be reviewed for additional background information on how to represent metadata for general datasets. See the Case Report Tabulation Data Definition Specification (define.xml), available on the CDISC website, for information about an xml representation of the data definition document.
- 4.1.1.2 Additional analysis variables should be added to analysis datasets if required by regulatory reviewers or if necessary to accommodate the scientific requirements of the submission. These should be named consistently with other variables in the model, described clearly, and presented in a format that is consistent across a submission. Specific assumptions regarding data types should also be applied. However, no new variables should be added to any tabulation dataset except through the Supplemental Qualifiers mechanism described in [Section 8](#).
- 4.1.1.3 Additional Timing variables from the general observation classes can be added as needed to a standard domain model based on the three general observation classes.
- 4.1.1.4 The order of variables in the Define data definition document should reflect the order of data in the dataset. The current order of variables in the CDISC domain models has been chosen to facilitate the review of the models and application of the models. Sponsors may thus wish to reorder timing and qualifier variables in order to place more emphasis on the most important variables, but are encouraged to apply a consistent variable ordering scheme for all domains in a submission wherever possible.
- 4.1.1.5 CDISC Core variables: In V3.0, CDISC identified all Core variables in domain models (identified by a 'Y' in the '[Core](#)' column). In V3.x, three categories of variables are specified in the 'Core' column:
- A **Required** variable is any variable that is basic to the identification and meaning of a data record (i.e., essential key Identifiers and a Topic variable). Values cannot be null.
 - An **Expected** variable is any variable necessary to make a record meaningful in the context of a specific domain. In most cases, some but not all values for expected variables may be null in a domain if unknown or not done. But when an expected variable has not been collected, a null column should still be included and a comment should be included in the define data definition document to state it was not collected.
 - A **Permissible** variable may be used as appropriate, when collected or derived.

- All timing variables are considered permissible.
- Any qualifier variables from the same general class are permissible for that domain.

Required and expected variables must be included in the dataset. Required variables cannot be null, but the expected variables may contain some null values. In studies where limited data collection has been agreed upon in advance with the agency, some expected variables might be submitted with all null values. Normally the sponsor should only include those variables designated as permissible that were collected or derived; however, the sponsor can choose whether permissible variables that are always null should be included.

- 4.1.1.6 Additional guidance on dataset naming will be provided in future versions of the CDISC SDTMIG.

4.1.2 General Variable Assumptions

- 4.1.2.1 Data variable names should be limited to 8 characters, and cannot start with a number, nor should they contain characters other than letters, numbers, or underscores. This is a SAS V5 Transport file limitation, and since use of SAS V5 is specified in the current guidance this limitation will be in effect until the use of other formats (such as XML) becomes acceptable to regulatory authorities.

Data variable descriptive names (labels), up to 40 characters, should be provided as data variable labels. This is a correction to the 32 characters noted as a limitation in the FDA guidance document, Providing Regulatory Submissions in Electronic Format – NDAs (IT-3, January 1999).

Use of variable names (other than domain prefixes), formats, decodes, terminology, and data types for the same type of data should be consistent within and across studies within a submission. In general, sponsors should use the CDISC-standard labels in all standard domains.

- 4.1.2.2 In order to minimize the risk of difficulty when merging/joining domains for reporting purposes, the two-character Domain Identifier is used as a prefix in most variable names.

The rule for applying the prefix is as follows: All variable names should be prefixed with the Domain Identifier in which they reside except:

- a. Required Identifiers (STUDYID, DOMAIN, USUBJID)
- b. Commonly used Merge Keys (VISIT, VISITNUM, VISITDY, and many of the variables in trial design such as ELEMENT and ARM)
- c. All Demography domain (DM) variables with the exception of the DM-specific Timing variables DMDTC, and DMDY
- d. All variables in RELREC and SUPPQUAL.

Required Identifiers are not prefixed because they are usually used as keys when merging/joining observations. The --SEQ and the optional Identifiers --GRPID and --REFID are prefixed because they are only used as keys when relating observations in special cases.

- 4.1.2.3 'Subject' should be used where applicable to generically refer to both 'patients' and 'healthy volunteers.' The term 'Subject' should be used consistently in all labels and comments.

- 4.1.2.4 It is recommended that textual data be submitted in upper case text. Exceptions may include long text data (such as comment text), values of --TEST in Findings datasets (which may be more readable in mixed case if used as labels in transposed views) and certain controlled

terminology (such as MedDRA terms) that are already in mixed case. The Sponsor's Define data definition document may indicate as a general note or assumption whether case sensitivity applies to text data for any or all variables in the dataset. If case sensitivity differs across variables, it should be documented at the variable level in the Define document.

- 4.1.2.5 If a special convention is used for missing values (or values that were not collected for certain records in a Tabulation), it should be described in the Define data definition document and should be used consistently in all datasets.
- 4.1.2.6 Sponsors may assign categories (--CAT variable) and subcategories (--SCAT variable) within domains. The categories and subcategories will provide additional context for the Topic variable and should be used consistently across all records in the domain. For example, a lab record with LBTEST = 'SODIUM' might have LBCAT = 'CHEMISTRY' and LBSCAT = 'ELECTROLYTES'. The values for these variables may be collected on the CRF or derived. --CAT and --SCAT values should not be redundant with the domain or dictionary classification provided by --DECOD and --BODSYS. That is, they should provide a different means of defining or classifying records. For example, a sponsor may be interested in identifying all adverse events that the investigator feels might represent an infection, and thus will collect that categorization on the CRF. This categorization might differ from the categorization derived from the adverse-event coding dictionary.

The intended relationship among the records grouped using --GRPID (see [Section 8.1](#) for details) may appear to duplicate a relationship already identified by other variables such as VISIT, --CAT, --SCAT, or --CLAS. These latter variables have meaning across subjects, whereas the values of --GRPID apply to a specific domain and subject, and are intended solely to facilitate grouping a set of records for a subject. For example, in the Laboratory Test Results dataset, LBTEST values of HEMATOCRIT, HEMOGLOBIN, and RBC might be assigned a value of HEMATOLOGY for LBCAT. LBCAT would be populated for all records falling into that category across all subjects. However, for one subject, the investigator commented on these specific records noting that the subject had a long history of anemia. For this subject, the three hematology records would be assigned a common LBGRPID value. That value would be used in the COMMENTS special-purpose domain to link the comment to the appropriate lab records.

- 4.1.2.7 Sponsors sometimes collect free text information on the CRF. When the CRF includes a list of values including "Other, Specify", then the free text value should be placed in a SUPQUAL dataset as [described in Section 8.4](#). For example, when a description of "Other Medically Important Serious Adverse Events" category is collected on a CRF then the free text description should be stored in the SUPQUAL dataset.

4.1.3 Coding and Controlled Terminology Assumptions

- 4.1.3.1 The Submission Data Standards Team has defined two types of Controlled Terminology (CT). The first consists of CT that has been published externally (e.g., MedDRA, LOINC or maintained by CDISC in [Appendix 10.4](#)). The second consists of CT that is defined by a sponsor, and is used consistently throughout a submission (e.g., a standard code list). Two asterisks (**) within the Controlled Terms or Format column indicate the respective variable should be populated with values from an external source. A single asterisk (*) within the Controlled Terms or Format column indicates the respective variable should be populated with a value from a sponsor-controlled list. The values that appear in the Controlled Terms or Format column are expected to be used in submissions. CDISC will be providing future guidance on the use of controlled terminology in separate documents available through the CDISC website. If applicable, additional information is provided within this column for common alpha-codes, decodes, and references.

- 4.1.3.2 It is recommended that controlled terminology be submitted in upper case text other than the two exceptions noted below. Some sponsors may choose to submit controlled terminology in mixed case text to enhance the appearance of reports. As noted in [item 4.1.2.4](#) above, case sensitivity should be described in the Define document.
- a. If the external reference for the controlled terminology is not in upper case then the data should conform to the case prescribed in the external reference (e.g., MedDRA and LOINC).
 - b. Units, which are considered symbols rather than abbreviated text (e.g., mg/dL).
- 4.1.3.3 The controlled terminology should be displayed for each applicable variable within the 'Controlled Terms or Format' column of the Define data definition document. Variables with numerous controlled terms may be placed within an appendix section of Define.
- 4.1.3.4 Controlled terminology or text should be used instead of, or in addition to, arbitrary number codes in order to reduce ambiguity for submission reviewers. For example, for concomitant medications, the verbatim term and/or dictionary term should be presented, rather than numeric codes. Separate coded values may be submitted as supplemental qualifiers or included in the define data definition document and may be necessary in analysis datasets.
- 4.1.3.5 Controlled terminology for domain topic variables should be stored as follows:
- For safety events such as AEs and Medical History, fill --DECOD with the dictionary's preferred term and fill --BODSYS with the preferred body system name. If a dictionary is multi-axial, the value in --BODSYS should represent the primary path. When using MedDRA, for example, --DECOD should contain the PT (Preferred Term), and --BODSYS should contain the primary SOC (System Organ Class).
 - For concomitant medications, fill --DECOD with the drug's generic name, and fill --CLAS with the drug class only if the dictionary used codes drugs to a single class. When using WHODRUG, for example, --CLAS would not be filled since a drug may have multiple classes.

In either case, no other intermediate levels or relationships should be stored in the dataset. By knowing the dictionary and version used, the reviewer will be able to obtain intermediate levels in a hierarchy (as in MedDRA), or a drug's ATC codes (as in WHO Drug). The sponsor may be required to submit the dictionary if it is not already available to the reviewer.

- 4.1.3.6 The topic variable for many of the general domain models is often stored as verbatim text. For an Event domain, the topic variable is --TERM. For an Interventions domain, the topic variable is --TRT. For a Findings domain, the topic variable, --TESTCD, should use Controlled Terminology (e.g., SYSBP for systolic blood pressure). If a CDISC-standard controlled terminology exists it should be used, otherwise the sponsor should define its own controlled list of terms. If the verbatim term is modified to facilitate coding, the modified text is stored in --MODIFY. In most cases (other than PE), the dictionary coded text is derived into --DECOD. Since the PEORRES variable is modified instead of the topic variable for PE, the dictionary-derived text would be placed in PESTRESC. The items used in each of the defined domains are:

AE:	AETERM/AEMODIFY/AEDECOD
DS:	DSTERM/DSDECOD
CM:	CMTRT/CMMODIFY/CMDECOD
MH:	MHTERM/MHMODIFY/MHDECOD
PE:	PEORRES/PEMODIFY/PESTRESC

- 4.1.3.7 Variables where the response is 'Yes' or 'No' ('Y'/'N') should normally be populated for both 'Y' and 'N' responses. This eliminates confusion regarding whether a blank response indicates 'N' or is a missing value. However, some variables are collected or derived in a manner that allows only one response, such as when a single check box indicates 'Yes'. In situations such as these, where it is unambiguous to only populate the response of interest, it is permissible to only populate one value ('Y' or 'N') and leave the alternate value blank. An example of when it would be acceptable to use only a value of 'Y' would be for Baseline Flag (BLFL) variables, where 'N' is not necessary to indicate that a value is not a baseline value.

Note: Permissible values for variables with controlled terms of 'Y' or 'N' may be extended to include 'U' if it is the sponsor's practice to explicitly collect or derive values indicating 'Unknown' for that variable.

4.1.4 Actual and Relative Time Assumptions

- 4.1.4.1 Date/Time formats for --DTC Variables: The CDISC SDS V2 models included date variables and some separate time variables. In Version 3.0, to address the FDA's request to provide a uniform date/time representation, all date and time variables were replaced with a single date/time variable ending in 'DTM'. These date and time values were stored as SAS date/times, which in SAS are stored as the number of seconds since midnight on January 1, 1960. A SAS format was required to display the date/times in human readable form. This was changed in version 3.1. Dates and time of day are now stored according to the international standard ISO 8601 (ISO, URL: <http://www.iso.ch/iso/en/ISOOnline.openerpage>).

The SDTMIG template uses ISO 8601 for calendar dates and times of day, which are expressed as follows:

YYYY-MM-DDThh:mm:ss

where:

[YYYY] = four-digit year

[MM] = two-digit representation of the month (01-12, 01=January, etc.)

[DD] = two-digit day of the month (01 through 31)

[T] = (time designator) indicates time information follows

[hh] = two digits of hour (00 through 23) (am/pm is NOT allowed)

[mm] = two digits of minute (00 through 59)

[ss] = two digits of second (00 through 59)

Other characters defined for use within the ISO 8601 standard are:

[-] (hyphen): to separate the time elements "year" and "month", "month" and "day"

[:] (colon): to separate the time elements "hour" and "minute", and "minute" and "second"

[/] (solidus): to separate components in the representation of time intervals

[P] (duration designator): precedes the components that represent the duration

Key aspects of the ISO 8601 standard are as follows:

- ISO 8601 represents dates as a text string using the notation YYYY-MM-DD.
- ISO 8601 represents times as a text string using the notation hh:mm:ss.

- ISO 8601 does not require dashes to separate date parts nor colons to separate time parts; however, these are required for the SDTM and V3.x SDTMIG to improve readability.
- When a date is stored with a time in the same variable (as a date-time), the date is written in front of the time and the time is preceded with 'T' using the notation YYYY-MM-DDT hh:mm:ss (e.g. 2001-12-26T00:00:01).

Implementation of the ISO standard means that nominal date/time variables are no longer numeric data types but will now have to be character/text data types. The SDS fragment employed for date/time variables is --DTC, where '--' is the two letter domain code [with or without the 'ST' or 'EN' strings indicating start or stop dates].

4.1.4.2 Date/Time Precision. In V3.0, a separate date/time precision variable ending in 'DTP' was paired with each date/time variable to address the FDA desire to easily detect any value that was partially or fully imputed. This has been eliminated in V3.x. According to ISO 8601, precision (also referred to by ISO 8601 as "completeness" or "representations with reduced accuracy"), is specified by the number of parts present in the date, time, or date-time value. Whatever is unknown is not included so if the date is completely missing then the date value should be null. For example, if month is unknown, then only year is included in the value string. If day is unknown, only year and month are included in the value string. If seconds are unknown, all parts of the date-time string are included except the last two numbers representing seconds. If day is known but month is unknown then neither day nor month is included in the date representation. Furthermore, one-digit numbers are always padded with a leading zero. Hence, the use of a separate precision variable, as defined in Version 3.0, is no longer required.

The table below provides examples of ISO 8601 representation complete date and truncated date/time values using ISO 8601 "appropriate right truncations" of incomplete datetime representations.

	Date and Time as Originally Recorded	Interval of Uncertainty	Nominal Date/Time (--DTC)
1	December 15, 2003 13:14:17	Complete date and time	2003-12-15T13:14:17
2	December 15, 2003 13:14	Unknown seconds	2003-12-15T13:14
3	December 15, 2003 13	Unknown minutes	2003-12-15T13
4	December 15, 2003	Unknown time	2003-12-15
5	December, 2003	Unknown day	2003-12
6	2003	Unknown month	2003

This date and date/time model provides for nonstandard intervals of uncertainty, such as those commonly seen in Medical History. To represent these intervals while applying the ISO8601 standard, it is recommended that the sponsor concatenate the date/time values (using the most complete representation of the date/time known) that describe the beginning and the end of the intervals of uncertainty and separate them with a solidus as shown in the table below:

	Interval of Uncertainty	Nominal Date/Time (--DTC)
1	Between 10:00 and 10:30 on the Morning of December 15, 2003	2003-12-15T10:00/2003-12-15T10:30
2	After the first of this year (2003) until "now" (February 15, 2003, noon)	2003-01-01/2003-02-15T12:00
3	Between the first and the tenth of December, 2003	2003-12-01/2003-12-10
4	Sometime in the first half of 2003	2003-01-01/2003-06-30

Other uncertainty intervals may be represented by the omission of components of the date when these components are unknown. Examples of this method of omitted component representation are represented in the table below:

	Date and Time as Originally Recorded	Interval of Uncertainty	Nominal Date/Time (--DTC)
1	December 15, 2003 13:15:17	Complete date and time	2003-12-15T13:15:17
2	December 15, 2003 __:15	Unknown hour with known minutes	2003-12-15T15T-:15
3	December 15, 2003 13:__:17	Unknown minutes with known date, hours and seconds	2003-12-15T13:-:17
4	The 15 th of some month in 2003	Unknown month and time with known year and day	2003---15
5	December 15, but can't remember the year	Unknown year with known month and day	--12-15

Using a character-based data type to implement the ISO 8601 date/time standard will ensure that the date/time information will be machine and human readable without the need for further manipulation, and will be platform and software independent.

NOTE FOR BACKWARD COMPATIBILITY WITH SDS V. 3.0:

Because the --DT fragment was used in the SDS Version 2 standards, and the DTM fragment was used to represent numeric based date/time values in Version 3, the "DTC" fragment has been used to identify character-based date/time information in Version 3.1. This will allow sponsors to continue to use the --DT and --DTM fragments for variables used in operational databases or analysis files.

The SAS format IS8601DT can be used to display SAS date/time information in the ISO 8601 format. Examples are shown below:

	Date and time as originally recorded	SAS Numeric Date/Time (--DTM)¹	Formatted SAS Date/Time²
1	29June, 1956 11:32:09	-110636871	1956-06-29T11:32:09
2	29June, 1956 11:32:00	-110636880	1956-06-29T11:32:00
3	29June, 1956 11:00:00	-110638800	1956-06-29T11:00:00
4	29June, 1956 00:00:00	-110678400	1956-06-29T00:00:00
5	June1956	Not calculated ³	Not calculated ³

¹ DTM = input(DTC, datetime18.)

² format of --DTM IS8601DT

³ SAS requires a full, valid date. Apply a standard convention to impute days.

4.1.4.3 Duration is frequently used during a review; however, the duration timing variable (--DUR) should generally be used in a domain if it was collected in lieu of a start date/time (--STDTC) and end date/time (--ENDTC). If both --STDTC and --ENDTC are collected, durations can be calculated by the difference in these two values, and need not be in the submission dataset.

In V3.0, there were separate duration (--DUR) and duration unit (--DURU) variables. In V3.x, the --DURU variable is eliminated, as both duration and duration units can be provided in the single --DUR variable according to ISO 8601. The values provided in --DUR should follow one of the following ISO 8601 formats:

PnYnMnDTnHnMnS
or
PnW

The letter "P" precedes other values in the duration format. The 'n' preceding each letter represents the number of Years, Months, Days, Hours, Minutes, Seconds, or Weeks. As with the date/time format,

'T' is used to separate the date and time components. Note, weeks cannot be mixed with days in representations

When duration is being measured after an event, the correct usage when you know the start date/time from which the duration is measured is:

YYYY-MM-DDThh:mm:ss/PnYnMnDTnHnMnS

For duration measured prior to an event (like 2 hours before last dose) the syntax is:

PnYnMnDTnHnMnS/YYYY-MM-DDThh:mm:ss

Note only the hyphen delimited extended format conforming to "Format with time-unit designators" (ISO 5.5.4.2.1 - second edition 2000-12-15) is considered valid for SDTM use.

Usage Notes:

- There should be no spaces in the representation.
- The 'T' designator must always precede Time components, but is not necessary if no time components are included.
- Missing components can be omitted.

The table below provides some examples of ISO 8601 compliant durations:

Duration as originally recorded	--DUR Value
2 Years	P2Y
10 weeks	P10W
3 Months 14 days	P3M14D
3 Days	P3D
6 Months 17 Days 3 Hours	P6M17DT3H
14 Days 7 Hours 57 Minutes	P14DT7H57M
42 Minutes 18 Seconds	PT42M18S
2 hours before a reference time point in RFTDTC	-PT2H

4.1.4.4 The Study Day variable (--DY) describes the relative day of the observation starting with the reference date as Day 1. It is determined by comparing the date portion of any DTC variable to the date portion of the Subject Reference Date (RFSTDTC from the Demography domain). It is considered a descriptive representation of a relative date within the study.

The subject reference date is designated as Study Day 1. The Study Day value is incremented by 1 for each date following RFSTDTC. Dates prior to RFSTDTC are decremented by 1, with the date preceding RFSTDTC designated as Study Day -1 (there is no Study Day 0). This algorithm for determining Study Day is consistent with how people typically describe sequential days relative to a fixed reference point, but creates problems if used for mathematical calculations because it does not allow for a Day 0. As such, Study Day is not suited for use in subsequent numerical computations, such as calculating duration. The raw date values should be used rather than Study Day in those calculations.

All Study Day values are integers. Thus, to calculate Study Day:

--DY = (date portion of --DTC) - (date portion of RFSTDTC) + 1 if --DTC is on or after RFSTDTC
 --DY = (date portion of --DTC) - (date portion of RFSTDTC) if --DTC precedes RFSTDTC

The algorithm for this calculation should be consistent across all domains.

4.1.4.5 Clinical encounters are described by the CDISC Visit variables. VISITNUM indicates the clinical encounter number, a numeric version of VISIT used for sorting. VISITDY indicates the planned study day of VISIT. A text description of the visit (VISIT) is recommended for intelligibility and consistency with the protocol and CRF.

VISITNUM is required when data are collected more than once per subject or at a discrete time point (e.g., Labs, ECG, Vital Signs or other domains with multiple assessment points).

VISITNUM is not included in Subject Characteristics because the data in this domain is collected only once per subject. It is permissible in Disposition if more than one disposition event is collected per subject. VISITNUM is not required in Adverse Events, Concomitant Medication, or Medical History.

VISIT and VISITDY are permissible, but not required when VISITNUM is used. The following table shows an example of how the visit identifiers might be used:

USUBJID	VISIT	VISITNUM	VISITDY	--DY
001	Week 1	2	7	7
001	Week 2	3	14	13
001	Week 2 Unscheduled	3.1	(Null)	17

4.1.4.6 The calculation of study days within subdivisions of time in a clinical trial may be based on one or more sponsor-defined reference dates not adequately represented by RFSTDTC. The Sponsor's Define data definition document should reflect the reference dates used to calculate such study days. CDISC will define an appropriate model for representing such sponsor-defined dates in a future version of this document.

4.1.4.7 The CDISC V2/V3 domain models included the timing variables --ONGO and --PRIOR, but the meaning of these variables was found to differ dramatically among different sponsors, and even across different trials for the same sponsor. For example, AEONGO often meant an event was ongoing as of the end of a study, while MHONGO meant a condition was ongoing at the beginning of the study. The meaning of these variables was further confused because it was not always apparent as to which time point to use to define the beginning or end of a study. To clarify this, V3.1 of the SDTMIG defined two new timing variables, --STRF, Start Relative to Reference Period, and --ENRF, End Relative to Reference Period.

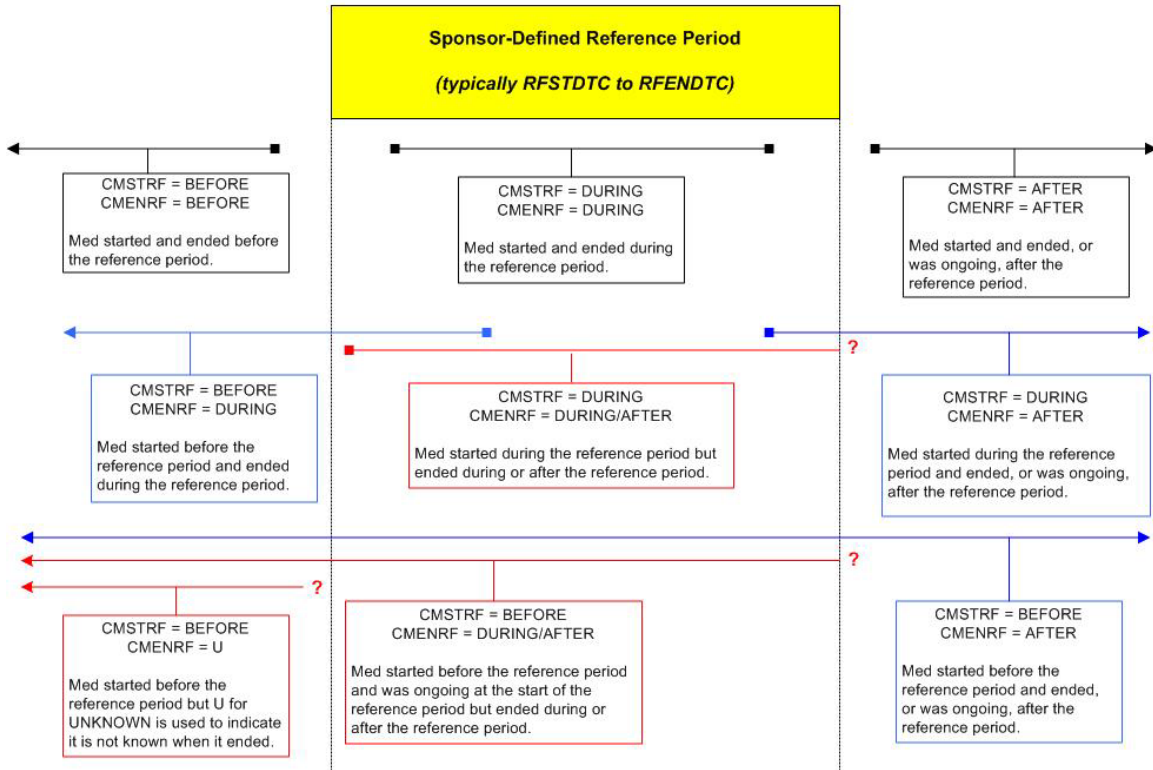
--STRF identifies the start of an observation (Event, Intervention, or Finding) as being 'BEFORE', 'DURING', or 'AFTER' the sponsor-defined reference period. The sponsor-defined reference period is a continuous period of time defined by a discrete starting point (RFSTDTC) and a discrete ending point (RFENDTC). --STRF should be populated for observations with information such as check boxes on the CRF (e.g., 'PRIOR', 'ONGOING', or 'CONTINUING') that indicates when they occur in reference to the study timeline. --STRF should be set to 'U' if its value is unknown.

--ENRF identifies the end of the observation (Event, Intervention, or Finding) as being 'BEFORE', 'DURING', 'DURING/AFTER' or 'AFTER' the sponsor-defined reference period. The sponsor-defined reference period is a continuous period of time defined by a discrete starting point (RFSTDTC) and a discrete ending point (RFENDTC). --ENRF should be populated for observations with information such as check boxes on the CRF (e.g. 'PRIOR', 'ONGOING', or 'CONTINUING') that indicate when they occur in reference to the study timeline. --ENRF should be set to 'U' if its value is unknown.

Figure 4.1.4.7 below illustrates how to populate these variables in a CM domain. For both variables, sponsors should define the reference period in the study metadata.

Figure 4.1.4.7 Example of CMSTRF and CMENRF Variables

Graphical depiction of use of CMSTRF and CMENRF timing variables



4.1.4.8 When the date and time of collection is reported in a domain based on the Findings general class, the date should go into the --DTC field -- e.g., EGDTC (Date/Time of ECG). When an interval (start and end) date/time is reported, the start date and time goes into --DTC and the end date and time goes into --ENDTC. For example, in the LB domain the LBDTC date variable is used for all single point collections or spot urine collections. For timed lab collections (e.g., 24 hour urine collections) the LBDTC variable is used for the start date/time of the collection and LBENDTC for the end date/time of the collection. This approach will allow the single point and interval collections to use the same date/time variables consistently across all datasets for the findings general class. The table below illustrates the proper use of these variables. Note --STDTC is not used for collection dates, so is blank in the following table.

	--DTC	--STDTC	--ENDTC
Date Collected	X		
Interval Collected	X		X

4.1.4.9 Dates are generally used only as timing variables to describe the timing of an event, intervention or collection activity, but there may be occasions when it may be preferable to model a date as a result in a Findings dataset. Note that using a date as a result to a findings question is unusual and atypical, and should be approached with caution, but this situation may occasionally occur when a) a group of questions (each of which has a date response) is asked and analyzed together; or b) the event(s) and intervention(s) in question are not medically significant (often the case when included in questionnaires). Consider the following cases:

- Calculated due date?
- Date of last day on the job?
- Date of birth of youngest child?
- End date of last reporting period?

One approach to modeling these data would be to place the text of the question in --TEST and the response to the question, a date represented in ISO 8601 format, in --ORRES and --STRESC as long as these date results do not contain the dates of medically significant events or interventions.

Again, use extreme caution when storing dates as the results of findings. Remember, in most cases, these dates should be timing variables associated with a record in an Intervention or Events dataset.

4.1.5 Other Assumptions

4.1.5.1 Original and Standardized Results of Findings

The --ORRES variable contains the result of the measurement or finding as originally received or collected. --ORRES is an expected variable, and should always be populated with two exceptions:

- When --STAT = 'NOT DONE'
- When a record is derived (e.g., to represent an average or sum of collected values), in such a way as to make it impossible to meaningfully populate --ORRES.

Derived records are flagged with the --DRVFL variable. When the derived record comes from more than one visit, the sponsor must define the value for VISITNUM, addressing the correct temporal sequence.

When --ORRES is populated, --STRESC must also be populated, regardless of whether the data values are character or numeric.--STRESC is derived either by the conversion of values in --ORRES to values with standard units, or by the assignment of the value of --ORRES (as in the PE Domain, where --STRESC could contain a dictionary-derived term). A further step is necessary when --STRESC contains numeric values. These are converted to numeric type and written to --STRESN. Because --STRESC may contain a mixture of numeric and character values, --STRESN may contain null values, as shown in the flowchart below.



When the original measurement or finding is a selection from a defined code list, the --ORRES or --STRESC variables should contain results in decoded format, that is, the textual interpretation of whichever code was selected from the code list.

Occasionally data that are intended to be numeric are collected with characters attached that cause the character-to-numeric conversion to fail. For example, numeric cell counts in the source data may be specified with a greater than (>) or less than (<) sign attached (e.g. >10,000 or <1). In order for these values to be reported as numeric in --STRESN, and assuming this is desirable, the sponsor must apply a rule to derive a numeric value for these strings and document this in the value-level metadata. For example the rule may be to remove the character symbol or to compensate for the symbol in some other way, such as adding to values using > or subtracting from values using <.

Tests Not Done

When an entire examination (Laboratory draw, ECG, Vital Signs or Physical Examination), or a group of tests (hematology or urinalysis), or an individual test (glucose, PR interval, blood pressure, or hearing) is not done, and this information is explicitly captured with a yes/no or done/not done question on the CRF, a record could be created in the dataset to record this information. See the examples below.

If the data on the CRF is missing and yes/no or done/not done was not explicitly captured a record should not be created to indicate that the data was not collected.

If a group of tests were not done

- --TESTCD should be --ALL
- --TEST should be <Name of the Module>
- --CAT should be <Name of Group of Tests>
- --ORRES should be NULL
- --STAT should be "NOT DONE"

For example, if urinalysis is not done then

- LBTESTCD should be "LBALL"
- LBTEST should be "LAB"
- LBCAT should be "URINALYSIS"
- LBORRES should be NULL
- LBSTAT should be "NOT DONE".

Laboratory Examples:

- Numeric values that have been converted or copied (Rows 1 & 3).
- A character result that has been copied (Row 2).
- A result of 'TRACE' converted to a numeric value of 1 (Row 4).
- Value of 1+ converted to numeric 1 (Row 5).
- A result of 'BLQ' converted to a numeric value of zero (Row 6).
- Some original character results in LBORRES such as '<10,000', which are converted to numeric results per sponsor-defined algorithms (Row 7).
- A result is missing because the observation was 'NOT DONE', as reflected in the --STAT variable; neither LBORRES nor LBSTRESC are populated (Row 8).
- A result is derived from multiple records such as an average of baseline measurements for a baseline value (Row 9).
- None of the scheduled tests were completed as planned (Row 10)
- A group of tests were not completed as planned (Row 11)

Row	LBTESTCD	LBSTAT	LBCAT	LBORRES	LBORRESU	LBSTRESC	LBSTRESN	LBSTRESU	LBDRVFL
1	GLUCOSE			6.0	mg/dL	60.0	60.0	mg/L	
2	UBAC			MODERATE		MODERATE			
3	SGPT			12.1	mg/L	12.1	12.1	mg/L	
4	URBC			TRACE		1	1		
5	UWBC			1+		1	1		
6	KETONE			BLQ		0	0		
7	WBC			<10,000		[sponsor-defined value]	[sponsor-defined value]		
8	HCT	NOT DONE							
9	MCHC					33.8	33.8	g/dL	Y
10	LBALL	NOT DONE							
11	LBALL	NOT DONE	HEMATOLOGY						

ECG Examples:

- Numeric and character values that have converted or copied (Rows 1, 2, and 3).
- A result is missing because the test was 'NOT DONE', as reflected in the EGSTAT variable; neither EGORRES nor EGSTRESC are populated (Row 4).
- The overall interpretation is included as a new record (Row 5)
- The entire ECG was not done (Row 6)

Row	EGTESTCD	EGSTAT	EGORRES	EGORRESU	EGSTRESC	EGSTRESN	EGSTRESU
1	QRS		362	SEC	362	362	SEC
2	QT		221	MSEC	.221	.221	SEC
3	RHYMRATE		ATRIAL FLUTTER		ATRIAL FLUTTER		
4	PR	NOT DONE					
5	INTP		ABNORMAL, CLINICALLY SIGNIFICANT		ABNORMAL, CLINICALLY SIGNIFICANT		
6	EGALL	NOT DONE					

Vital Signs Examples:

- Numeric and character values that have converted or copied (Rows 1 and 2).
- A result is missing because the Vital Signs test was 'NOT DONE', as reflected in the VSSTAT variable; neither VSORRES nor VSSTRESC are populated (Row 3).
- The result is derived by having multiple records for one measurement and the derived value is recorded in a new row with the derived record flagged. (Rows 4, 5, and 6).
- The entire examination was not done (Row 7)

Row	VSTESTCD	VSSTAT	VSORRES	VSORRESU	VSSTRESC	VSSTRESN	VSSTRESU	VSDRVFL
1	HT		60	IN	152	152	cm	
2	WT		110	LB	50	50	kg	
3	HR	NOT DONE						
4	SYSBP		96	mmHg	96	96	mmHg	
5	SYSBP		100	mmHg	100	100	mmHg	
6	SYSBP				98	98	mmHg	Y
7	VSALL	NOT DONE						

Questionnaire Examples:

- Character values that have converted to standard scores (Row1).
- A result is derived from multiple records (Row 2)
- A result is missing because the observation was 'NOT DONE', as reflected in the QSSTAT variable; neither QSORRES nor QSSTRESC are populated (Row 3).
- The entire questionnaire was not done (Row 4)

Row	QSTEST	QSSTAT	QSORRES	QSSTRESC	QSSTRESN	QSDRVFL
1	HEALTH		VERY GOOD	4.4	4.4	
2	HEALTH PERCEPTIONS (0-100)				82	Y
3	HEALTH	NOT DONE				
4	QSALL	NOT DONE				

4.1.5.2 Multiple observations may be linked (associated with each other) using the RELREC dataset described in [Section 8 of this document](#). Comments should be stored in the Comments domain (CO). Additional Qualifier variables that were not specified in the model should be included in the Supplemental Qualifiers (SUPPQUAL) dataset.

4.1.5.3 CDISC realizes that some sponsors may have text strings longer than 200 characters for some variables. Because of the current requirement for Version 5 SAS transport file format, it will not be possible to store those long text strings using only one variable. Therefore, CDISC has defined a convention for storing a long text string by using a combination of the standard domain dataset and the Supplemental Qualifiers (SUPPQUAL) dataset. For a text string of more than 200 characters in length, the first 200 characters of text should be stored in the standard domain variable and each additional 200 characters of text should be stored as a record in the SUPPQUAL dataset. In this dataset, the value for QNAM should contain a sequential variable name, which is formed by appending a one-digit integer, beginning with 1, to the original standard domain variable name. When splitting a text string into several records, the text should be split between words to improve readability.

As an example, if a sponsor had text for Medical History Reported Term (MHTERM) for one subject of 500 characters in length, the sponsor would put the first 200 characters of text in the standard domain variable and dataset (MHTERM in MH), the next 200 characters of text as a first supplemental record in the SUPPQUAL dataset, and the final 100 characters of text as a second record in the SUPPQUAL dataset. Variable QNAM would have the values MHTERM1 and MHTERM2 for the first and second records in SUPPQUAL, respectively, for this one particular text string. Sponsors should place the text itself into variable QVAL and the label of the original standard domain variable into variable QLABEL. In this case, IDVAR and IDVARVAL should be used in SUPPQUAL to relate the associated supplemental text records to the first 200 characters of text in the standard domain.

In cases where the standard domain variable name is already 8 characters in length, sponsors should replace the last character with a digit when creating values for QNAM. As an example, for Other Action Taken in Adverse Events (AEACNOTH), values for QNAM for the SUPPQUAL records would have the values AEACNOT1, AEACNOT2, and so on.

Examples of how SUPPQUAL would be populated are presented below:

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
12345	MH	99-123	MHSEQ	6	MHTERM1	Reported Term for Condition or Procedure	2nd 200 chars of text	CRF	
12345	MH	99-123	MHSEQ	6	MHTERM2	Reported Term for Condition or Procedure	last 200 chars of text	CRF	
12345	AE	99-123	AESEQ	4	AEACNOT1	Other Action Taken	last 200 chars of text	CRF	

4.1.5.4 For the Interventions and Events observation classes, all data are assumed to be attributed to the Principal Investigator, or derived from such data by the sponsor. The observations recorded in the Findings class include the --EVAL qualifier because the observation may originate from more than one source (e.g., an Investigator or Central Reviewer). The QEVAL variable can be used to describe the evaluator for any data item in SUPPQUAL, but is not required when the data is objective. For all observations that have a primary and supplemental evaluation, sponsors should always put data from the primary evaluation into the standard domain dataset and data from the supplemental evaluation into the Supplemental Qualifiers dataset (SUPPQUAL – see [Section 8.4](#)). In SUPPQUAL, each variable for the evaluation should be represented as a record. Within each record, the value for QNAM should be formed by appending a “1” to the corresponding standard domain variable name. In cases where the standard domain variable name is already 8 characters in length, sponsors should replace the last character with a “1” (incremented for each additional attribution). The result of, response to, or value associated with QNAM should be placed in variable QVAL. The label of the corresponding standard domain variable should be placed in variable QLABEL. As an example, if an adjudication committee evaluates an adverse event, data in SUPPQUAL for this evaluation would look as below. (Note that QNAM takes on the value AERELNS1, as the corresponding standard domain variable AERELNST is already 8 characters in length.) The adverse event data as determined by the primary investigator would reside in the standard AE dataset.

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
12345	AE	99-123	AESEQ	3	AESEV1	Severity/Intensity	MILD	CRF	ADJUDICATION COMMITTEE
12345	AE	99-123	AESEQ	3	AEREL1	Causality	POSSIBLY RELATED	CRF	ADJUDICATION COMMITTEE
12345	AE	99-123	AESEQ	3	AERELNS1	Relationship to Non-Study Treatment	Possibly related to aspirin use.	CRF	ADJUDICATION COMMITTEE

5 Models for Special Purpose Domains

5.1.1 Demographics Domain Model — DM

dm.xpt, Demographics — Version 3.1.1, August 26, 2005. One record per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**DM	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission. This can be a unique number or a compound identifier formed by concatenating STUDYID-SITEID-SUBJID. Since a subject may participate in more than one study (as in a follow-on), the USUBJID may differ from the SUBJID in some cases.	Req	SDTM 2.2.4
SUBJID	Subject Identifier for the Study	Char		CRF	Topic	Subject identifier used within the study. Often the ID of the subject as recorded on a CRF.	Req	
RFSTDTC	Subject Reference Start Date/Time	Char	ISO 8601	Sponsor Defined	Timing	Reference Start Date/time for the subject in ISO 8601 character format. Usually equivalent to date/time of first intake of drug. Required for all randomized subjects; null for screen failures (if screen failures are submitted).	Exp	SDTMIG 4.1.4.1
RFENDTC	Subject Reference End Date/Time	Char	ISO 8601	Sponsor Defined	Timing	Reference End Date/time for the subject in ISO 8601 character format. Usually equivalent to the date/time when subject was determined to have ended the trial, and often equivalent to date/time of last intake of drug. Required for all randomized subjects; null for screen failures (if screen failures are submitted).	Exp	SDTMIG 4.1.4.1
SITEID	Study Site Identifier	Char		CRF or Derived	Record Qualifier	Unique identifier for a study site within a submission.	Req	
INVID	Investigator Identifier	Char	*	CRF or Derived	Record Qualifier	An identifier to describe the Investigator for the study. May be used in addition to the SITEID. Not needed if SITEID is equivalent to INVID.	Perm	
INVNAM	Investigator Name	Char		CRF or Derived	Synonym Qualifier	Name of the investigator for a site.	Perm	
BIRTHDTC	Date/Time of Birth	Char	ISO 8601	CRF or Derived	Result Qualifier	Date/time of birth of the subject.	Perm	SDTMIG 4.1.4.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
AGE	Age in AGEU at RFSTDTC	Num		CRF or Derived	Result Qualifier	Usually derived as (RFSTDTC -BRTHDTC), but BRTHDTC may not be available in all cases (due to subject privacy concerns).	Exp	
AGEU	Age Units	Char	** YEARS, MONTHS, DAYS, HOURS, WEEKS	CRF or Derived	Variable Qualifier		Exp	
SEX	Sex	Char	**M, F, U	CRF	Result Qualifier	M, F, U for Male, Female, Unknown	Req	
RACE	Race	Char	**	CRF	Result Qualifier	Race of the subject. In cases of mixed race, primary race should be listed. If additional race values are collected, they should be represented as separate records in the Subject Characteristics domain with TESTCD of 'RACEOTH'. The variable itself is required by FDA, but with ongoing discussions in Europe about the permissibility of collecting this data, the variable may be optional in future.	Exp	
ETHNIC	Ethnicity	Char	**	CRF	Result Qualifier	Ethnicity of subject.	Perm	
ARMCD	Planned Arm Code	Char	*	CRF or Derived	Result Qualifier	Short version of ARM, used for sorting and programming (Formerly TRTCD).	Req	
ARM	Description of Planned Arm	Char	*	CRF or Derived	Synonym Qualifier	Name given to the arm the subject was assigned to (Formerly TRTGRP).	Req	
COUNTRY	Country	Char	**ISO 3166 3-char. code	CRF or Derived	Result Qualifier	Country of the investigational site in which the subject participated in the trial.	Req	
DMDTC	Date/Time of Collection	Char	ISO 8601	CRF or Derived	Timing	Use if collected.	Perm	SDTMIG 4.1.4.1
DMDY	Study Day of Collection	Num		Derived	Timing	1. Study day of collection of the demographic information, measured as integer days. 2. Algorithm for calculations must be relative to the sponsor-defined RFSTDTC variable in Demographics. This formula should be consistent across the submission.	Perm	SDTMIG 4.1.4.4

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

5.1.1.1 Assumptions for Demographics (DM) domain model

1. Investigator (INVID) and/or site (SITEID) identification: Companies use different methods to distinguish sites and/or investigators. CDISC assumes that site will always be present, with investigator used as necessary to group data. This should be done consistently for the submission and the meaning of the variable made clear in the Define data definition document. Investigator name (INVNAM) may also be included as a support variable.
2. Subject identification: It is presumed that every subject in a study will have a subject identifier (SUBJID) and that in some cases a subject may be included in more than one study within a submission. To identify a subject uniquely across a submission, a unique identifier (USUBJID) should be assigned and included in all datasets within the submission.
3. Recent ongoing discussions regarding subject privacy suggest that care must be taken with regard to the collection of variables like BRTHDTC. This variable is included in the Demographics model as an example of a variable that might be included in a submission; however, sponsors should follow regulatory guidelines and guidance as appropriate.
4. Arm/Treatment identification: Note that in Version 3.x the variables ARM and ARMCD have replaced TRTGRP and TRTCD for consistency with the trial design, and to more clearly distinguish the planned arm from the actual treatment. When a sponsor is submitting trial design information, the values of ARM and ARMCD should be identical to the values defined for that subject in the Subject Elements (SE) dataset. The assignment of values should be consistent, if possible, within a submission.
5. Justification for using SEX vs. GENDER: Page 71 of 'Providing Regulatory Submissions in Electronic Format - NDAs' (IT-3, January, 1999), available at <http://www.fda.gov/cder/guidance/2353fml.pdf> specifically lists SEX as part of demographic data. Similarly, page 60 of 'Guidance for Industry, Providing Regulatory Submissions to the Center for Biologics Evaluation in Electronic Format - Biologics Marketing Applications' (November, 1999), available at <http://www.fda.gov/cber/guidelines.htm> specifically lists SEX as part of demographic data. SEX is used consistently in both documents except for one instance where GENDER is used (page 30 for Table 6 which may have been from another writer). 'ICH E3: Structure and Content of Clinical Study Reports' (November 30, 1999) only uses SEX (not GENDER).
6. Attributions used to classify study populations for review and analysis, such as the COMPLT, SAFETY, ITT and PPROT variables discussed in the V3 DM domain, should be placed in the 'suppdm.xpt' SUPPQUAL dataset as [described in Section 8.4](#).
7. When additional free text information is reported about subject's RACE using 'Other, Specify', sponsors should place the free text value in the Subject Characteristics (SC) dataset with TESTCD of 'RACEOTH' as described in the CDISC note for RACE.
8. Data for screen failure subjects, if submitted, should be included in the Demographics dataset, with ARMCD = "SCRNFAIL" and ARM = "Screen Failure" to distinguish screen failures from randomized subjects. Sponsors may include a record in the Disposition dataset indicating when the screen dropout event occurred. See [section "9.3.2 DS Examples"](#) for examples of screen dropout records.

5.1.2 Comments Domain Model — CO

The COMMENTS special-purpose domain is a fixed domain that provides a solution for submitting free-text comments related to data in one or more SDS domains or collected on a separate CRF page dedicated to comments. Comments are generally not responses to specific questions; instead, comments usually consist of voluntary, free-text, unsolicited observations. COMMENTS has a structure that is similar to the Supplemental Qualifiers (SUPPQUAL) special-purpose dataset but it allows for one comment to span multiple variables (COVAL-COVALn) in order to accommodate comments longer than 200 characters.¹ Comments may be related to a Subject or to specific parent records in an SDTM domain. The approach for relating comments is described in [Section 8.5](#).

co.xpt -- Version 3.1.1. August 26, 2005. One record per comment per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**CO	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
RDOMAIN	Related Domain Abbreviation	Char	**	Derived	Identifier	Domain abbreviation of the parent record(s). Null for comments collected on a general comments or additional information CRF page.	Exp	
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
COSEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence Number given to ensure uniqueness within a domain. Can be used to join related records	Req	
IDVAR	Identifying Variable	Char	**	Derived	Identifier	Identifying variable in the dataset that identifies the related record(s). Examples --SEQ, or --GRPID. Null for comments collected on separate CRFs.	Perm	
IDVARVAL	Identifying Variable Value	Char		Derived	Identifier	Value of identifying variable of the parent record(s). Used only when individual comments are related to domain records. Null for comments collected on separate CRFs	Perm	
COREF	Comment Reference	Char	*	CRF	Record Qualifier	CRF page of parent record(s) to which the comment refers, or on which the comment was collected. COREF may be the page number, e.g. 650, or the page name e.g. DEMOG, or a combination of information that uniquely identifies the reference both e.g. 650-VITALS- VISIT 2.	Perm	
CODTC	Date/Time of Comment	Char	ISO 8601	CRF	Timing		Perm	SDTMIG 4.1.4.1
COVAL	Comment	Char		CRF	Result Qualifier	The text of the comment. COVAL cannot be null – a value is required for the record to be valid. The first 200 characters of the comment will be in COVAL, then next 200 in COVAL1, and continuing as needed to COVALn.	Req	
COEVAL	Evaluator	Char	*	CRF or Derived	Record Qualifier	Used to describe the originator of the comment. Controlled terminology will consist of values such as ADJUDICATION COMMITTEE, STATISTICIAN, DATABASE ADMINISTRATOR, CLINICAL COORDINATOR, and PRINCIPAL INVESTIGATOR.	Perm	

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

¹ Allowing for multiple variables per comment is an interim solution until the limitations posed by SAS® Version 5 transport files are eliminated.

5.1.2.1 Assumptions for Comments (CO) domain model

1. The COMMENTS dataset accommodates two sources of comments: 1) those collected alongside other data on topical CRF pages such as Adverse Events; and 2) those collected on a separate page specifically dedicated to comments. The value of the variable RDOMAIN for comments of the first type should be the domain code of the parent record(s). The value of the variable RDOMAIN of the second type should be null. Assumptions for the use of GRPID and SEQ as values for IDVAR are the same as for the Supplemental Qualifiers (SUPPQUAL) special purpose dataset described in [Section 8](#), except that when there is no parent record (i.e., for comments collected on a dedicated comments CRF) IDVAR and IDVARVAL should be NULL. COMMENTS records without a parent record (i.e. those collected on a dedicated Comments CRF) should have CODTC filled if it is collected on the CRF. COMMENTS records with a parent record inherit timing variables from the parent, thus CODTC is NULL in COMMENTS. The variable COREF may be NULL unless it is used to identify the source of the comment.
2. Example of use of COMMENTS special-purpose dataset --Subject with four comments: One comment collected on each of the AE, EX, and VS pages, and another collected on a separate Comments CRF
 - 1) Columns 3 and 6 (RDOMAIN and IDVAR) represent information from the parent record, if applicable.
 - 2) Row 4 contains general comments from the Comments CRF Page. There is no parent record for this comment, thus DOMAIN has the value CO and RDOMAIN, IDVAR and IDVARVAL are NULL.
 - 3) Row 1 contains a comment text > 400 characters; Rows 2 and 3 contains text between 200 and 400 characters, Row 4 contains 150 characters of comment text.

ROW	STUDYID	DOMAIN	RDOMAIN	USUBJID	COSEQ	IDVAR	IDVARVAL	COREF	COVAL	COVAL1	COVAL2	COEVAL
1	1234	CO	AE	AB-99	1	AESEQ	7	650	First 200 characters	Next 200 characters	More Text	PRINCIPAL INVESTIGATOR
2	1234	CO	EX	AB-99	2	EXGRPID	17	320-355	First 200 characters	Next 200 characters		PRINCIPAL INVESTIGATOR
3	1234	CO	VS	AB-99	3	VSGRPID	5		First 200 characters	More text		PRINCIPAL INVESTIGATOR
4	1234	CO		AB-99	4			800	Total 150 characters			PRINCIPAL INVESTIGATOR

6 Domain Models Based on the General Classes

6.1 INTERVENTIONS

6.1.1 Concomitant Medications — CM

cm.xpt, Concomitant Medications — Interventions, Version 3.1.1, August 26, 2005. One record per medication intervention episode per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**CM	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
CMSEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
CMGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to tie together a block of related records in a single domain to support relationships within the domain and between domains.	Perm	SDTMIG 2.1 ; SDTM 2.2.4
CMSPID	Sponsor-Defined Identifier	Char		Sponsor Defined	Identifier	Optional sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: line number on a concomitant medication page.	Perm	SDTM 2.2.4

CDISC SDTM Implementation Guide (Version 3.1.1)

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
CMTRT	Reported Name of Drug, Med, or Therapy	Char		CRF	Topic	Verbatim medication name that is either pre-printed or collected on a CRF.	Req	SDTMIG 4.1.3.6
CMMODIFY	Modified Reported Name	Char		Sponsor Defined	Synonym Qualifier	If CMTRT is modified, then CMMODIFY will contain the modified text.	Perm	SDTMIG 4.1.3.5
CMDECOD	Standardized Medication Name	Char	**	Derived	Synonym Qualifier	Standardized or dictionary-derived text description of CMTRT or CMMODIFY. Equivalent to the generic medication name in WHO Drug. The sponsor should specify the dictionary name and version in the Sponsor Comments column of the Define document. If an intervention term does not have a decode value in the dictionary then CMDECOD will be left blank.	Perm	SDTMIG 4.1.3.5
CMCAT	Category for Medication	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of medications/treatments. Examples: ANTI-CANCER MEDICATION, or GENERAL CONMED.	Perm	SDTMIG 2.1
CMSCAT	Subcategory for Medication	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of medications/ treatment. Examples: CHEMOTHERAPY, RADIOTHERAPY, HORMONAL THERAPY, ALTERNATIVE THERAPY.	Perm	SDTMIG 2.1
CMOCCUR	CM Occurrence	Char	**Y, N or Null	CRF or Sponsor Defined	Record Qualifier	Used when the use of specific medications is solicited To indicate whether a medication was taken or not. Values are null for spontaneously reported medications. <i>Note: Use of this variable is under evaluation by the SDS Team.</i>	Perm	
CMSTAT	Concomitant Medication Status	Char	**NOT DONE	CRF or Sponsor Defined	Record Qualifier	The status indicates that the question was not asked.	Perm	SDTMIG 4.1.5.1.
CMREASND	Reason Medication Not Collected	Char		CRF	Record Qualifier	Describes the reason concomitant medication was not collected. Used in conjunction with CMSTAT when value is NOT DONE.	Perm	
CMINDC	Indication	Char		CRF or Derived	Record Qualifier	Denotes why a medication was taken or administered. Examples: NAUSEA, HYPERTENSION.	Perm	
CMCLAS	Medication Class	Char	**	Derived	Variable Qualifier	Use only when the dictionary used codes to a single class. If using a dictionary that allows links to multiple classes, then omit CMCLAS from the dataset. For example, sponsors who use WHO Drug, which allows links from a medication to multiple ATC codes, would not include CMCLAS.	Perm	SDTMIG 4.1.3.5

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
CMCLASCD	Medication Class Code	Char	**	Derived	Variable Qualifier	Use only when the dictionary used codes to a single class.	Perm	SDTMIG 4.1.3.5
CMDOSE	Dose per Administration	Num		CRF or Sponsor Defined	Record Qualifier	Amount of CMTRT taken.	Perm	
CMDOSTXT	Dose Description	Char		CRF or Derived	Record Qualifier	Dosing amounts or a range of dosing information collected in text form. Example: 200-400.	Perm	
CMDOSU	Dose Units	Char	*	CRF or Sponsor Defined	Variable Qualifier	Units for CMDOSE and CMDOSTOT. Examples: ng, mg, or mg/kg.	Perm	
CMDOSFRM	Dose Form	Char	*	CRF or Sponsor Defined	Variable Qualifier	Dose form for CMTRT. Examples: TABLET, LOTION.	Perm	
CMDOSFRQ	Dosing Frequency Per Interval	Char	*	CRF or Sponsor Defined	Variable Qualifier	Usually expressed as the number of dosings given per a specific interval. Examples: BID, QID.	Perm	
CMDOSTOT	Total Daily Dose using CMDOSU	Num		CRF or Derived	Record Qualifier	Total daily dose of CMTRT using the units in CMDOSU.	Perm	
CMDOSRGM	Intended Dose Regimen	Char		CRF or Derived	Variable Qualifier	Text description of the (intended) schedule or regimen for the Intervention. Examples: TWO WEEKS ON, TWO WEEKS OFF.	Perm	
CMROUTE	Route of Administration	Char	*	CRF or Sponsor Defined	Variable Qualifier	Route of administration for CMTRT. Examples: ORAL, INTRAVENOUS.	Perm	
CMSTDTC	Start Date/Time of Medication	Char	ISO 8601	CRF	Timing		Perm	SDTMIG 4.1.4.1
CMENDTC	End Date/Time of Medication	Char	ISO 8601	CRF	Timing		Perm	SDTMIG 4.1.4.1
CMSTDY	Study Day of Start of Medication	Num		Derived	Timing	Study day of start of medication relative to the sponsor-defined RFSTDTC.	Perm	SDTMIG 4.1.4.1
CMENDY	Study Day of End of Medication	Num		Derived	Timing	Study day of end of medication relative to the sponsor-defined RFSTDTC.	Perm	SDTMIG 4.1.4.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
CMDUR	Duration of Medication	Char	ISO 8601	CRF	Timing	Collected duration and unit of a treatment. Used only if collected on the CRF and not derived from start and end date/times.	Perm	
CMSTRF	Start Relative to Reference Period	Char	** BEFORE, DURING, AFTER	Derived	Timing	Identifies the start of the medication with respect to the sponsor-defined reference period. Sponsors should define the reference period in the study metadata. CMSTRF should be populated when a start date is not collected. If information such as "PRIOR", "ONGOING", or "CONTINUING" was collected, this information should be translated into CMSTRF.	Perm	SDTMIG 4.1.4.7
CMENRF	End Relative to Reference Period	Char	** BEFORE, DURING, AFTER, DURING/AFTER, U	Derived	Timing	Identifies the end of the medication with respect to the sponsor-defined reference period. Sponsors should define the reference period in the study metadata. Medications that are ongoing at the end of the reference period should have a value of AFTER for this variable. CMENRF should be populated when an end date is not collected. If information such as "PRIOR", "ONGOING", or "CONTINUING" was collected, this information should be translated into CMENRF.	Perm	SDTMIG 4.1.4.7

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

6.1.1.1 Assumptions for Concomitant Medications (CM) domain model

1. CM Definition

- a. CRF data that captures the Concomitant and Prior Medications/Therapies used by the subject. Examples are the Concomitant Medications/Therapies given on an as needed basis and the usual and background medications/therapies given for a condition.
- b. This domain should contain one record per each Concomitant Medication/Therapy intervention episode.

2. Concomitant Medications Description and Coding

- a. CMTRT captures the name of the Concomitant Medications/Therapy and it is the topic variable. It is a required variable and must have a value.
- b. CMMODIFY should be included if the sponsor's procedure permits modification of a verbatim term for coding. The modified term is listed in CMMODIFY.
- c. CMDECOD is the standardized medication/therapy term derived by the sponsor from the coding dictionary. It is expected that the reported term (CMTRT) or the modified term (CMMODIFY) will be coded using a standard dictionary. The sponsor is expected to provide the dictionary name and version used to map the terms in the metadata (using the Comments column in the Define document). If a sponsor chooses to use more than one dictionary to code different types of data within the same dataset (e.g., drug and non-drug therapies), then the variables CMCAT or CMSCAT can be used to categorize the data by type. The metadata will specify which dictionary is applied to each category.
- d. CMCLAS is used only when the dictionary used codes to a single class. If using a dictionary that allows links to multiple classes, then omit CMCLAS from the submission dataset.

3. Additional Categorization and Grouping
 - a. CMCAT and CMSCAT may be used as categorization variables to group data across subjects. Examples of CMCAT could include values such as ANTI-CANCER MEDICATION, or GENERAL CONMED. CMSCAT may be used to further categorize the Concomitant Medications/Therapy taken by the subject. Examples: CHEMOTHERAPY, RADIOTHERAPY, HORMONAL THERAPY or ALTERNATIVE THERAPY. Neither CMCAT nor CMSCAT are expected variables in the submission dataset.
 - b. CMGRPID is used to link a block of related records at the subject level in a single domain to support relationships within the domain and between domains.

4. Presence or Absence of Concomitant Medications
 - a. Information regarding Concomitant Medications/Therapy is generally collected in two different ways, either by recording free text or using a pre-specified list of terms. In the latter case, the solicitation of information on specific concomitant medications may affect the frequency at which they are reported; therefore, the fact that use of a specific medication was solicited may be of interest to reviewers. Consequently, the SDS Team is proposing the addition of a new variable --PRESP to the SDTM to indicate that the use of those specific medications was solicited. Until then, CMPRESP can be used as a supplemental qualifier.
 - b. CMOCCUR may be used when a CRF includes a pre-specified list of Concomitant Medications/Therapies to indicate whether or not a medication/therapy was used. Proper use of this variable is under evaluation by the SDS Team.

5. ConMed Structure
 - a. The structure of the CM domain is one record per medication intervention episode per subject. It is the sponsor's responsibility to define an intervention episode. This definition may vary based on the sponsor's requirements for review and analysis. The submission dataset structure may differ from the structure at the time of collection. One common approach is to submit a new record when there is a change in the dosing regimen. Another approach is to collapse all records for a medication to a summary level with either a dose range or the highest dose level. Other approaches, including a visit-based record, may also be reasonable as long as they meet the sponsor's evaluation requirements.

6.1.2 Exposure — EX

ex.xpt, Exposure — Interventions, Version 3.1.1 August 26, 2005. One record per constant dosing interval per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**EX	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
EXSEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
EXGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to tie together a block of related records in a single domain to support relationships within the domain and between domains.	Perm	SDTMIG 2.1 ; SDTM 2.2.4
EXSPID	Sponsor-Defined Identifier	Char		Sponsor Defined	Identifier	Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Line number on a CRF Page.	Perm	SDTM 2.2.4
EXTRT	Name of Actual Treatment	Char		CRF or Randomization File or Derived	Topic	Name of the intervention treatment — usually the verbatim name of the investigational treatment given during the dosing period for the observation.	Req	SDTMIG 4.1.3.6
EXCAT	Category for Treatment	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records. Example: COMPARATOR CLASS.	Perm	SDTMIG 2.1
EXSCAT	Subcategory for Treatment	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of treatment.	Perm	SDTMIG 2.1
EXDOSE	Dose per Administration	Num		CRF or Sponsor Defined	Record Qualifier	Amount of EXTRT administered or given.	Exp	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
EXDOSTXT	Dose Description	Char		CRF or Derived	Record Qualifier	Dosing amounts or a range of dosing information collected in text form. Example: 200-400.	Perm	
EXDOSU	Dose Units	Char	*	CRF or Sponsor Defined	Variable Qualifier	Units for EXDOSE and EXDOSTOT. Examples: ng, mg, or mg/kg.	Exp	
EXDOSFRM	Dose Form	Char	*	CRF or Sponsor Defined	Variable Qualifier	Dose form for EXTRT. Examples: TABLET, LOTION,.	Req	
EXDOSFRQ	Dosing Frequency Per Interval	Char	*	CRF or Sponsor Defined	Variable Qualifier	Usually expressed as the number of dosings given per a specific interval. Examples: BID, QID.	Perm	
EXDOSTOT	Total Daily Dose Using EXDOSU	Num		CRF or Derived	Record Qualifier	Total daily dose of EXTRT using the units in EXDOSU.	Perm	
EXDOSRGM	Intended Dose Regimen	Char		CRF or Derived	Variable Qualifier	Text description of the (intended) schedule or regimen for the Intervention. Examples: two weeks on, two weeks off.	Perm	
EXROUTE	Route of Administration	Char	*	CRF or Sponsor Defined	Variable Qualifier	Route of administration for EXTRT. Examples: ORAL, INTRAVENOUS.	Perm	
EXLOT	Lot Number	Char		CRF	Record Qualifier	Lot Number of the EXTRT product.	Perm	
EXLOC	Location of Dose Administration	Char	*	CRF or Sponsor Defined	Record Qualifier	Specifies anatomical location of administration Example: LEFT ARM for a topical application.	Perm	
EXTRTV	Treatment Vehicle	Char	*	CRF or Sponsor Defined	Record Qualifier	Describes vehicle used for treatment. Example: SALINE.	Perm	
EXADJ	Reason for Dose Adjustment	Char		CRF or Sponsor Defined	Record Qualifier	Describes reason or explanation of why a dose is adjusted – used only when an adjustment is represented in EX. May be used for variations from protocol-specified doses, or changes from expected doses.	Perm	
TAETORD	Order of Element within Arm	Num		Sponsor Defined / Protocol	Timing	Number that gives the order of the Element within the arm.	Perm	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
EXSTDTC	Start Date/Time of Treatment	Char	ISO 8601	CRF or Derived	Timing	The time when the dose was administered.	Req	SDTMIG 4.1.4.1
EXENDTC	End Date/Time of Treatment	Char	ISO 8601	CRF or Derived	Timing	May not be relevant for injections.	Perm	SDTMIG 4.1.4.1
EXSTDY	Study Day of Start of Treatment	Num		Derived	Timing	Study day of start of treatment relative to the sponsor-defined RFSTDTC.	Perm	SDTMIG 4.1.4.1
EXENDY	Study Day of End of Treatment	Num		Derived	Timing	Study day of end of treatment relative to the sponsor-defined RFSTDTC.	Perm	SDTMIG 4.1.4.1
EXDUR	Duration of Treatment	Char	ISO 8601	CRF	Timing	Collected duration and unit of a treatment. Used only if collected on the CRF and not derived from start and end date/times.	Perm	
EXTPT	Planned Time Point Name	Char		CRF or Derived	Timing	1. Text Description of time when a dose should be given. 2. This may be represented as an elapsed time relative to a fixed reference point, such as time of last dose. See EXTPTNUM and EXPTREF. Examples: Start or 5 min post.	Perm	
EXTPTNUM	Planned Time Point Number	Num		Sponsor Defined	Timing	Numerical version of EXTPT to aid in sorting.	Perm	
EXELTM	Planned Elapsed Time from Reference Pt	Char	ISO 8601	Sponsor Defined	Timing	Elapsed time (in ISO 8601) format relative to the planned fixed reference (EXPTREF). This variable is useful where there are repetitive measures. Not a clock time.	Perm	
EXPTREF	Time Point Reference	Char		Sponsor Defined	Timing	Name of the fixed reference point referred to by EXELTM, EXTPTNUM, and EXTPT. Examples: Previous Dose, Previous Meal.	Perm	

- *indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.*

6.1.2.1 Assumptions for Exposure (EX) domain model

1. EX Definition

- The Exposure domain model records the details of a subject's exposure to protocol-specified study treatment. Study treatment may be any intervention that is prospectively defined as a test material within a study, and is typically but not always supplied to the subject. Examples include but are not limited to placebo, active comparator, and investigational product. Treatments that are not protocol-specified should be recorded in the Concomitant Medication (CM) domain.

- b. This domain should contain one record per constant dosing interval per subject. "Constant dosing interval" is sponsor-defined, and may include any period of time that can be described in terms of a known treatment given at a consistent dose. e.g., for a study with once-a-week administration of a standard dose for 6 weeks, exposure may be represented with a single record per subject, spanning the entire treatment phase. Or if the sponsor monitors each treatment administration and deviations in treatment or dose occur, there could be up to six records (one for each weekly administration).

2. Categorization and Grouping

- a. EXCAT and EXSCAT may be used when appropriate to categorize treatments into categories and subcategories. For example, if a study contains several active comparator medications, EXCAT may be set to 'ACTIVE COMPARATOR.' Since such categorization may not apply to many studies, these variables are permissible but not expected.

6.1.3 Substance Use - SU

su.xpt, Substance Use — Interventions, Version 3.1.1, August 26, 2005. One record per substance type per visit per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**SU	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
SUSEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
SUGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to tie together a block of related records in a single domain to support relationships within the domain and between domains.	Perm	SDTMIG 2.1 ; SDTM 2.2.4
SUSPID	Sponsor-Defined Identifier	Char		Sponsor Defined	Identifier	Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Line number on a Tobacco & Alcohol use CRF page.	Perm	SDTM 2.2.4
SUTRT	Reported Name of Substance	Char		CRF	Topic	Substance name. Examples: Cigarettes, Coffee.	Req	SDTMIG 4.1.3.6
SUMODIFY	Modified Substance Name	Char		Sponsor Defined	Synonym Qualifier	If SUTRT is modified, then the modified text is placed here.	Perm	SDTMIG 4.1.3.5
SUDECOD	Standardized Substance Name	Char	**	Derived	Synonym Qualifier	Standardized or dictionary-derived text description of SUTRT or SUMODIFY if the sponsor chooses to code the substance use. The sponsor should specify the dictionary name and version in the Sponsor Comments column of the Define document. If an intervention term does not have a decode value in the dictionary then SUDECOD will be null	Perm	SDTMIG 4.1.3.5
SUCAT	Category for Substance Use	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records. Examples: TOBACCO, ALCOHOL, or CAFFEINE.	Perm	SDTMIG 2.1
SUSCAT	Subcategory for Substance Use	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of substance use. Examples: CIGARS, CIGARETTES, BEER, WINE	Perm	SDTMIG 2.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
SUOCCUR	SU Occurrence	Char	**Y, N or Null	CRF or Sponsor Defined	Record Qualifier	Used when the use of specific substances is solicited to indicate whether a substance was taken or not. Values are null for spontaneously reported substances. <i>Note: Use of this variable is under evaluation by the SDS Team.</i>	Perm	
SUSTAT	Substance Use Status	Char	**NOT DONE	CRF	Record Qualifier	Used to indicate that the question was not asked.	Perm	
SUREASND	Reason Substance Use Not Collected	Char		CRF	Record Qualifier	Describes the reason substance use was not collected. Used in conjunction with SUSTAT when value is NOT DONE.	Perm	SDTMIG 4.1.5.3
SUCLAS	Substance Use Class	Char	*	Derived	Variable Qualifier	May be used when coding certain substance use cases such as alcoholism or drug abuse.	Perm	SDTMIG 4.1.3.5
SUCLASCD	Substance Use Class Code	Char	*	Derived	Variable Qualifier	May apply when coding substance abuse use cases.	Perm	SDTMIG 4.1.3.1
SUDOSE	Substance Use Consumption	Num		CRF or Sponsor Defined	Record Qualifier	Amount of SUTRT consumed.	Perm	
SUDOSTXT	Substance Use Consumption Text	Char		CRF or Derived	Record Qualifier	Substance use consumption amounts or a range of consumption information collected in text form.	Perm	
SUDOSU	Consumption Units	Char	*	CRF or Sponsor Defined	Variable Qualifier	Units for SUDOSE. Examples: OUNCES, CIGARETTE EQUIVALENTS, or GRAMS.	Perm	
SUDOSFRM	Dose Form	Char	*	CRF or Sponsor Defined	Variable Qualifier	Dose form for SUTRT. Examples: INJECTABLE, LIQUID, or POWDER.	Perm	
SUDOSFRQ	Use Frequency Per Interval	Char	*	CRF or Sponsor Defined	Variable Qualifier	Usually expressed as the number of uses consumed per a specific interval. Examples: PER DAY, PER WEEK, OCCASIONAL.	Perm	
SUDOSTOT	Total Daily Consumption using SUDOSU	Num		CRF or Derived	Record Qualifier	Total daily use of SUTRT using the units in SUDOSU.	Perm	
SURROUTE	Route of Administration	Char	*	CRF or Sponsor Defined	Variable Qualifier	Route of administration for SUTRT. Examples: ORAL, INTRAVENOUS, INHALATION.	Perm	
VISIT	Visit Name	Char		CRF or Derived	Timing	1. Protocol-defined description of clinical encounter. 2. May be used in addition to VISITNUM and/or VISITDY.	Perm	SDTMIG 7.8; SDTMIG 4.1.4.5

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
VISITNUM	Visit Number	Num		CRF or Derived	Timing	1. Clinical encounter number. 2. Numeric version of VISIT, used for sorting.	Req	SDTMIG 7.8; SDTMIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDTMIG 7.8; SDTMIG 4.1.4.5
SUSTDTC	Start Date/Time of Substance Use	Char	ISO 8601	CRF or Derived	Timing		Perm	SDTMIG 4.1.4.1
SUENDTC	End Date/Time of Substance Use	Char	ISO 8601	CRF or Derived	Timing		Perm	SDTMIG 4.1.4.1
SUSTDY	Study Day of Start of Substance Use	Num		Derived	Timing	Study day of start of substance use relative to the sponsor-defined RFSTDTC.	Perm	SDTMIG 4.1.4.1
SUENDY	Study Day of End of Substance Use	Num		Derived	Timing	Study day of end of substance use relative to the sponsor-defined RFSTDTC.	Perm	SDTMIG 4.1.4.1
SUDUR	Duration of Substance Use	Char	ISO 8601	CRF	Timing	Collected duration and unit of substance use. Used only if collected on the CRF and not derived from start and end date/times.	Perm	
SUSTRF	Start Relative to Reference Period	Char	** BEFORE, DURING, AFTER	Derived	Timing	Identifies the start of the substance use period with respect to the sponsor-defined reference period. Sponsors should define the reference period in the study metadata. SUSTRF should be populated when a start date is not collected. If information such as "PRIOR", "ONGOING", or "CONTINUING" was collected, this information should be translated into SUSTRF.	Perm	SDTMIG 4.1.4.7
SUENRF	End Relative to Reference Period	Char	** BEFORE, DURING, AFTER, DURING/ AFTER, U	Derived	Timing	Identifies the end of the substance use period with respect to the sponsor-defined reference period. Sponsors should define the reference period in the study metadata. Use of substances that is ongoing at the end of the reference period should have a value of AFTER for this variable. SUENRF should be populated when an end date is not collected. If information such as "PRIOR", "ONGOING", or "CONTINUING" was collected, this information should be translated into SUENRF.	Perm	SDTMIG 4.1.4.7

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

6.1.3.1 Assumptions for Substance Use (SU) domain model

1. The intent of the domain is to capture substance use information that may be used to assess the efficacy and/or safety of therapies that look to mitigate the effects of chronic substance use, or that could be used as covariates to support the efficacy and/or safety analyses of other topic variables.
2. SU Definition
 - a. This information may be independent of planned study evaluations or may be a key outcome (e.g., planned evaluation) of a clinical trial.
 - b. In many clinical trials detailed substance use information as represented by this domain may not be required (e.g., the only information collected may be a response to the question ‘Have you ever smoked tobacco?’); in this case many of the qualifying variables would not be submitted.
3. Substance Use Description and Coding
 - a. SUTRT captures the verbatim term collected for the substance. It is the topic variable for the SU dataset. SUTRT is a required variable and must have a value.
 - b. SUMODIFY is a permissible variable and should be included if the sponsor’s procedure permits modification of a verbatim substance use term for coding. The modified term is listed in SUMODIFY. The variable may be populated as per the sponsor’s procedures; null values are permitted.
 - c. SUDECOD is the preferred term derived by the sponsor from the coding dictionary. It is a permissible variable; Null values are permitted. Where deemed necessary by the sponsor the verbatim term (SUTRT) should be coded using a standard dictionary such as WHODRUG. The sponsor is expected to provide the dictionary name and version used to map the terms in the metadata (using the Comments column in the Define document).
 - d. SUCLAS and SUCLASCD can be used to classify substances using a dictionary that codes to a single class. These variables would not be used if the sponsor were to use WHODRUG to classify substances.
4. Additional Categorization and Grouping
 - a. SUCAT and SUSCAT should not be redundant with the domain code or dictionary classification provided by SUDECOD. That is, they should provide a different means of defining or classifying SU records. For example, a sponsor may be interested in identifying all SU records that the investigator feels might represent opium use and thus will collect that categorization on the CRF. This categorization might differ from the categorization derived from the coding dictionary. It is not necessary to populate these variables.
 - b. SUGRPID may be used to link (or associate) different records together to form a block of related records at the subject level within SU. It should not be used in place of SUCAT or SUSCAT, which are used to group data across subjects. It is not necessary to populate this variable.
5. Presence or Absence of Substance Use
 - a. If specific substances are solicited (i.e., a pre-specified list of terms is collected), SUOCCUR may be used to indicate whether or not each substance was consumed. Use of this variable is under evaluation by the SDS Team.
 - b. For cases when it is important to know whether the occurrence of specific substances is solicited (i.e., a pre-specified list of terms is collected), the SDS Team is proposing the addition of a new variable --PRESF to the SDTM. Until then, SUPRESF can be used as a supplemental qualifier.
6. Timing Variables
 - a. SUSTDTC and SUENDTC may be populated as required. Null values are permissible.
 - b. If substance use information is collected more than once within the CRF (indicating that the data are visit-based) or collected only once within the CRF then VISITNUM is required. VISITDY and VISIT are always permissible.
7. Other Qualifier Variables
 - a. In those instances where detailed dosing information is collected, the variables SUDOSE, SUDOSTXT, SUDOSU, SUDOSFRM, SUDOSFRQ, SUDOSTOT and SUROUTE may be populated; null values are permitted.
 - b. When the response for a substance use question on a CRF was not collected, then the value of SUSTAT = 'NOT DONE' and SUOCCUR will be null.

6.2 EVENTS

6.2.1 Adverse Events — AE

ae.xpt, Adverse Events — Events, Version 3.1.1, August 26, 2005. One record per adverse event per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**AE	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
AESEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
AEGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to tie together a block of related records in a single domain to support relationships within the domain and between domains.	Perm	SDTM 2.2.4
AEREFID	Reference ID	Char		Sponsor Defined	Identifier	Optional internal or external identifier such as a serial number on an SAE reporting form	Perm	SDTM 2.2.4
AESPID	Sponsor-Defined Identifier	Char		Sponsor Defined	Identifier	Optional Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Line number on a Adverse Events page.	Perm	SDTM 2.2.4
AETERM	Reported Term for the Adverse Event	Char		CRF	Topic	Verbatim name of the event.	Req	SDTMIG 4.1.3.6
AEMODIFY	Modified Reported Term	Char		Sponsor Defined	Synonym Qualifier	If AETERM is modified, then AEMODIFY will contain the modified text.	Perm	SDTMIG 4.1.3.6
AEDECOD	Dictionary-Derived Term	Char	**	Derived	Synonym Qualifier	Dictionary-derived text description of AETERM or AEMODIFY. Equivalent to the Preferred Term (PT in MedDRA). The sponsor should specify the dictionary name and version in the Sponsor Comments column of the Define document.	Req	SDTMIG 4.1.3.5; SDTMIG 4.1.3.6
AECAT	Category for Adverse Event	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records. Example: BLEEDING, HYPOGLYCEMIA.	Perm	SDTMIG 4.1.2.6

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
AESCAT	Subcategory for Adverse Event	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of adverse event. Example: NEUROLOGIC.	Perm	SDTMIG 4.1.2.6
AEOCCUR	Adverse Event Occurrence	Char	**Y, N or Null	CRF or Sponsor Defined	Record Qualifier	Used when the occurrence of specific adverse events is solicited to indicate whether an adverse event occurred or not. Values are null for spontaneously reported events. <i>Note: Use of this variable is under evaluation by the SDS Team.</i>	Perm	
AEBODSYS	Body System or Organ Class	Char	**	CRF or Derived	Record Qualifier	Body system or organ class (Primary SOC) that is involved in an event or measurement from the standard hierarchy (e.g., MedDRA).	Exp	SDTMIG 4.1.3.5
AELOC	Location of the Reaction	Char	*	CRF or Derived	Record Qualifier	Describes anatomical location relevant for the event. (e.g., LEFT ARM for skin rash).	Perm	
AESEV	Severity/Intensity	Char	*	CRF	Record Qualifier	The severity or intensity of the event. Examples: MILD, MODERATE, SEVERE.	Perm	
AESER	Serious Event	Char	**Y, N	CRF or Derived	Record Qualifier	Is this is a serious event?	Exp	
AEACN	Action Taken with Study Treatment	Char	*	CRF	Record Qualifier	Describes changes to the study treatment as a result of the event. Examples include ICH E2B values: DRUG WITHDRAWN, DOSE REDUCED, DOSE INCREASED, DOSE NOT CHANGED, UNKNOWN or NOT APPLICABLE	Exp	
AEACNOTH	Other Action Taken	Char		CRF	Record Qualifier	Describes other actions taken as a result of the event. Usually reported as free text. Example: "Treatment unblinded. Primary care physician notified."	Perm	
AEREL	Causality	Char	*	CRF	Record Qualifier	Records the investigator's opinion as to the causality of the event to the treatment. Examples: NOT RELATED, UNLIKELY RELATED, POSSIBLY RELATED, RELATED	Exp	
AERELNST	Relationship to Non-Study Treatment	Char		CRF	Record Qualifier	Records the investigator's opinion as to whether the event may have been due to a treatment other than study drug. Reported as free text. Example: "More likely related to aspirin use".	Perm	
AEPATT	Pattern of Adverse Event	Char	*	CRF	Record Qualifier	Used to indicate the pattern of the event over time. Examples: INTERMITTENT, CONTINUOUS, SINGLE EVENT.	Perm	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
AEOUT	Outcome of Adverse Event	Char	*	CRF	Record Qualifier	Description of the outcome of an event. Examples include ICH E2B values: RECOVERED/RESOLVED, RECOVERING/RESOLVING, NOT RECOVERED/NOT RESOLVED, RECOVERED/RESOLVED WITH SEQUELAE, FATAL or UNKNOWN.	Perm	
AESCAN	Involves Cancer	Char	**Y, N or Null	CRF	Record Qualifier	Was the event associated with the development of cancer?	Perm	
AESCONG	Congenital Anomaly or Birth Defect	Char	**Y, N or Null	CRF	Record Qualifier	Was the event associated with congenital anomaly or birth defect?	Perm	
AESDISAB	Persist or Signif Disability/Incapacity	Char	**Y, N or Null	CRF	Record Qualifier	Did the event result in persistent or significant disability/incapacity?	Perm	
AESDTH	Results in Death	Char	**Y, N or Null	CRF or Derived	Record Qualifier	Did the event result in death?	Perm	
AESHOSP	Requires or Prolongs Hospitalization	Char	*Y, N or Null	CRF	Record Qualifier	Did the event require or prolong hospitalization?	Perm	
AESLIFE	Is Life Threatening	Char	**Y, N or Null	CRF	Record Qualifier	Was the event life threatening?	Perm	
AESOD	Occurred with Overdose	Char	**Y, N or Null	CRF	Record Qualifier	Did the event occur with an overdose?	Perm	
AESMIE	Other Medically Important Serious Event	Char	**Y, N or Null	CRF	Record Qualifier	Do additional categories for seriousness apply?	Perm	
AECONTRT	Concomitant or Additional Trtmnt Given	Char	**Y, N	CRF	Record Qualifier	Was another treatment given because of the occurrence of the event?	Perm	
AETOXGR	Standard Toxicity Grade	Char	**	CRF or Derived	Record Qualifier	Toxicity grade according to a standard toxicity scale such as Common Terminology Criteria for Adverse Events v3.0 (CTCAE). Sponsor should specify scale and version used in the Comments column of the Define data definition document. Value should contain valid numbers only; datatype will be changed to numeric in future version of SDTM.	Perm	
AESTDTC	Start Date/Time of Adverse Event	Char	ISO 8601	CRF or Derived	Timing		Exp	SDTMIG 4.1.4.1 ; SDTMIG 4.1.4.2
AEENDTC	End Date/Time of Adverse Event	Char	ISO 8601	CRF or Derived	Timing		Exp	SDTMIG 4.1.4.1 ; SDTMIG 4.1.4.2

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
AESTDY	Study Day of Start of Adverse Event	Num		Derived	Timing	Study day of start of adverse event relative to the sponsor-defined RFSTDTC.	Perm	SDTMIG 4.1.4.4
AEENDY	Study Day of End of Adverse Event	Num		Derived	Timing	Study day of end of event relative to the sponsor-defined RFSTDTC.	Perm	SDTMIG 4.1.4.4
AEDUR	Duration of Adverse Event	Char	ISO 8601	CRF	Timing	Collected duration and unit of an adverse event. Used only if collected on the CRF and not derived from start and end date/times. Example: P1DT2H (for 1 day, 2 hours).	Perm	SDTMIG 4.1.4.3
AEENRF	End Relative to Reference Period	Char	** BEFORE, DURING, AFTER, DURING/ AFTER, U	Derived	Timing	Identifies the end of the event as being BEFORE, DURING, DURING/AFTER or AFTER the sponsor-defined reference period. The sponsor-defined reference period is a continuous period of time defined by a discrete starting point and a discrete ending point. Typically, this period is defined by the start (RFSTDTC) and end (RFENDTC) of the trial. Sponsors should define the reference period in the study metadata. Events that are ongoing at the end of the reference period should have a value of AFTER for this variable. If information such as "PRIOR", "ONGOING", or "CONTINUING" was collected, this information should be translated into AEENRF.	Perm	SDTMIG 4.1.4.7

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

6.2.1.1 Assumptions for Adverse Events (AE) domain model

1. AE Definition

- a. The Adverse Events dataset includes "any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have to have a causal relationship with this treatment" (ICH E2A). Adverse events may be captured either as free text or a pre-specified list of terms.

2. Adverse Event Description and Coding

- a. AETERM captures the verbatim term collected for the event. It is the topic variable for the AE dataset. AETERM is a required variable and must have a value.
- b. AEMODIFY is a permissible variable and should be included if the sponsor's procedure permits modification of a verbatim term for coding. The modified term is listed in AEMODIFY. The variable should be populated as per the sponsor's procedures; null values are permitted.

- c. AEDECOD is the preferred term derived by the sponsor from the coding dictionary. It is a required variable and must have a value. It is expected that the reported term (AETERM) will be coded using a standard dictionary such as MedDRA. The sponsor is expected to provide the dictionary name and version used to map the terms in the metadata (using the Comments column in the Define document).
 - d. AEBODSYS is the primary system organ class derived by the sponsor from the coding dictionary. It is expected that this variable will be populated.
3. Additional Categorization and Grouping
- a. AECAT and AESCAT should not be redundant with the domain code or dictionary classification provided by AEDECOD and AEBODSYS. That is, they should provide a different means of defining or classifying AE records. For example, a sponsor may be interested in identifying all adverse events that the investigator feels might represent an infection and thus will collect that categorization on the CRF. This categorization might differ from the categorization derived from the coding dictionary.
 - b. AEGRPID may be used to link (or associate) different records together to form a block of related records at the subject level within AE. It should not be used in place of AECAT or AESCAT, which are used to group data across subjects..
4. Presence or Absence of Adverse Events
- a. Adverse events are generally collected in two different ways, either by recording free text or using a pre-specified list of terms. In the latter case, the solicitation of information on specific adverse events may affect the frequency at which they are reported; therefore, the fact that a specific adverse event was solicited may be of interest to reviewers. Consequently, the SDS Team is proposing the addition of a new variable -- PRESP to the SDTM to indicate that the use of those specific medications was solicited. Until then, AEPRESP can be used as a supplemental qualifier.
 - b. If specific adverse events or a pre-specified list of terms are collected, AEOCCUR may be used to indicate whether or not each event occurred. Use of this variable is under evaluation by the SDS Team.
 - c. If a study collects both pre-specified events as well as subject reported events, the value of AEOCCUR should be populated for all pre-specified events and null for subject reported events.
 - a. When adverse events are collected with the recording of free text, some sponsors may enter a record into the data management system to indicate 'no adverse events,' for a specific subject. For these subjects, do not include a record to indicate that there were no events unless AEOCCUR is being used to represent records for non-occurring events. Otherwise, records should be included in the submission dataset only for adverse events that have actually occurred.
5. Timing Variables
- a. Either AEENDTC or AEENRF should be populated.
 - b. If adverse events are collected more than once within the CRF (indicating that the data are visit-based), then VISITNUM is expected. If the adverse events are collected only once within the CRF (indicating that the data are event-based), then VISITNUM is permissible. VISITDY and VISIT are always permissible.

6. Other Qualifier Variables

- a. If categories of serious events are collected secondarily to a leading question as in the example below, the values may be null. For example, if Serious is answered 'No,' the values for AESDTH, AESLIFE, AESHOSP, etc. may be null. However, if Serious is answered 'Yes,' at least one of AESDTH, AESLIFE, AESHOSP, etc. will have a 'Y' response.

Serious? Yes No

If yes, check all that apply: Fatal Life-threatening Inpatient hospitalization... etc.

On the other hand, if the CRF were structured so that a response is collected for each serious category, all category variables (e.g., AESDTH, AESHOSP) would be populated and AESER could be derived.

- b. When a description of Other Medically Important Serious Adverse Events category is collected on a CRF, sponsors should place the description in the SUPQUAL dataset as [described in Section 8.4](#).
- c. The serious categories “Involves cancer” (AESCAN) and “Occurred with overdose” (AESOD) are not part of the ICH definition of a serious adverse event, but these categories are available for use for studies conducted under guidelines that existed prior to the FDA’s adoption of the ICH definition.
- d. In studies using toxicity grade according to a standard toxicity scale such as Common Terminology Criteria for Adverse Events v3.0 (CTCAE - The Common Terminology is the current version for toxicity criteria published by the NCI (National Cancer Institute) at <http://ctep.cancer.gov/reporting/ctc.html>), AETOXGR should be used instead of AESEV. AESEV would not be included on the dataset. The sponsor is expected to specify which scale and version are used in the metadata (using the Comments column in the Define document).

7. AE Structure.

- a. The structure of the AE domain is one record per adverse event per subject. It is the sponsor's responsibility to define an event. This definition may vary based on the sponsor's requirements for characterizing and reporting product safety and is usually described in the protocol. The submission dataset structure may differ from the structure at the time of collection. One common approach is to collapse all records for an event to a summary level with the highest level of severity, causality, seriousness, etc. If there is a need to evaluate AEs at a more granular level, sponsors may choose to create a new record when severity, causality, or seriousness changes or worsens. Other approaches, including a visit-based record, may also be reasonable as long as they meet the sponsor's safety evaluation requirements.

6.2.2 Disposition — DS

ds.xpt, Disposition — Events Version 3.1.1, August 26, 2005. One record per disposition status or protocol milestone per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**DS	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
DSSEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
DSGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to tie together a block of related records in a single domain to support relationships within the domain and between domains.	Perm	SDTMIG 2.1 ; SDTM 2.2.4
DSREFID	Reference ID	Char		Sponsor Defined	Identifier	Optional internal or external identifier.	Perm	SDTM 2.2.4
DSSPID	Sponsor-Defined Identifier	Char		Sponsor Defined	Identifier	Optional Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Line number on a Disposition page.	Perm	SDTM 2.2.4
DSTERM	Reported Term for the Disposition Event	Char		CRF	Topic	Verbatim name of the event or protocol milestone. Some terms in DSTERM will match DSDECOD, but others, such as 'Subject moved,' will map to controlled terminology in DSDECOD, such as 'LOST TO FOLLOW-UP.'	Req	SDTMIG 4.1.3.6

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
DSDECOD	Standardized Disposition Term	Char	*	CRF or Derived	Synonym Qualifier	Controlled terminology for the name of disposition event or protocol milestone. Examples of disposition events: COMPLETED, ADVERSE EVENT, DEATH, LACK OF EFFICACY, LOST TO FOLLOW-UP, NON-COMPLIANCE WITH STUDY DRUG, PHYSICIAN DECISION, PREGNANCY, PROGRESSIVE DISEASE, PROTOCOL VIOLATION, SCREEN FAILURE, STUDY TERMINATED BY SPONSOR, TECHNICAL PROBLEMS, WITHDRAWAL OF CONSENT, OTHER. Examples of protocol milestones: INFORMED CONSENT OBTAINED, RANDOMIZED	Req	SDTMIG 4.1.3.5
DSCAT	Category for Disposition Event	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records. Examples: DISPOSITION EVENT or PROTOCOL MILESTONE.	Perm	SDTMIG 2.1
DSSCAT	Subcategory for Disposition Event	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of disposition event.	Perm	SDTMIG 2.1
VISIT	Visit Name	Char		CRF or Derived	Timing	1. Protocol-defined description of clinical encounter. 2. May be used in addition to VISITNUM and/or VISITDY.	Perm	SDTMIG 7.8; SDTMIG 4.1.4.5
VISITNUM	Visit Number	Num		CRF or Derived	Timing	1. Clinical encounter number. 2. Numeric version of VISIT, used for sorting.	Perm	SDTMIG 7.8; SDTMIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDTMIG 7.8; SDTMIG 4.1.4.5
EPOCH	Trial Epoch	Char	*	CRF or Derived	Timing	EPOCH may be used when DSCAT = 'DISPOSITION EVENT'. Examples: TREATMENT PHASE, SCREENING, FOLLOW-UP	Perm	SDTMIG 7.4.4
DSDTC	Date/Time of Collection	Char	ISO 8601	CRF or Derived	Timing		Perm	SDTMIG 4.1.4.1
DSSTDTC	Start Date/Time of Disposition Event	Char	ISO 8601	CRF or Derived	Timing		Exp	SDTMIG 4.1.4.1
DSSTDY	Study Day of Start of Disposition Event	Num		Derived	Timing	Study day of start of the disposition event relative to the sponsor-defined RFSTDTC.	Perm	SDTMIG 4.1.4.4

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

6.2.2.1 Assumptions for Disposition (DS) domain model

1. DS Definition

- a. The Disposition dataset provides an accounting for all subjects who entered the study and may include protocol milestones, such as randomization, as well as the subject's completion status or reason for discontinuation for each phase or segment of the study, including screening and post treatment follow-up. Sponsors may choose which disposition events and milestones to submit for a study.

2. DS Description and Coding
 - a. DSTERM contains the disposition or other event information such as a protocol milestone. It is a required variable and must have a value. DSTERM contains either specific verbatim information of the disposition event and/or sponsor defined controlled terms with the same values as DSDECOD. Some terms in DSTERM will match those in DSDECOD, but others such as 'Subject moved' will map to controlled terminology in DSDECOD such as 'LOST TO FOLLOW-UP'.
 - b. DSDECOD is the sponsor defined controlled term for the disposition event or other event information such as protocol milestones. It is a required variable and will always have a value. The value in DSDECOD may be equivalent to the value in DSTERM. For disposition events, DSDECOD = 'COMPLETE' when the subject completed the phase or segment described in EPOCH; otherwise, DSDECOD indicates the reason the subject did not complete the EPOCH.

3. Other Qualifier Variables
 - a. DSCAT may be used to distinguish between disposition events and protocol milestones. It is a permissible variable, but its use is recommended for clarity when both types of events are submitted.
 - b. Additional information about a death such as autopsy data could be modeled in a separate 'autopsy' domain based on the Findings class.

4. Timing Variables
 - a. DSSTDTC is expected as the date/time of the disposition event. Disposition events do not have start and end dates since disposition events do not span an interval (e.g. randomization date) but occur on a single date/time.
 - b. EPOCH may be used to indicate when the disposition event occurred. EPOCH is populated when DSCAT= 'DISPOSITION EVENT'. EPOCH should not be populated when DSCAT = 'PROTOCOL MILESTONE'.

6.2.3 Medical History — MH

mh.xpt, Medical History — Events, Version 3.1.1, August 26, 2005. One record per medical history event per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**MH	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
MHSEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
MHGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to tie together a block of related records in a single domain to support relationships within the domain and between domains.	Perm	SDTMIG 2.1 ; SDTM 2.2.4
MHREFID	Reference ID	Char		Sponsor Defined	Identifier	Optional internal or external medical history identifier.	Perm	SDTM 2.2.4
MHSPID	Sponsor-Defined Identifier	Char		Sponsor Defined	Identifier	Optional Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Line number on a Medical History page.	Perm	SDTM 2.2.4
MHTERM	Reported Term for the Medical History	Char		CRF	Topic	Verbatim or preprinted CRF term for the medical condition or event.	Req	SDTMIG 4.1.3.6
MHMODIFY	Modified Reported Term	Char		Sponsor Defined	Synonym Qualifier	If MHTERM is modified, then MHMODIFY will contain the modified text.	Perm	SDTMIG 4.1.3.5
MHDECOD	Dictionary-Derived Term	Char	**	Derived	Synonym Qualifier	Dictionary-derived text description of MHTERM or MHMODIFY. Equivalent to the Preferred Term (PT in MedDRA). The sponsor should specify the dictionary name and version in the Sponsor Comments column of the Define data definition document.	Perm	SDTMIG 4.1.3.5
MHCAT	Category for Medical History	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records. Examples: CARDIAC or GENERAL	Perm	SDTMIG 2.1

CDISC SDTM Implementation Guide (Version 3.1.1)

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
MHSCAT	Subcategory for Medical History	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of the condition or event.	Perm	SDTMIG 2.1
MHOCCUR	Medical History Occurrence	Char	**Y, N or Null	CRF or Sponsor Defined	Record Qualifier	Used when the occurrence of specific medical history events is solicited to indicate whether or not a medical condition had ever occurred. Values are null for spontaneously reported events. <i>Note: Use of this variable is under evaluation by the SDS Team.</i>	Perm	
MHSTAT	Medical History Status	Char	**NOT DONE	CRF or Sponsor Defined	Record Qualifier	The status indicates that the question was not asked.	Perm	SDTMIG 4.1.5.1
MHREASND	Reason Medical History Not Collected	Char		CRF	Record Qualifier	Describes the reason medical history was not collected. Used in conjunction with MHSTAT when value is NOT DONE.	Perm	SDTMIG 4.1.5.3
MHBODSYS	Body System or Organ Class	Char	**	Derived	Record Qualifier	Body system or organ class (Primary SOC) that is involved in an event or measurement from the standard hierarchy (e.g., MedDRA).	Perm	SDTMIG 4.1.3.5
VISIT	Visit Name	Char		CRF or Derived	Timing	1. Protocol-defined description of clinical encounter. 2. May be used in addition to VISITNUM and/or VISITDY.	Perm	SDTMIG 7.8 ; SDTMIG 4.1.4.5
VISITNUM	Visit Number	Num		CRF or Derived	Timing	1. Clinical encounter number. 2. Numeric version of VISIT, used for sorting.	Perm	SDTMIG 7.8 ; SDTMIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDTMIG 7.8 ; SDTMIG 4.1.4.5
MHDTC	Date/Time of History Collection	Char	ISO 8601	CRF or Derived	Timing		Perm	SDTMIG 4.1.4.1
MHSTDTC	Start Date/Time of Medical History Event	Char	ISO 8601	CRF or Derived	Timing		Perm	
MHENDTC	End Date/Time of Medical History Event	Char	ISO 8601	CRF or Derived	Timing		Perm	SDTMIG 4.1.4.1
MHDY	Study Day of History Collection	Num		Derived	Timing	1. Study day of medical history collection, measured as integer days. 2. Algorithm for calculations must be relative to the sponsor-defined RFSTDTC variable in Demographics. This formula should be consistent across the submission.	Perm	SDTMIG 4.1.4.4

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
MHENRF	End Relative to Reference Period	Char	** BEFORE, DURING, AFTER, DURING /AFTER, U	Derived	Timing	Identifies the end of the event as being BEFORE, DURING, DURING/AFTER or AFTER the sponsor-defined reference period. Sponsors should define the reference period in the study metadata. Events that are ongoing at the end of the reference period should have a value of AFTER for this variable. If information such as "PRIOR", "ONGOING", or "CONTINUING" was collected, this information should be translated into MHENRF.	Perm	SDTMIG 4.1.4.7

** indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.*

6.2.3.1 Assumptions for Medical History (MH) domain model

1. MH Definition

- a. The Medical History dataset generally includes the subject's prior history at the start of the trial. Note that treatments for prior and concomitant conditions should usually be submitted in an appropriate dataset from the Interventions class such as CM or SG.
- b. Examples of subject medical history information could include general medical history, gynecological history, and primary diagnosis.

2. Medical History Description and Coding

- a. MHTERM captures the verbatim term collected for the condition or event. It is the topic variable for the MH dataset. MHTERM is a required variable and must have a value.
- b. MHMODIFY is a permissible variable and should be included if the sponsor's procedure permits modification of a verbatim term for coding. The modified term is listed in MHMODIFY. The variable should be populated as per the sponsor's procedures; null values are permitted.
- c. If the sponsor codes the reported term (MHTERM) using a standard dictionary, then MHDECOD will be populated with the preferred term derived from the dictionary. The sponsor is expected to provide the dictionary name and version used to map the terms in the metadata (using the Comments column in the Define document).
- d. If the sponsor codes the reported term (MHTERM) using a standard dictionary, then MHBODSYS will be populated with the primary system organ class derived from the dictionary.
- e. If a CRF collects medical history by pre-specified body systems and the sponsor also codes reported terms using a standard dictionary, then MHDECOD and MHBODSYS are populated using the information from the standard dictionary. MHCAT and MHSCAT should be used for the pre-specified body systems.

3. Additional Categorization and Grouping

- a. MHCAT and MHSCAT may be populated with the sponsor's categorization of medical history events, which are often displayed on the CRF. MHCAT and MHSCAT should not be redundant with the domain code or dictionary classification provided by MHDECOD and/or MHBODSYS. That is, they should provide a different means of defining or classifying MH records.
 - i. This categorization should not group all records (within the MH Domain) into one generic group such as 'Medical History' or 'General Medical History' because this is redundant information with the domain code. If no smaller categorization can be applied, then it is not necessary to include or populate this variable.
 - ii. Examples of MHCAT could include 'General Medical History,' 'Allergy Medical History,' and 'Reproductive Medical History.'
- b. MHGRPID may be used to link (or associate) different records together to form a block of related records at the subject level within MH. It should not be used in place of MHCAT or MHSCAT, which are used to group data across subjects. For example, if a group of syndromes were related to a particular disease then this variable could be populated with the appropriate text.

4. Presence or Absence of History

- a. Information on medical history is generally collected in two different ways, either by recording free text or using a pre-specified list of terms. In the latter case, the solicitation of information on specific medical history events may affect the frequency at which they are reported; therefore, the fact that a specific medical history event was solicited may be of interest to reviewers. Consequently, the SDS Team is proposing the addition of a new variable --PRESP to the SDTM to indicate that the occurrence of those specific medical history events was solicited. Until then, MHPRESP can be used as a supplemental qualifier.
- b. If history is collected by body system or category, MHOCCUR may be used to indicate the presence or absence of events for a given system or category.
- c. If specific medical history events are solicited (i.e., a pre-specified list of terms is collected), MHOCCUR may be used to indicate whether or not each event occurred. Use of this variable is under evaluation by the SDS Team.
- d. When the response for a medical history question or group of questions on a CRF was not collected, then the value of MHSTAT = 'NOT DONE' and MHOCCUR will be null. The reason that medical history was not done may be documented in MHREASND.
- e. If a study collects both pre-specified events as well as subject reported events, the value of MHOCCUR should be populated for all pre-specified events and null for subject-reported events unless the response was not collected.

6.3 FINDINGS

6.3.1 ECG Test Results — EG

eg.xpt, ECG — Findings, Version 3.1.1, August 26, 2005. One record per ECG observation per time point per visit per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**EG	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
EGSEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
EGGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to tie together a block of related records in a single domain to support relationships within the domain and between domains.	Perm	SDTMIG 2.1 ; SDTM 2.2.4
EGREFID	ECG Reference ID	Char		Sponsor Defined or Derived	Identifier	Internal or external ECG identifier. Example: UUID for external ECG Waveform File.	Perm	
EGSPID	Sponsor-Defined Identifier	Char		Sponsor Defined	Identifier	Optional Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Line number from the ECG page.	Perm	SDTM 2.2.4
EGTESTCD	ECG Test or Examination Short Name	Char	**	CRF or Derived	Topic	Short name of the measurement, test, or examination described in EGTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in EGTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g., '1TEST'). EGTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: PR, QT, INTP.	Req	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
EGTEST	ECG Test or Examination Name	Char	**	CRF	Synonym Qualifier	Verbatim name of the test or examination used to obtain the measurement or finding. The value in EGTEST cannot be longer than 40 characters. Examples: PR Interval, QT Interval, etc.	Req	
EGCAT	Category for ECG	Char	*	Sponsor Defined	Grouping Qualifier	Used to categorize ECG observations. Examples: MEASUREMENT or FINDING.	Perm	SDTMIG 2.1
EGSCAT	Subcategory for ECG	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of the ECG.	Perm	SDTMIG 2.1
EGPOS	ECG Position of Subject	Char	*	CRF or Derived	Record Qualifier	Position of the subject during a measurement or examination. Examples: SUPINE, STANDING: SITTING.	Perm	
EGORRES	Result or Finding in Original Units	Char		CRF or Derived	Result Qualifier	Result of the ECG measurement or finding as originally received or collected. Examples of expected values are 62 or 0.151 when EGTEST is a measured finding, or ATRIAL FIBRILLATION or QT PROLONGATION when EGTEST is 'interpretation'.	Exp	SDTMIG 4.1.5.1
EGORRESU	Original Units	Char	*	CRF or Derived	Variable Qualifier	Original units in which the data were collected. The unit for EGORRES. Examples: SECONDS or MILLISECONDS.	Perm	SDTMIG 4.1.5.1
EGNRIND	Reference Range Indicator	Char	*	CRF or Derived	Variable Qualifier	1. Indicates where value falls with respect to reference range which could be defined by EGORNRL0 and EGORNRI (if those variables are used) or by EGSTNRC. Examples: NORMAL, ABNORMAL, HIGH, LOW. 2. Should not be used to indicate clinical significance.	Perm	
EGSTRESC	Character Result/Finding in Std Format	Char		Derived	Result Qualifier	Contains the result value for all findings, copied or derived from EGORRES in a standard format or standard units. EGSTRESC should store all results or findings in character format; if results are numeric, they should also be stored in numeric format in EGSTRESN. For example, if a test has results 'NONE', 'NEG', and 'NEGATIVE' in EGORRES and these results effectively have the same meaning, they could be represented in standard format in EGSTRESC as 'NEGATIVE'. For other examples, see general assumptions . Additional examples of result data; SINUS BRADYCARDIA, ATRIAL FLUTTER, ATRIAL FIBRILLATION.	Exp	SDTMIG 4.1.5.1
EGSTRESN	Numeric Result/Finding in Standard Units	Num		Derived	Result Qualifier	Used for continuous or numeric results or findings in standard format; copied in numeric format from EGSTRESC. EGSTRESN should store all numeric test results or findings.	Perm	SDTMIG 4.1.5.1
EGSTRESU	Standard Units	Char	*	CRF or Derived	Variable Qualifier	Standardized unit used for EGSTRESC or EGSTRESN.	Perm	SDTMIG 4.1.5.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
EGSTAT	ECG Status	Char	**NOT DONE	CRF or Derived	Record Qualifier	Used to indicate an ECG was not done, or an ECG measurement was not taken. Should be null if a result exists in EGORRES.	Perm	SDTMIG 4.1.5.1
EGREASND	Reason ECG Not Performed	Char		CRF or Derived	Record Qualifier	Describes why a measurement or test was not performed. Examples: BROKEN EQUIPMENT or SUBJECT REFUSED. Used in conjunction with EGSTAT when value is NOT DONE.	Perm	
EGXFN	ECG External file Name	Char		Sponsor Defined or Derived	Record Qualifier	File name for the external ECG Waveform file	Perm	
EGNAM	Vendor Name	Char		CRF or Derived	Record Qualifier	Name or identifier of the laboratory or vendor who provides the test results.	Perm	
EGLOINC	LOINC Code	Char	**	Derived	Synonym Qualifier	1. LOINC Code for EGTEST. LOINC codes are available for many but not all commonly used ECG test observations. 2. The LOINC version should be provided in the Sponsor Comments column in the data definition document.	Perm	SDTMIG 4.1.3.2
EGMETHOD	Method of ECG Test	Char	*	CRF	Record Qualifier	Method of the ECG test. Examples: MONITORING (1-LEAD), 12-LEAD.	Perm	
EGBLFL	Baseline Flag	Char	**Y or Null	CRF or Derived	Record Qualifier	Indicator used to identify a baseline value.	Exp	SDTMIG 4.1.3.7
EGDRVFL	Derived Flag	Char	**Y or Null	Derived	Record Qualifier	Used to indicate a derived record. The value should be Y or null. Records which represent the average of other records or questionnaire subscores, which do not come from the CRF are examples of records which would be derived for the submission datasets. If EGDRVFL=Y, then EGORRES should be NULL, with EGSTRESC, and (if numeric) EGSTRESN having the derived value.	Perm	SDTMIG 4.1.3.7
EGEVAL	Evaluator	Char	*	CRF	Record Qualifier	Role of the person who provided the evaluation. Used only for results that are subjective (e.g., assigned by a person or a group). Should be null for records that contain collected or derived data. Examples: INVESTIGATOR, ADJUDICATION COMMITTEE, VENDOR.	Exp	
VISIT	Visit Name	Char		CRF or Derived	Timing	1. Protocol-defined description of clinical encounter. 2. May be used in addition to VISITNUM and/or VISITDY.	Perm	SDTMIG 7.8; SDTMIG 4.1.4.5

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
VISITNUM	Visit Number	Num		CRF or Derived	Timing	1. Clinical encounter number. 2. Numeric version of VISIT, used for sorting.	Req	SDTMIG 7.8; 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDTMIG 7.8; SDTMIG 4.1.4.5
EGDTC	Date/Time of ECG	Char	ISO 8601	CRF or Derived	Timing		Exp	SDTMIG 4.1.4.1
EGDY	Study Day of ECG	Num		Derived	Timing	1. Study day of the ECG, measured as integer days. 2. Algorithm for calculations must be relative to the sponsor-defined RFSTDTC variable in Demographics. This formula should be consistent across the submission.	Perm	SDTMIG 4.1.4.4 SDTM 2.2.6
EGTPTNUM	Planned Time Point Number	Num		CRF or Derived	Timing	Numerical version of EGTPT to aid in sorting.	Perm	
EGTPT	Planned Time Point Name	Char		CRF or Derived	Timing	1. Text Description of time when measurement should be taken. 2. This may be represented as an elapsed time relative to a fixed reference point, such as time of last dose. See EGTPTNUM and EGTPTREF. Examples: Start, 5 min post.	Perm	
EGELTM	Elapsed Time from Reference Point	Char	ISO 8601	Derived	Timing	Elapsed time (in ISO 8601) relative to a planned fixed reference (EGTPTREF). This variable is useful where there are repetitive measures. Not a clock time or a date time variable. Examples: '-P15M' to represent the period of 15 minutes prior to the reference point indicated by EGTPTREF, or 'P8H' to represent the period of 8 hours after the reference point indicated by EGTPTREF.	Perm	SDTMIG 4.1.4.3
EGTPTREF	Time Point Reference	Char		Sponsor Defined	Timing	Name of the fixed reference point referred to by EGELTM, EGTPTNUM, and EGTPT. Examples: Previous Dose, Previous Meal.	Perm	

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

6.3.1.1 Assumptions for ECG (EG) domain model

1. The method for QT interval correction is included in the test name by controlled terminology. The controlled terminology would be QTCF for Fridericia and QTCB for Bazett in EGTESTCD. EGTEST would contain either 'QT interval correction Fridericia' or 'QT interval correction Bazett'. The value of the calculated QT interval results goes into EGORRES.
2. Although EGLOINC is included to allow use of LOINC codes for ECG tests where appropriate, other coding schemes may be recommended by the CDISC terminology team in the future.

6.3.2 Inclusion/Exclusion Exceptions — IE

ie.xpt, Inclusion/Exclusion Exceptions — Findings, Version 3.1.1, August 26, 2005. One record per inclusion/exclusion criteria exception per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**IE	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
IESEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
IESPID	Sponsor-Defined Identifier	Char		Sponsor Defined	Identifier	Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Question number on a questionnaire.	Perm	SDTM 2.2.4
IETESTCD	Inclusion/Exclusion Criterion Short Name	Char	*	CRF	Topic	Short name of the criterion described in IETEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in IETESTCD cannot be longer than 8 characters, nor can it start with a number (e.g. '1TEST'). IETESTCD cannot contain characters other than letters, numbers, or underscores. Examples: IN01, EX01.	Req	
IETEST	Inclusion/Exclusion Criterion	Char		CRF	Synonym Qualifier	Verbatim description of the inclusion or exclusion criterion that was the exception for the subject within the study.	Req	
IECAT	Inclusion//Exclusion Category	Char	**INCLUSION, EXCLUSION	Sponsor Defined	Grouping Qualifier	Used to define a category of related records.	Req	SDTMIG 2.1
IESCAT	Inclusion//Exclusion Subcategory	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of the exception criterion. Can be used to distinguish criteria for a sub-study.	Perm	SDTMIG 2.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
IEORRES	I/E Criterion Original Result	Char	**Y, N	CRF or Derived	Result Qualifier	Original response to Inclusion/Exclusion Criterion question. Inclusion or Exclusion criterion met?	Req	SDTMIG 4.1.5.1
IESTRESC	I/E Criterion Result in Std Format	Char	**Y, N	Derived	Result Qualifier	Response to Inclusion/Exclusion criterion result in standard format. (Same as IEORES for the IE Domain).	Req	SDTMIG 4.1.5.1
VISIT	Visit Name	Char		CRF or Derived	Timing	1. Protocol-defined description of clinical encounter. 2. May be used in addition to VISITNUM and/or VISITDY.	Perm	SDTMIG 7.8; SDTMIG 4.1.4.5
VISITNUM	Visit Number	Num		CRF or Derived	Timing	1. Clinical encounter number. 2. Numeric version of VISIT, used for sorting.	Perm	SDTMIG 7.8; SDTMIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDTMIG 7.8; SDTMIG 4.1.4.5
IEDTC	Date/Time of Collection	Char	ISO 8601	CRF or Derived	Timing		Perm	SDTMIG 4.1.4.1
IEDY	Study Day of Collection	Num		Derived	Timing	1. Study day of collection of the inclusion/exclusion exceptions, measured as integer days. 2. Algorithm for calculations must be relative to the sponsor-defined RFSTDTC variable in Demographics. This formula should be consistent across the submission.	Perm	SDTMIG 4.1.4.4 SDTM 2.2.5

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

6.3.2.1 Assumptions for Inclusion/Exclusion Exceptions (IE) domain model

1. The intent of the domain model is to only collect those criteria that cause the subject to be in violation of the Inclusion/Exclusion criteria not a response to each criterion.
2. This domain should not be used to collect protocol deviations/violations incurred during the course of the study. A separate domain model (DV, available separately) will be developed for protocol deviations/violations in a future version of the Implementation Guide.
3. Note that the SQL statement query "select distinct IETESTCD from IE;" may not contain an exhaustive list of the inclusion / exclusion criteria because this domain only lists those not met by subjects. The complete list would be found in the TI trial inclusion/exclusion criteria dataset described in [Section 7.9](#).

6.3.3 Laboratory Test Results — LB

lb.xpt, Labs — Findings, Version 3.1.1, August 26, 2005. One record per lab test per time point per visit per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**LB	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
LBSEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
LBGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to tie together a block of related records in a single domain to support relationships within the domain and between domains.	Perm	SDTMIG 2.1; SDTM 2.2.4
LBREFID	Specimen ID	Char		Sponsor Defined or Derived	Identifier	Internal or external specimen identifier. Example: Specimen ID.	Perm	
LBSPID	Sponsor-Defined Identifier	Char		Sponsor Defined	Identifier	Optional Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Line number on the Lab page.	Perm	SDTM 2.2.4
LBTESTCD	LAB Test or Examination Short Name	Char	*	CRF or Derived	Topic	Short name of the measurement, test, or examination described in LBTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in LBTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g. '1TEST'). LBTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: ALT, LDH.	Req	
LBTEST	LAB Test or Examination Name	Char	*	CRF	Synonym Qualifier	Verbatim name of the test or examination used to obtain the measurement or finding. Note any test normally performed by a clinical laboratory is considered a lab test. The value in LBTEST cannot be longer than 40 characters. Examples: Alanine Aminotransferase, Lactate Dehydrogenase.	Req	
LBCAT	Category for Lab Test	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records. Examples: such as HEMATOLOGY, URINALYSIS, CHEMISTRY.	Exp	SDTMIG 2.1
LBSCAT	Subcategory for Lab Test	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of a test category such as DIFFERENTIAL, COAGULATON, LIVER FUNCTION or ELECTROLYTES.	Perm	SDTMIG 2.1.
LBORRES	Result or Finding in Original Units	Char		CRF or Derived	Result Qualifier	Result of the measurement or finding as originally received or collected.	Exp	SDTMIG 4.1.5.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
LBORRESU	Original Units	Char	*	CRF or Derived	Variable Qualifier	Original units in which the data were collected. The unit for LBORRES. Example: g/L.	Exp	SDTMIG 4.1.3.2
LBORNRLO	Reference Range Lower Limit in Orig Unit	Char		CRF or Derived	Variable Qualifier	Lower end of reference range for continuous measurements in original units. Should be populated only for continuous results.	Exp	
LBORNRHI	Reference Range Upper Limit in Orig Unit	Char		CRF or Derived	Variable Qualifier	Upper end of reference range for continuous measurements in original units. Should be populated only for continuous results.	Exp	
LBSTRESC	Character Result/Finding in Std Format	Char		Derived	Result Qualifier	Contains the result value for all findings, copied or derived from LBORRES in a standard format or standard units. LBSTRESC should store all results or findings in character format; if results are numeric, they should also be stored in numeric format in LBSTRESN. For example, if a test has results 'NONE', 'NEG', and 'NEGATIVE' in LBORRES and these results effectively have the same meaning, they could be represented in standard format in LBSTRESC as 'NEGATIVE'. For other examples, see general assumptions .	Exp	SDTMIG 4.1.5.1
LBSTNRC	Reference Range for Char Rslt-Std Units	Char		CRF or Derived	Variable Qualifier	For normal range values that are character in ordinal scale (e.g., 'NEGATIVE TO TRACE', '-1 to +1', etc.). If categorical ranges were supplied.	Perm	
LBSTRESN	Numeric Result/Finding in Standard Units	Num		Derived	Result Qualifier	Used for continuous or numeric results or findings in standard format; copied in numeric format from LBSTRESC. LBSTRESN should store all numeric test results or findings.	Exp	SDTMIG 4.1.5.1
LBSTRESU	Standard Units	Char	*	CRF or Derived	Variable Qualifier	Standardized unit used for LBSTRESC or LBSTRESN.	Exp	
LBSTNRLO	Reference Range Lower Limit-Std Units	Num		CRF or Derived	Variable Qualifier	Lower end of reference range for continuous measurements in standardized units. Should be populated only for continuous results.	Exp	
LBSTNRHI	Reference Range Upper Limit-Std Units	Num		CRF or Derived	Variable Qualifier	Upper end of reference range for continuous measurements in standardized units. Should be populated only for continuous results.	Exp	
LBNRIND	Reference Range Indicator	Char	*	CRF or Derived	Variable Qualifier	1. Indicates where value falls with respect to reference range defined by LBORNRLO and LBORNRHI or by LBSTNRC. Examples: NORMAL, ABNORMAL, HIGH, LOW. 2. Should not be used to indicate clinical significance.	Exp	
LBSTAT	Lab Status	Char	**NOT DONE	CRF or Derived	Record Qualifier	Used to indicate exam not done. Should be null if a result exists in LBORRES.	Perm	SDTMIG 4.1.5.1
LBREASND	Reason Test Not Done	Char		CRF or Derived	Record Qualifier	Describes why a measurement or test was not performed such as BROKEN EQUIPMENT, SUBJECT REFUSED, or SPECIMEN LOST. Used in conjunction with LBSTAT when value is NOT DONE.	Perm	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
LBNAM	Vendor Name	Char		CRF or Derived	Record Qualifier	Name or identifier of the laboratory or vendor who provides the test results.	Perm	
LBLOINC	LOINC Code	Char	**	Derived	Synonym Qualifier	1. LOINC Code for LBTEST. 2. The LOINC version should be provided in the comments column in the data definition file.	Perm	SDTMIG 4.1.3.2
LBSPEC	Specimen Type	Char	*	CRF	Record Qualifier	Defines the type of specimen used for a measurement. Examples: SERUM, PLASMA, URINE.	Perm	
LBSPCCND	Specimen Condition	Char		CRF	Record Qualifier	Free or standardized text describing the condition of the specimen e.g. HEMOLYZED, ICTERIC, LIPEMIC etc.	Perm	
LBMETHOD	Method of Test or Examination	Char	*	CRF	Record Qualifier	Method of the test or examination. Examples EIA (Enzyme Immunoassay), ELECTROPHORESIS, DIPSTICK	Perm	
LBBLFL	Baseline Flag	Char	**Y or Null	CRF or Derived	Record Qualifier	Indicator used to identify a baseline value.	Exp	SDTMIG 4.1.3.7
LBDRVFL	Derived Flag	Char	**Y or Null	CRF or Derived	Record Qualifier	Used to indicate a derived record. The value should be Y or null. Records which represent the average of other records or questionnaire subscores, which do not come from the CRF, are examples of records which would be derived for the submission datasets. If LBDRVFL=Y, then LBORRES should be NULL, with LBSTRESC, and (if numeric) LBSTRESN having the derived value.	Perm	SDTMIG 4.1.5.1 SDTMIG 4.1.3.7
LBFAST	Fasting Status	Char	**Y, N, U or Null	CRF	Record Qualifier	Indicator used to identify fasting status such as Y, N, U, or null if not relevant.	Perm	
LBTOX	Toxicity	Char	*	Derived	Variable Qualifier	Description of toxicity quantified by LBTOXGR. Sponsor should specify which scale and version is used in sponsor comments. Example: NCI CTCAE short name	Perm	
LBTOXGR	Standard Toxicity Grade	Char	**	CRF or Derived	Variable Qualifier	Records toxicity grade value using a standard toxicity scale (such as the NCI CTCAE). Value should contain valid numbers only; datatype will be changed to numeric in future version of SDTM.	Perm	
VISIT	Visit Name	Char		CRF or Derived	Timing	1. Protocol-defined description of clinical encounter 2. May be used in addition to VISITNUM and/or VISITDY	Perm	SDTMIG 7.8 ; SDTMIG 4.1.4.5
VISITNUM	Visit Number	Num		CRF or Derived	Timing	1. Clinical encounter number. 2. Numeric version of VISIT, used for sorting.	Req	SDTMIG 7.8 ; SDTMIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDTMIG 7.8 ; SDTMIG 4.1.4.5
LBDC	Date/Time of Specimen Collection	Char	ISO 8601	CRF or Derived	Timing		Exp	SDTMIG 4.1.4.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
LBENDTC	End Date/Time of Specimen Collection	Char	ISO 8601	CRF or Derived	Timing		Perm	SDTMIG 4.1.4.1 SDTMIG 4.1.4.8
LBDY	Study Day of Specimen Collection	Num		Derived	Timing	1. Study day of specimen collection, measured as integer days. 2. Algorithm for calculations must be relative to the sponsor-defined RFSTDTC variable in Demographics. This formula should be consistent across the submission.	Perm	SDTMIG 4.1.4.4
LBTPPT	Planned Time Point Name	Char		CRF or Derived	Timing	1. Text Description of time when specimen should be taken. 2. This may be represented as an elapsed time relative to a fixed reference point, such as time of last dose. See LBTPPTNUM and LBTPPTREF. Examples: Start, 5 min post.	Perm	
LBTPPTNUM	Planned Time Point Number	Num		CRF or Derived	Timing	Numerical version of LBTPPT to aid in sorting.	Perm	
LBELTM	Elapsed Time from Reference Point	Char	ISO 8601	Derived	Timing	Elapsed time (in ISO 8601) relative to a planned fixed reference (LBTPPTREF).. This variable is useful where there are repetitive measures. Not a clock time or a date time variable. Examples: '-P15M' to represent the period of 15 minutes prior to the reference point indicated by LBTPPTREF, or 'P8H' to represent the period of 8 hours after the reference point indicated by LBTPPTREF.	Perm	SDTMIG 4.1.4.3
LBTPPTREF	Time Point Reference	Char		Sponsor Defined	Timing	Name of the fixed reference point referred to by LBELTM, LBTPPTNUM, and LBTPPT. Examples: PREVIOUS DOSE, PREVIOUS MEAL.	Perm	

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

6.3.3.1 Assumptions for Laboratory Test Results (LB) domain model

1. For lab tests that do not have continuous numeric results (e.g., urine protein as measured by dipstick, descriptive tests such as urine color), LBSTNRC could be populated either with normal range values that are character in an ordinal scale (e.g., 'NEGATIVE to TRACE') or a delimited set of values that are considered to be normal (e.g., 'YELLOW', 'AMBER'). LBORNRL0, LBORNRI, LBSTNRLO, and LBSTNRHI should be null for these type of tests.
2. For lab tests where the specimen is collected over time, i.e., 24-hour urine collection, the start date/time of the collection goes into LBDTC and the end date/time of collection goes into LBENDTC.

6.3.4 Physical Examinations — PE

pe.xpt, Physical Examination — Findings, Version 3.1.1, August 26, 2005. One record per body system or abnormality per visit per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**PE	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
PESEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
PEGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to tie together a block of related records in a single domain to support relationships within the domain and between domains.	Perm	SDTMIG 2.1 ; SDTM 2.2.4
PESPID	Sponsor-Defined Identifier	Char		Sponsor Defined	Identifier	Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Question number on a questionnaire.	Perm	SDTM 2.2.4
PETESTCD	Body System Examined Short Name	Char	*	CRF or Derived	Topic	Topic variable for PE. Short name for the value in PETEST, which can be used as a column name when converting the dataset from a vertical format to a horizontal format. The short names can be up to 8 characters. Examples: CV, GI. For subject-level exam, value should be 'ALL'.	Req	
PETEST	Body System Examined	Char	*	CRF	Synonym Qualifier	Verbatim term part of the body examined. The value in PETEST cannot be longer than 40 characters. Examples: CARDIOVASCULAR and GASTROINTESTINAL. For subject-level exam, value should be 'ALL'.	Req	
PEMODIFY	Modified Reported Term	Char		Sponsor Defined	Synonym Qualifier	If PEORRES is modified as part of a defined procedure, then PEMODIFY will contain the modified text.	Perm	SDTMIG 4.1.3.6
PECAT	Category for Examination	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of examination. Examples: GENERAL, NEUROLOGICAL.	Perm	SDTMIG 2.1
PESCAT	Subcategory for Examination	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of the examination. Used if needed to add further detail to PECAT.	Perm	SDTMIG 2.1
PELOC	Location of Physical Exam Finding	Char	*	CRF or Sponsor Defined	Record Qualifier	Can be used to specify where a physical exam finding occurred. Example: LEFT ARM for skin rash.	Perm	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
PEBODSYS	Body System or Organ Class	Char	**	CRF or Derived	Result Qualifier	1. Body system or organ class (Primary SOC) that is involved in an event or measurement from the standard hierarchy (e.g., MedDRA). 2. May be predefined from CRF page or obtained from coding.	Perm	SDTMIG 4.1.3.5
PEORRES	Verbatim Examination Finding	Char		CRF or Derived	Result Qualifier	Text description of any abnormal findings. If the examination was completed and there were no abnormal findings, the value should be NORMAL. If the examination was not performed on a particular body system, or at the subject level, then the value should be null, and NOT DONE should appear in PESTAT.	Exp	SDTMIG 4.1.5.1
PEORRESU	Original Units	Char		CRF or Derived	Variable Qualifier	Original units in which the data were collected. The unit for PEORRES.	Perm	
PESTRESC	Character Result/Finding in Std. Format	Char		Derived	Result Qualifier	If there are findings for a body system, then either the dictionary preferred term (if findings are coded using a dictionary) or PEORRES (if findings are not coded) should appear here. If PEORRES is null, PESTRESC should be null	Exp	SDTMIG 4.1.5.1
PESTRESN	Numeric Result/Finding in Standard Units	Num		Derived	Result Qualifier	Used for continuous or numeric results or findings in standard format; copied in numeric format from PESTRESC. PESTRESN should store all numeric test results or findings.	Exp	SDTMIG 4.1.5.1
PESTRESU	Standard Units	Char	*	Derived	Variable Qualifier	Standardized unit used for PESTRESC or PESTRESN.	Exp	
PESEV	Severity/Intensity	Char	*	CRF	Record Qualifier	The severity or intensity of the exam finding. PESEV is only used to qualify the results of an observation -- it should not be used to record the answer to the question posed in PETEST.. Example: "MODERATE" could be used to describe a PEORRES of "ACNE".	Perm	
PESTAT	Examination Status	Char	**NOT DONE	CRF or Derived	Record Qualifier	Used to indicate exam not done. Should be null if a result exists in ORRES.	Perm	SDTMIG 4.1.5.1
PEREASND	Reason Not Examined	Char		CRF or Derived	Record Qualifier	Describes why an examination was not performed or why a body system was not examined. Example: SUBJECT REFUSED. Used in conjunction with STAT when value is NOT DONE.	Perm	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
PEBLFL	Baseline Flag	Char	**Y or Null	CRF or Derived	Record Qualifier	Indicator used to identify a baseline value.	Exp	SDTMIG 4.1.3.7
PEEVAL	Evaluator	Char	*	CRF	Record Qualifier	Role of the person who provided the evaluation. Used only for results that are subjective (e.g., assigned by a person or a group). Should be null for records that contain collected or derived data. Examples: INVESTIGATOR, ADJUDICATION COMMITTEE, VENDOR.	Perm	
VISIT	Visit Name	Char		CRF or Derived	Timing	1. Protocol-defined description of clinical encounter. 2. May be used in addition to VISITNUM and/or VISITDY.	Perm	SDTMIG 7.8; SDTMIG 4.1.4.5
VISITNUM	Visit Number	Num		CRF or Derived	Timing	1. Clinical encounter number. 2. Numeric version of VISIT, used for sorting.	Exp	SDTMIG 7.8; SDTMIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDTMIG 7.8; SDTMIG 4.1.4.5
PEDTC	Date/Time of Examination	Char	ISO 8601	CRF or Derived	Timing		Exp	SDTMIG 4.1.4.1
PEDY	Study Day of Examination	Num		Derived	Timing	1. Study day of physical exam, measured as integer days. 2. Algorithm for calculations must be relative to the sponsor-defined RFSTDTC variable in Demographics. This formula should be consistent across the submission.	Perm	SDTMIG 4.1.4.4

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

6.3.4.1 Assumptions for Physical Examinations (PE) domain model

1. The PE domain provides an example where the result, PEORRES, is coded. This is in contrast to other domains (e.g., AE, CM, and MH), in which the topic variable (AETERM, CMTRT, and EXTRT, respectively) is the one coded.

6.3.5 Questionnaires — QS

qs.xpt, Questionnaires — Findings, Version 3.1.1, August 26, 2005. One record per question per time point per visit per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**QS	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
QSSEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a domain.	Req	SDTM 2.2.4
QSGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to tie together a block of related records in a single domain to support relationships within the domain and between domains.	Perm	SDTMIG 2.1 ; SDTM 2.2.4
QSSPID	Sponsor-Defined Identifier	Char		Sponsor Defined	Identifier	Optional Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Question number on a questionnaire.	Perm	SDTM 2.2.4
QSTESTCD	Question Short Name	Char	*	CRF or Derived	Topic	Topic variable for QS. Short name for the value in QSTEST, which can be used as a column name when converting the dataset from a vertical format to a horizontal format. The value in QSTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g. '1TEST'). QSTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: COG01, GH1, PF1.	Req	
QSTEST	Question Name	Char		CRF or Derived	Synonym Qualifier	Verbatim name of the question or group of questions used to obtain the measurement or finding. The value in QSTEST cannot be longer than 40 characters. Examples: In General, Would You Say your Health Is.	Req	
QSCAT	Category of Question	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records that will be meaningful to the Reviewer. Examples: HAMILTON DEPRESSION SCLAE, SF36, ADAS.	Req	SDTMIG 2.1
QSSCAT	Subcategory for Question	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of the questions within the category. Examples: MENTAL HEALTH DOMAIN, DEPRESSION DOMAIN, WORD RECALL.	Perm	SDTMIG 2.1
QSORRES	Finding in Original Units	Char		CRF or Derived	Result Qualifier	Finding as originally received or collected (e.g. RARELY, SOMETIMES).	Exp	SDTMIG 4.1.5.1
QSORRESU	Original Units	Char	*	CRF or Derived	Variable Qualifier	Original units in which the data were collected. The unit for QSORRES, such as minutes or seconds or the units associated with a visual analog scale.	Perm	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
QSSTRESC	Character Result/Finding in Std Format	Char		CRF or Derived	Result Qualifier	Contains the finding for all questions or sub-scores, copied or derived from QSORRES in a standard format or standard units. QSSTRESC should store all findings in character format; if findings are numeric, they should also be stored in numeric format in QSSTRESN. If question scores are derived from the original finding, then the standard format is the score. Examples: 0, 1.	Exp	SDTMIG 4.1.5.1
QSSTRESN	Numeric Finding in Standard Units	Num		Derived	Result Qualifier	Used for continuous or numeric findings in standard format; copied in numeric format from QSSTRESC. QSSTRESN should store all numeric results or findings.	Perm	SDTMIG 4.1.5.1
QSSTRESU	Standard Units	Char	*	CRF or Derived	Variable Qualifier	Standardized unit used for QSSTRESC or QSSTRESN.	Perm	
QSSTAT	Status of Question	Char	**NOT DONE	CRF or Derived	Record Qualifier	Used to indicate a questionnaire or response to a questionnaire was not done. Should be null if a result exists in QSORRES.	Perm	SDTMIG 4.1.5.1
QSREASND	Reason Not Performed	Char		CRF or Derived	Record Qualifier	Describes why a question was not answered. Example: subject refused. Used in conjunction with QSSTAT when value is NOT DONE.	Perm	
QSBFL	Baseline Flag	Char	**Y or Null	CRF or Derived	Record Qualifier	Indicator used to identify a baseline value.	Exp	SDTMIG 4.1.3.7
QSDRVFL	Derived Flag	Char	**Y or Null	CRF or Derived	Record Qualifier	The value should be Y or null. Records which represent the average of other records or questionnaire sub-scores which do not come from the CRF are examples of records which would be derived for the submission datasets. If QSDRVFL=Y, then QSORRES should be NULL, with QSSTRESC and (if numeric) QSSTRESN having the derived value.	Perm	SDTMIG 4.1.5.1 SDTMIG 4.1.3.7
VISIT	Visit Name	Char		CRF or Derived	Timing	1. Protocol-defined description of clinical encounter. 2. May be used in addition to VISITNUM and/or VISITDY.	Perm	SDTMIG 7.8; SDTMIG 4.1.4.5
VISITNUM	Visit Number	Num		CRF or Derived	Timing	1. Clinical encounter number. 2. Numeric version of VISIT, used for sorting.	Exp	SDTMIG 7.8; SDTMIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDTMIG 7.8; SDTMIG 4.1.5
QSDTC	Date/Time of Finding	Char	ISO 8601	CRF or Derived	Timing		Exp	SDTMIG 4.1.4.
QSDY	Study Day of Finding	Num		Derived	Timing	1. Study day of finding collection, measured as integer days. 2. Algorithm for calculations must be relative to the sponsor-defined RFSTDTC variable in Demographics. This formula should be consistent across the submission.	Perm	SDTMIG 4.1.4.4
QSTPT	Planned Time Point Name	Char		CRF or Derived	Timing	1. Text Description of time when questionnaire is administered. 2. This may be represented as an elapsed time relative to a fixed reference point, such as time of last dose. See QSTPTNUM and QSTPTREF.	Perm	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
QSTPTNUM	Planned Time Point Number	Num		CRF or Derived	Timing	Numerical version of QSTPT to aid in sorting.	Perm	
QSELTM	Elapsed Time from Reference Point	Char	ISO 8601	Derived	Timing	Elapsed time (in ISO 8601) relative to a planned fixed reference (QSTPTREF). This variable is useful where there are repetitive measures. Not a clock time or a date time variable. Examples: '-P15M' to represent the period of 15 minutes prior to the reference point indicated by QSTPTREF, or 'P8H' to represent the period of 8 hours after the reference point indicated by QSTPTREF.	Perm	SDTMIG 4.1.4.3
QSTPTREF	Time Point Reference	Char		Sponsor Defined	Timing	Name of the fixed reference point referred to by QSELTM, QSTPTNUM, and QSTPT. Examples: PREVIOUS DOSE, PREVIOUS MEAL.	Perm	
QSEVLINT	Evaluation Interval	Char	ISO 8601	CRF or Derived	Timing	Evaluation Interval associated with a QSTEST question represented in ISO 8601 character format. Example: "P2Y" to represent an interval of 2 years in the question "Have you experienced any episodes in the past 2 years?"	Perm	

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

6.3.5.1 Assumptions for Questionnaire (QS) domain model

1. Questionnaire data may include, but are not limited to the following: subject reported outcomes, validated or non-validated questionnaires, and validated or non-validated diaries (i.e., data with a common topicality). When objective numeric data with result qualifiers are collected in a questionnaire or diary format, the sponsor should consider whether this data actually belongs in a separate (new) domain. It is realized that questionnaires may include some objective questions and some of these may trigger additional data collection in other domains.
2. The names of the questionnaires should be described under the variable QSCAT in the questionnaire domain. For example, ADAS for Alzheimer's Disease Assessment Scale, SF36 for SF-36 Health Survey, PANSS for Positive and Negative Syndrome Scale, etc.
3. Names of categories/subcategories for groups of items/questions should be described under QSSCAT. Derived information such as, total scores, and sub scores, etc, should be stored in the QS domain as derived records with appropriate category/subcategory names (QSCAT/QSSCAT), item names (QSTESTCD/QSTEST), and results (QSSTRESN/QSSTRESC). Derived records should be flagged by QSDRVFL.

6.3.6 Subject Characteristics — SC

sc.xpt, Subject Characteristics — Findings, Version 3.1.1, August 26, 2005. One record per characteristic per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**SC	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
SCSEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
SCGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to tie together a block of related records in a single domain to support relationships within the domain and between domains.	Perm	SDTMIG 2.1 ; SDTM 2.2.4
SCSPID	Sponsor-Defined Identifier	Char		Sponsor Defined	Identifier	Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Question number on a questionnaire.	Perm	SDTM 2.2.4
SCTESTCD	Subject Characteristic Short Name	Char	*	CRF or Derived	Topic	Short name of the measurement, test, or examination described in SCTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in SCTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g. '1TEST'). SCTESTCD cannot contain characters other than letters, numbers, or underscores. Example: SUBJINIT, EYECD.	Req	
SCTEST	Subject Characteristic	Char	*	CRF	Synonym Qualifier	Verbatim name of the test or examination used to obtain the measurement or finding. The value in SCTEST cannot be longer than 40 characters. Examples: Subject Initials, Eye Color.	Req	
SCCAT	Category for Subject Characteristic	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records.	Perm	SDTMIG 2.1
SCSCAT	Subcategory for Subject Characteristic	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of the subject characteristic.	Perm	SDTMIG 2.1
SCORRES	Result or Finding in Original Units	Char		CRF or Derived	Result Qualifier	Result of the subject characteristic as originally received or collected.	Exp	SDTMIG 4.1.5.1
SCORRESU	Original Units	Char	*	CRF or Derived	Variable Qualifier	Original Unit in which the data were collected. The unit for SCORRES.	Perm	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
SCSTRESC	Character Result/Finding in Std Format	Char		Derived	Result Qualifier	Contains the result value for all findings, copied or derived from SCORRES in a standard format or standard units. SCSTRESC should store all results or findings in character format; if results are numeric, they should also be stored in numeric format in SCSTRESN. For example, if a test has results 'NONE', 'NEG', and 'NEGATIVE' in SCORRES and these results effectively have the same meaning, they could be represented in standard format in SCSTRESC as 'NEGATIVE'. For other examples, see general assumptions .	Exp	SDTMIG 4.1.5.1
SCSTRESN	Numeric Result/Finding in Standard Units	Num		Derived	Result Qualifier	Used for continuous or numeric results or findings in standard format; copied in numeric format from SCSTRESC. SCSTRESN should store all numeric test results or findings.	Perm	SDTMIG 4.1.5.1
SCSTRESU	Standard Units	Char	*	CRF or Derived	Variable Qualifier	Standardized unit used for SCSTRESC or SCSTRESN.	Perm	
SCSTAT	Status of SC Measurement	Char	**NOT DONE	CRF or Derived	Record Qualifier	Used to indicate that the measurement was not done. Should be null if a result exists in SCORRES.	Perm	SDTMIG 4.1.5.1
SCREASND	Reason Not Performed	Char		CRF or Derived	Record Qualifier	Describes why a question was not answered. Example: subject refused. Used in conjunction with SCSTAT when value is NOT DONE.	Perm	
SCDTC	Date/Time of Collection	Char	ISO 8601	CRF or Derived	Timing		Exp	SDTMIG 4.1.4.1
SCDY	Study Day of Examination	Num		Derived	Timing	1. Study day of collection, measured as integer days. 2. Algorithm for calculations must be relative to the sponsor-defined RFSTDTC variable in Demographics. This formula should be consistent across the submission.	Perm	SDTMIG 4.1.4.4

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

6.3.6.1 Assumptions for Subject Characteristics (SC) domain model

1. Subject Characteristics is for data not collected in other domains that is subject-related. Examples: subject initials, other race details, eye color, childbearing status, etc.
2. The structure for demographic data is fixed and includes date of birth, age, sex, race, ethnicity and country. The structure of subject characteristics is based on the Findings general class and is an extension of the demographics data. Subject Characteristics consists of data that is collected once per subject (per test) and that is not expected to change during the trial. Subject characteristics contains data such as additional information about race, subject initials, economic information, and eye color.

6.3.7 Vital Signs — VS

vs.xpt, Vital Signs — Findings, Version 3.1.1, August 26, 2005. One record per vital sign measurement per time point per visit per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	Reference
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	**VS	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req	SDTM 2.2.4
VSSEQ	Sequence Number	Num		CRF or Derived	Identifier	Sequence number given to ensure uniqueness within a dataset for a subject. Can be used to join related records.	Req	SDTM 2.2.4
VSGRPID	Group ID	Char		Sponsor Defined	Identifier	Used to tie together a block of related records in a single domain to support relationships within the domain and between domains.	Perm	SDTMIG 2.1 ; SDTM 2.2.4
VSSPID	Sponsor-Defined Identifier	Char		Sponsor Defined	Identifier	Optional Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database.	Perm	
VSTESTCD	Vital Signs Test Short Name	Char	**	CRF or Derived	Topic	Short name of the measurement, test, or examination described in VSTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in VSTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g. '1TEST'). VSTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: SYSBP, DIABP, BMI.	Req	
VSTEST	Vital Signs Test Name	Char	**	CRF	Synonym Qualifier	Verbatim name of the test or examination used to obtain the measurement or finding. The value in VSTEST cannot be longer than 40 characters. Examples: Systolic Blood Pressure, Diastolic Blood Pressure, Body Mass Index.	Req	SDTMIG 2.1
VSCAT	Category for Vital Signs	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records.	Perm	SDTMIG 2.1
VSSCAT	Subcategory for Vital Signs	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of a measurement or examination.	Perm	SDTMIG 2.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	Reference
VSPOS	Vital Signs Position of Subject	Char	*	CRF or Derived	Record Qualifier	Position of the subject during a measurement or examination. Examples: SUPINE, STANDING, SITTING.	Perm	
VSORRES	Result or Finding in Original Units	Char		CRF or Derived	Result Qualifier	Result of the vital signs measurement as originally received or collected.	Exp	SDTMIG 4.1.5.1
VSORRESU	Original Units	Char	*	CRF or Derived	Variable Qualifier	Original units in which the data were collected. The unit for VSORRES. Examples: INCHES, FEET, POUNDS, BEATS PER MINUTE.	Exp	
VSSTRESC	Character Result/Finding in Std Format	Char		Derived	Result Qualifier	Contains the result value for all findings, copied or derived from VSORRES in a standard format or standard units. VSSTRESC should store all results or findings in character format; if results are numeric, they should also be stored in numeric format in VSSTRESN. For example, if a test has results 'NONE', 'NEG', and 'NEGATIVE' in VSORRES and these results effectively have the same meaning, they could be represented in standard format in VSSTRESC as 'NEGATIVE'. For other examples, see general assumptions .	Exp	SDTMIG 4.1.5.1
VSSTRESN	Numeric Result/Finding in Standard Units	Num		Derived	Result Qualifier	Used for continuous or numeric results or findings in standard format; copied in numeric format from VSSTRESC. VSSTRESN should store all numeric test results or findings.	Exp	SDTMIG 4.1.5.1
VSSTRESU	Standard Units	Char	*	CRF or Derived	Variable Qualifier	Standardized unit used for VSSTRESC or VSSTRESN.	Exp	SDTMIG 4.1.5.1
VSNRIND	Reference Range Indicator	Char	*	CRF or Derived	Variable Qualifier	1. Indicates where value falls with respect to reference range (which could be defined by VSORNRLO and VSORNRHI or by VSSTNRC (if those variables are used). Examples: NORMAL, ABNORMAL, HIGH, LOW. 2. Should not be used to indicate clinical significance.	Perm	
VSSTAT	Vitals Status	Char	**NOT DONE	CRF or Derived	Record Qualifier	Used to indicate that a vital sign measurement was not done. Should be null if a result exists in VSORRES.	Perm	SDTMIG 4.1.5.1
VSREASND	Reason Not Performed	Char		CRF or Derived	Record Qualifier	Describes why a measurement or test was not performed. Examples: BROKEN EQUIPMENT or SUBJECT REFUSED. Used in conjunction with VSSTAT when value is NOT DONE.	Perm	SDTMIG 4.1.5.1
VSLOINC	LOINC Code	Char	**	Derived	Synonym Qualifier	1. LOINC Code for VSTEST. 2. The LOINC version should be provided in the Sponsor Comments column in the data definition document.	Perm	SDTMIG 4.1.3.2
VSLOC	Location of Vital Signs Measurement	Char	*	CRF or Derived	Record Qualifier	Location relevant to the collection of Vital Signs measurement. Example: LEFT ARM for blood pressure.	Perm	SDTMIG 4.1.3.7

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	Reference
VSBLFL	Baseline Flag	Char	**Y or Null	CRF or Derived	Record Qualifier	Indicator used to identify a baseline value	Exp	SDTMIG 4.1.3.7
VSDRVFL	Derived Flag	Char	**Y or Null	Derived	Record Qualifier	Used to indicate a derived record. The value should be Y or null. Records which represent the average of other records or questionnaire sub-scores which do not come from the CRF are examples of records which would be derived for the submission datasets. If VSDRVFL=Y, then VSORRES should be NULL, with VSSTRESC and (if numeric) VSSTRESN having the derived value.	Perm	SDTMIG 4.1.5.1 SDTMIG 4.1.3.7
VISIT	Visit Name	Char		CRF or Derived	Timing	1. Protocol-defined description of clinical encounter. 2. May be used in addition to VISITNUM and/or VISITDY.	Perm	SDTMIG 7.8; SDTMIG 4.1.4.5
VISITNUM	Visit Number	Num		CRF or Derived	Timing	1. Clinical encounter number. 2. Numeric version of VISIT, used for sorting.	Req	SDTMIG 7.8; SDTMIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDTMIG 7.8; SDTMIG 4.1.4.5
VSDTC	Date/Time of Measurements	Char	ISO 8601	CRF or Derived	Timing		Exp	SDTMIG 4.1.4.1
VSDY	Study Day of Vital Signs	Num		Derived	Timing	1. Study day of vital signs measurements, measured as integer days. 2. Algorithm for calculations must be relative to the sponsor-defined RFSTDTC variable in Demographics. This formula should be consistent across the submission.	Perm	SDTMIG 4.1.4.4
VSTPT	Planned Time Point Name	Char		CRF or Derived	Timing	1. Text Description of time when measurement should be taken. 2. This may be represented as an elapsed time relative to a fixed reference point, such as time of last dose. See VSTPTNUM and VSTPTREF. Examples: Start, 5 min post.	Perm	
VSTPTNUM	Planned Time Point Number	Num		CRF or Derived	Timing	Numerical version of VSTPT to aid in sorting.	Perm	
VSELTM	Elapsed Time from Reference Point	Char	ISO 8601	Derived	Timing	Elapsed time (in ISO 8601) relative to a planned fixed reference (VSTPTREF). This variable is useful where there are repetitive measures. Not a clock time or a date time variable. Examples: '-P15M' to represent the period of 15 minutes prior to the reference point indicated by VSTPTREF, or 'P8H' to represent the period of 8 hours after the reference point indicated by VSTPTREF.	Perm	SDTMIG 4.1.4.3
VSTPTREF	Time Point Reference	Char		Sponsor Defined	Timing	Name of the fixed reference point referred to by VSELTM, VSTPTNUM, and VSTPT. Examples: PREVIOUS DOSE, PREVIOUS MEAL.	Perm	

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

7 Trial Design Datasets

7.1 INTRODUCTION

The Trial Design Model, which is included in the SDTM, allows description of key aspects of the planned conduct of a clinical trial in a standardized way. These standardized descriptions will allow reviewers to:

- clearly and quickly grasp the design of a clinical trial
- compare the designs of different trials
- search a data warehouse for clinical trials with certain features
- compare planned and actual treatments and visits for subjects in a clinical trial.

Modeling a clinical trial in this standardized way requires the explicit statement of certain decision rules that may not be addressed or may remain vague or ambiguous in the usual prose protocol document. Prospective modeling of the design of a clinical trial should lead to a clearer, better protocol. Retrospective modeling of the design of a clinical trial should ensure a clear description of how the trial was interpreted by the sponsor.

The aspects of clinical trial design currently included in the Trial Design Model are as follows:

- the planned arms of the trial
- what happens to a subject in each arm (i.e., what series of treatment and non-treatment time periods [trial Elements] are planned for a subject assigned to that arm)
- the planned schedule of visits
- the inclusion and exclusion criteria for the trial.

Future versions of the Trial Design Model are expected to include additional aspects of clinical trials, such as planned exposures and planned assessments.

The Trial Design Model is built upon the concepts of Elements, Arms, Epochs and Visits. An Element is the basic building block for time within a clinical trial, and has the following characteristics:

- A description of what happens to the subject during the Element.
- A definition of the start of the Element.
- A rule for ending the Element.

An Arm is a planned sequence of Elements, typically equivalent to a treatment group. Branches may take place between one Element and the next and some designs allow for some flexibility of Elements within an Arm. The term EPOCH (which is discussed in [Section 7.4.4](#)) is used to refer to the portion of a blinded trial that corresponds to an individual element (when the actual element may not be known).

A Visit is defined as a clinical encounter that encompasses planned and unplanned trial interventions, procedures, and assessments that may be performed on a subject. A Visit has a start and an end, each described with a rule. A Visit need not be nested within an Element. In other words, it may start in one Element and end in another.

In most blinded trials, the timing of Visits is the same for all subjects, regardless of the Arm to which they have been assigned. In these cases, the Arm is not needed to describe the timing of Visits, and is left blank in the Trial Visits domain. If the timing of Visits depends on Arm, then the complete set of Visits for each Arm should be represented in the domain.

The Trial Design Model also includes the Trial Inclusion/Exclusion (TI) dataset to describe the inclusion/exclusion criteria used to screen subjects. The IE domain (subject specific inclusion/exclusion criteria not met) described in [section 6.3.2](#) contains the actual exceptions to those criteria for enrolled subjects.

7.2 PLANNED ELEMENTS, ARMS, AND VISITS

Under the Trial Design Model, planned information is presented in a series of three tables:

- The Trial Element table (Table 7.2.1)
- The Trial Arms table (Table 7.2.2)
- The Trial Visits table (Table 7.2.3).

7.2.1 Trial Elements

te.xpt, Trial Elements, Version 3.1.1, August 26, 2005. One record per Planned Element

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req
DOMAIN	Domain Abbreviation	Char	**TE	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req
ETCD	Element Code	Char	*	Sponsor Defined	Topic	Short name of ELEMENT used for programming and sorting.	Req
ELEMENT	Description of Element	Char	*	Sponsor Defined / Protocol	Synonym Qualifier	The name of the Element.	Req
TESTRL	Rule for Start of Element	Char		Sponsor Defined / Protocol	Rule	Expresses rule for beginning Element.	Req
TEENRL	Rule for End of Element	Char		Sponsor Defined / Protocol	Rule	Expresses rule for ending Element. Either TEENRL or TEDUR must be present for each element.	Perm
TEDUR	Planned Duration of Element	Char	ISO 8601	Sponsor Defined / Protocol	Timing	Planned Duration of Element in ISO 8601 format. Used when the rule for ending the Element is to applied after a fixed duration.	Perm

* indicates variable may be subject to sponsor-controlled terminology;

** indicates variable may be subject to external controlled terminology.

7.2.2 Trial Arms

ta.xpt, Trial Arms, Version 3.1.1, August 26, 2005. One record per Planned Element per Arm

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req
DOMAIN	Domain Abbreviation	Char	**TA	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req
ARMCD	Arm Code	Char	*	Derived	Topic	Short name of ARM., used for programming and sorting.	Req
ARM	Description of Arm	Char	*	Derived	Synonym Qualifier	Name given to an Arm or treatment group.	Req
TAETORD	Order of Element within Arm	Num		Sponsor Defined	Identifier	Number that gives the order of the Element within the Arm.	Req
ETCD	Element Code	Char	*	Sponsor Defined	Record Qualifier	Short name of ELEMENT used for programming and sorting.	Req
ELEMENT	Description of Element	Char	*	Sponsor Defined / Protocol	Synonym Qualifier	The name of the Element.	Perm
TABRANCH	Branch	Char			Rule	Condition subject met, at a 'branch' in the trial design at the end of this Element, to be included in this Arm; (e.g., randomization to Drug A).	Exp
TATRANS	Transition Rule	Char			Rule	If the trial design allows a subject to transition to an Element other than the next Element in sequence, then the conditions for transitioning to those other Elements, and the alternative Element sequences, are specified in this rule (e.g., responders go to washout).	Exp
EPOCH	Trial Epoch	Char	*		Timing	Name of the Trial Epoch with which this Element of the Arm is associated.	Perm

Note: The same Element may occur more than once within an Arm.

** indicates variable may be subject to sponsor-controlled terminology;*

*** indicates variable may be subject to external controlled terminology.*

7.2.3 Trial Visits

tv.xpt, Trial Visits, Version 3.1.1, August 26, 2005. One Record per Planned Visit per Arm

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req
DOMAIN	Domain Abbreviation	Char	**TV	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation	Req
VISITNUM	Visit Number	Num		CRF or Derived	Topic	1. Clinical encounter number 2. Numeric version of VISIT, used for sorting.	Req
VISIT	Visit Name	Char		CRF or Derived	Synonym Qualifier	1. Protocol-defined description of clinical encounter. 2. May be used in addition to VISITNUM and/or VISITDY as a text description of the clinical encounter.	Perm
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing	1. Planned study day of VISIT. 2. Due to its sequential nature, used for sorting.	Perm
ARMCD	Arm Code	Char	*	Derived	Record Qualifier	1 Short name of ARM, used for programming and sorting. 2. If the timing of visits for a trial does not depend on which ARM a subject is in, then ARMCD should be null.	Exp
ARM	Description of Arm	Char	*	Derived	Synonym Qualifier	1. Name given to an Arm or Treatment Group. 2. If the timing of Visits for a trial does not depend on which Arm a subject is in, then Arm should be left blank.	Perm
TVSTRL	Visit Start Rule	Char		Derived	Rule	Rule describing when the Visit starts, in relation to the sequence of Elements. Used only when Visits are dependent on occurrences within the study, not fixed by protocol.	Perm
TVENRL	Visit End Rule	Char		Sponsor Defined / Protocol	Rule	Rule describing when the Visit ends, in relation to the sequence of Elements.	Perm

* indicates variable may be subject to sponsor-controlled terminology;

** indicates variable may be subject to external controlled terminology.

7.3 SUBJECT ELEMENTS AND VISITS

While the Trial Elements, Trial Arms and Trial Visits datasets describe the planned design of the study, it is also necessary to collect the corresponding actual data. Subject assignment to an Arm is captured in the ARM variable in Demographics. Actual Elements and Visits data for each subject are described in two additional datasets:

- The Subject Elements dataset (Table 7.3.1)
- The Subject Visits dataset (Table 7.3.2).

7.3.1 Subject Elements

se.xpt ,Subject Elements, Version 3.1.1, August 26, 2005. One record per Actual Element per Subject

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req
DOMAIN	Domain Abbreviation	Char	**SE	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req
SESEQ	Sequence Number	Num		Derived	Identifier	Sequence number given to ensure uniqueness within dataset. Can be used to join related records.	Req
ETCD	Subject Element Code	Char	*	Sponsor Defined	Topic	1. Short name of ELEMENT, used for programming and sorting. 2. If an encountered Element differs from the planned Element to the point that it is considered a new Element, then use 'UNPLAN' as the value for ETCD to represent this Element.	Req
ELEMENT	Description of Subject Element	Char	*	Sponsor Defined / Protocol	Synonym Qualifier	The name of the Element. If an encountered Element differs from the planned ELEMENT to the point that it is considered a new ELEMENT, then ELEMENT should be Null	Perm
SESTDTC	Start Date/ Time of Element	Char	ISO 8601	CRF or Derived	Timing	Start date/time for an Element for each subject.	Exp
SEENDTC	End Date/ Time of Element	Char	ISO 8601	CRF or Derived	Timing	End date/time of an Element for each subject.	Exp
SEUPDES	Description of Unplanned Element	Char		Sponsor Defined	Synonym Qualifier	Description of what happened to the subject during this unplanned Element. Used only if ETCD has the value of 'UNPLAN'.	Perm

* indicates variable may be subject to sponsor-controlled terminology;

** indicates variable may be subject to external controlled terminology.

7.3.2 Subject Visits

sv.xpt, Subject Visits, Version 3.1.1, August 26, 2005. One record per Subject per Actual Visit

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req
DOMAIN	Domain Abbreviation	Char	**SV	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req
VISITNUM	Visit Number	Num		CRF or Derived	Topic	1. Clinical encounter number. (Decimal numbering may be useful for inserting unplanned visits.) 2. Numeric version of VISIT, used for sorting.	Req
VISIT	Visit Name	Char		CRF or Derived	Synonym Qualifier	1. Protocol-defined description of clinical encounter. 2. May be used in addition to VISITNUM and/or VISITDY as a text description of the clinical encounter.	Perm
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing	Planned study day of VISIT.	Perm
SVSTDTC	Start Date/Time of Visit	Char	ISO 8601	CRF or Derived	Timing	Start date/time for a Visit.	Exp
SVENDTC	End Date/Time of Visit	Char	ISO 8601	CRF or Derived	Timing	End date/time of a Visit.	Exp
SVUPDES	Description of Unplanned Visit	Char		CRF or Derived	Synonym Qualifier	Description of what happened to the subject during an unplanned visit.	Perm

* indicates variable may be subject to sponsor-controlled terminology;

** indicates variable may be subject to external controlled terminology.

7.4 TRIAL ARMS

The core of the Trial Design models is the Trial Arms domain (TA). It contains one record for each occurrence of an Element in each Arm of the trial.

7.4.1 Identifying Trial Arms

Arms are a familiar concept in Clinical Trials. They usually correspond to treatment groups.

Example: Figure 1

Flowchart, Parallel Trial

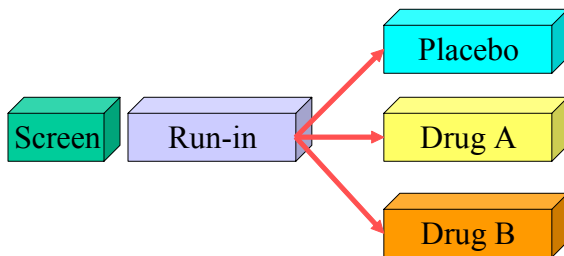


Figure 1

In this example, the number of Arms is three. However, in some trials with multiple branchings, the number of Arms in the trial may not be obvious.

Example: Figure 2

Counting Arms in More Complex Trials

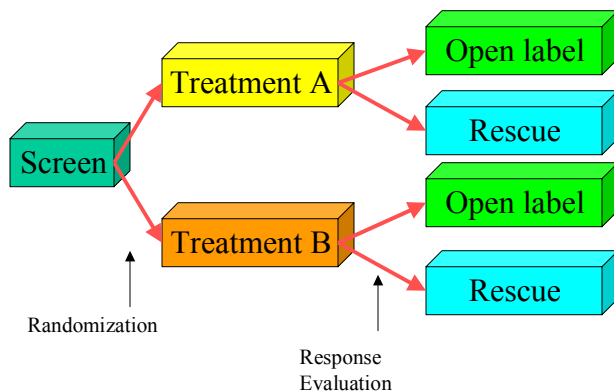


Figure 2

In this example, subjects are randomized to one of two treatment groups. At the end of this first treatment period, they are assigned to one of two further treatment groups on the basis of their response to their initial period of treatment. The combination of these two major branchings results in four Arms for this trial.

See [Section 7.4.7](#) for additional considerations that arise in a multi-stage trial design like this one.

See [Section 7.4.3](#) for additional discussion of when a decision point in a trial design should be considered to give rise to a new Arm.

7.4.2 Developing the Trial Arms Table

In Figure 1 (shown previously), the trial has three Arms. This diagram is similar to the kind of study schema or flowchart usually included in most trial protocols. The 'building blocks' shown in these diagrams are referred to within the Trial Design Model as 'Elements.' Further details on how Elements are defined are in [Section 7.5](#)

Parallel Trial Example: Figure 3

Arms Representation, Parallel Trial

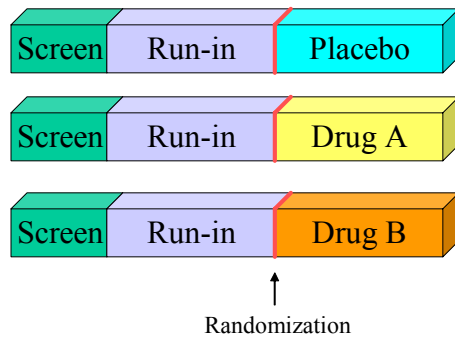


Figure 3

The trial design can be represented as three parallel Arms, as in this diagram. The screen and run-in Elements, which occur before the branching at the point of randomization, are repeated in each Arm.

Note that the randomization is represented by red lines at the ends of the run-in Elements in the three Arms. Randomizations, and other branchings in the trial design, are represented by such markers between Elements.

One kind of branching, other than a randomization, is an assignment based on the subject's clinical response to treatment. This was mentioned in the example in [Section 7.4.1](#). Branchings could also depend on a subject's disease severity, or any other clinical finding.

Another kind of branching occurs in dose-escalation cohort studies. In this kind of trial design, there are rules that say whether the next subject should be in the current cohort (receive the current dose level) depending on how many subjects are in the current cohort and what their response to treatment has been. For example, the rule might say that five subjects should be entered in each cohort, and that additional subjects would be added to a cohort if a certain proportion of subjects experienced dose-limiting toxicity. The rule might say that entry of subjects (and the addition of new cohorts) would stop when a certain proportion of subjects in such an enlarged cohort experiences dose-limiting toxicity.

The diagram in Figure 3 can be further transformed into a table to aid in creating the trial design datasets. There is one row in the table for each Element in each Arm in Figure 3. Note how the column labeled 'Branch' captures the information represented by the red markers in the diagram. Note also that the value in the Branch column shows the randomization result that is needed for a subject to be in that Arm.

ARMCD	TAETORD	ELEMENT	TABRANCH
Placebo	1	Screen	
Placebo	2	Run-In	Randomized to Placebo
Placebo	3	Placebo	
A	1	Screen	
A	2	Run-In	Randomized to Drug A
A	3	Drug A	
B	1	Screen	
B	2	Run-In	Randomized to Drug B
B	3	Drug B	

Crossover Trial Example: Figure 4

Flowchart, Crossover Trial

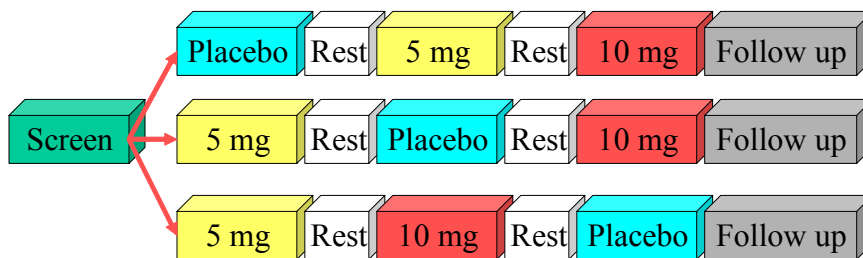


Figure 4

This trial is similar to the Parallel-Trial example, except that there are multiple treatment periods. Note that the Arms are distinguished by the order of treatments, rather than the presence of one treatment or another. Note also that rest Elements occur twice within each Arm.

CrossOver Trial Example: Figure 5

Arms Representation, Crossover Trial

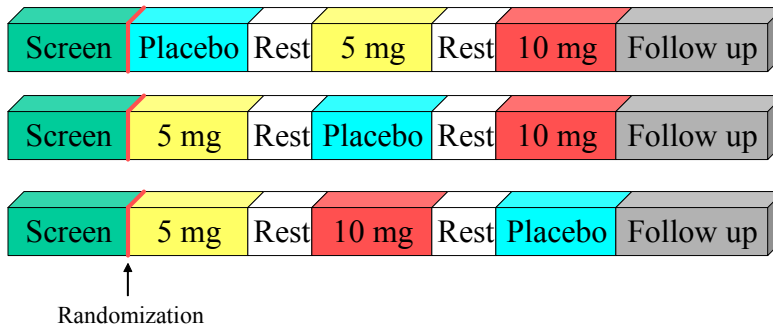


Figure 5

The trial design is represented with three parallel Arms. This transformation is similar to that shown in the Parallel Trial example (previously shown Figures 1-3).

As for the previous example, the diagram in Figure 5 can be further transformed into a table. Only the first 10 of the 21 rows of the table for this trial are shown in this example table.

ARM	TAETORD	ELEMENT	TABRANCH
P-5-10	1	Screen	Randomized to Placebo - 5 mg - 10 mg
P-5-10	2	Placebo	
P-5-10	3	Rest	
P-5-10	4	5 mg	
P-5-10	5	Rest	
P-5-10	6	10 mg	
P-5-10	7	Follow-up	
5-P-10	1	Screen	Randomized to 5 mg - Placebo - 10 mg
5-P-10	2	5 mg	
5-P-10	3	Rest	

Oncology Trial Example: Figure 6

Flowchart, Oncology Trial

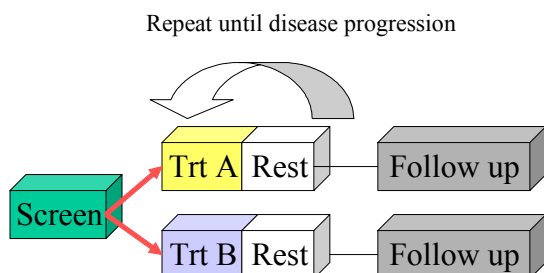
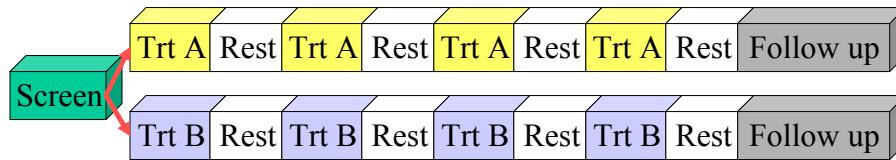


Figure 6

Many oncology trials feature repeating cycles of treatment. As indicated, there is a condition for repeating the cycle.

Oncology Trial Example: Figure 7

Flowchart with Repeats, Oncology Trial



Note: If protocol does not specify maximum number of cycles, use actual maximum number of cycles that occurred in trial.

Figure 7

The first step in transforming this design into tabular format is to put in the repeating cycles, replacing the arrow indicating cycling through the treatment Element.

If the protocol indicates a maximum number of permitted cycles, that number of cycles should be included. If the protocol does not specify a maximum, then use maximum number of cycles that actually occurred in the trial. At the time of submission, the trial has completed, and the maximum number of cycles is known.

Oncology Trial Example: Figure 8

Arm Representation, Oncology Trial

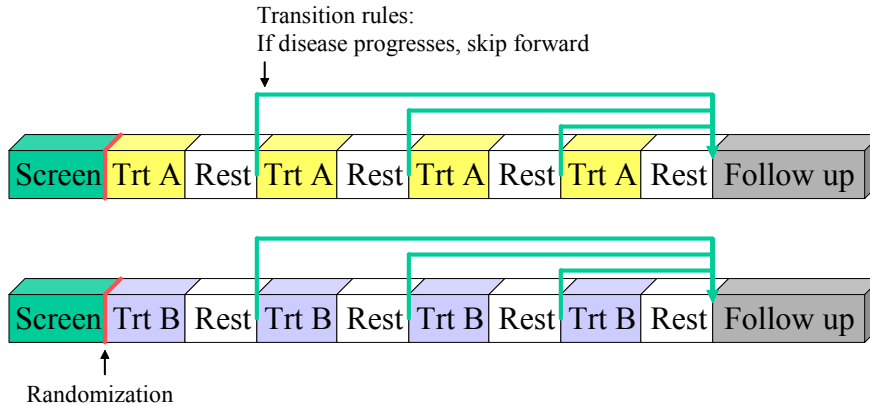


Figure 8

This diagram shows parallel Arms, as before. The randomization is represented by red markers, as before.

The green arrows represent the fact that subjects may not pass through all the cycles in the trial. In this case, a subject whose disease progresses will skip further treatment cycles and jump to the follow-up Element. In previous examples, the implicit rule was 'when the subject finishes one Element, move to the next.' In this trial, at the end of each Rest Element, the subject may move to one of two Elements. Thus there is a need for an explicit transition rule.

As before, this diagram can be transformed into a table. However, this table requires another column, for the transition rule. If this column is blank, then the default transition rule, 'go to the next Element in sequence' is assumed. Only the rows for one Arm are shown in this example table.

ARM	TAETORD	ELEMENT	TABRANCH	TATRANS
A	1	Screen	Randomized to A	
A	2	Trt A		
A	3	Rest		If disease progression, go to 10 th Element.
A	4	Trt A		
A	5	Rest		If disease progression, go to 10 th Element.
A	6	Trt A		
A	7	Rest		If disease progression, go to 10 th Element.
A	8	Trt A		
A	9	Rest		
A	10	Follow-up		

7.4.3 Distinguishing between Branches and Transitions

Both the Branch and Transition columns contain rules, but there are two columns because they represent two different types of rules. Branch rules represent forks in the trial flowchart, giving rise to separate Arms. The Branch rules thus appear in multiple Arms in the Trial Arms dataset. Within any one Arm, there is no choice in the value of the Branch condition. e.g., the subject must have been randomized to Arm A. Transition rules are used for choices within an Arm. In the Oncology example shown in Figure 8, subjects who receive 1, 2, 3, or 4 cycles are all considered to belong to Arm A.

In modeling a trial, decisions may have to be made about whether a decision point in the flow chart represents the separation of two or more separate Arms, or simply to variations within the same Arm. This decision will depend on the comparisons of interest in the trial.

7.4.4 Trial Epoch Concept

In all the trials we have considered thus far, the several trial Arms are similar in that they have the same numbers of Elements and the same pattern of treatment and non-treatment Elements.

Example: Figure 9

Epochs, Crossover Trial

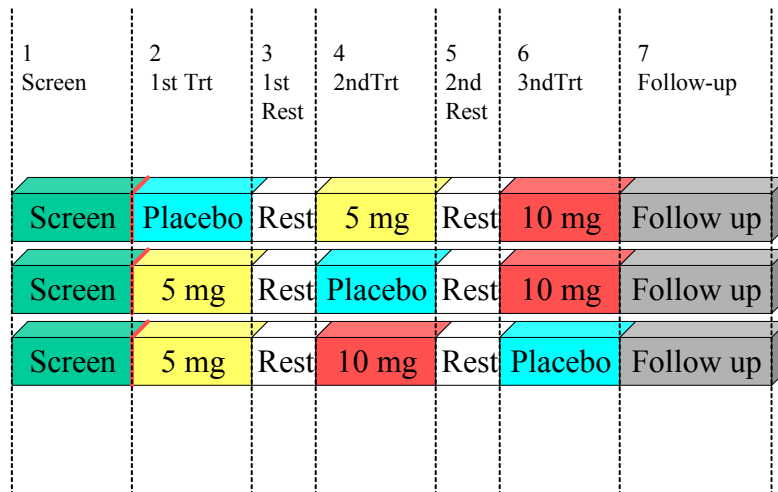


Figure 9

This diagram shows the similarity between Arms by means of dotted vertical lines that divide the trial as a whole, across all Arms, into 'Epochs.' When the trial elements form a grid in this way, the rows are Arms and the columns are Epochs. The numbers assigned to the columns correspond to TAETORD, the 'Order' column in the example tables. The names assigned to the columns are the names of the Epochs. The names suggested in this example represent familiar concepts. Only the terminology 'Epoch' is new.

The great majority of clinical trials, particularly blinded clinical trials, have this grid structure of Epochs and Arms. During the conduct of a blinded study, when the Arm assignment of a subject is unknown, a subject's progress through the trial must be spoken of in terms of Epochs, since the subject's specific treatment Elements are unknown.

The concept of Epochs is not universally applicable. The following examples show trials for which Epochs are applicable and are not applicable and some that fall in a gray area.

In Figure 9, the Elements of the Arms fall into columns because the blocks representing the Elements in a column all have the same width. The widths of these blocks represent the Element durations, which are determined by their

End Rules. (Element Start and End Rules are described in [Section 7.5.](#)) The following is a join of the TA and TE datasets for this trial (i.e., the TA dataset, to which the Start Rule (TESTRL) and Duration (TEDUR) from the TE dataset have been added).

ARM	TAETORD	ELEMENT	TESTRL	TEENRL	TEDUR	TABRANCH
P-5-10	1	Screen	Informed consent		P14D	Randomized to Placebo - 5 mg - 10 mg
P-5-10	2	Placebo	First dose of placebo		P14D	
P-5-10	3	Rest	48 hrs after last dose drug		7D	
P-5-10	4	5 mg	First dose of 5 mg drug		P14D	
P-5-10	5	Rest	48 hrs after last dose drug		7D	
P-5-10	6	10 mg	First dose of 10 mg drug		P14D	
P-5-10	7	Follow-up	48 hrs after last dose drug		P21D	
5-P-10	1	Screen	Informed consent		P14D	Randomized to 5 mg - Placebo - 10 mg
5-P-10	2	5 mg	First dose of 5 mg drug		P14D	
5-P-10	3	Rest	48 hrs after last dose drug		7D	
5-P-10	4	Placebo	First dose of placebo		P14D	
5-P-10	5	Rest	48 hrs after last dose drug		7D	
5-P-10	6	10 mg	First dose of 10 mg drug		P14D	
5-P-10	7	Follow-up	48 hrs after last dose drug		P21D	
5-10-P	1	Screen	Informed consent		P14D	Randomized to 5 mg - 10 mg - Placebo
5-10-P	2	5 mg	First dose of 5 mg drug		P14D	
5-10-P	3	Rest	48 hrs after last dose drug		7D	
5-10-P	4	10 mg	First dose of 10 mg drug		P14D	
5-10-P	5	Rest	48 hrs after last dose drug		7D	
5-10-P	6	Placebo	First dose of placebo		P14D	
5-10-P	7	Follow-up	48 hrs after last dose drug		P21D	

Sorting this dataset by TAETORD displays the similarity between Elements with the same value of TAETORD. For each value of TAETORD, the End Rules are the same (not shown with the example) and the Start Rules are either the same or differ only by drug dose. Each value of TAETORD corresponds to an Epoch, and the associated Start and End Rules of the Elements associated with that value of TAETORD represent, in effect, the Start and End Rules for the associated Epoch.

ARM	TAETORD	EPOCH	ELEMENT	TESTRL	TEDUR	TABRANCH
5-P-10	1	Trial Screen	Screen	Informed consent	P14D	Randomized to 5 mg/Placebo/10 mg
P-5-10	1	Trial Screen	Screen	Informed consent	P14D	Randomized to Placebo/5 mg/10 mg
5-10-P	1	Trial Screen	Screen	Informed consent	P14D	Randomized to 5 mg/10 mg/Placebo
5-P-10	2	First Treatment	5 mg	First dose of 5 mg drug	P14D	
P-5-10	2	First Treatment	Placebo	First dose of placebo	P14D	
5-10-P	2	First Treatment	5 mg	First dose of 5 mg drug	P14D	
5-P-10	3	First Rest	Rest	48 hrs after last dose drug	7D	
P-5-10	3	First Rest	Rest	48 hrs after last dose drug	7D	
5-10-P	3	First Rest	Rest	48 hrs after last dose drug	7D	
5-P-10	4	Second Treatment	Placebo	First dose of placebo	P14D	
P-5-10	4	Second Treatment	5 mg	First dose of 5 mg drug	P14D	
5-10-P	4	Second Treatment	10 mg	First dose of 10 mg drug	P14D	
5-P-10	5	Second Rest	Rest	48 hrs after last dose drug	7D	
P-5-10	5	Second Rest	Rest	48 hrs after last dose drug	7D	
5-10-P	5	Second Rest	Rest	48 hrs after last dose drug	7D	
5-P-10	6	Third Treatment	10 mg	First dose of 10 mg drug	P14D	
P-5-10	6	Third Treatment	10 mg	First dose of 10 mg drug	P14D	
5-10-P	6	Third Treatment	Placebo	First dose of placebo	14D	
5-P-10	7	Trial Follow-up	Follow-up	48 hrs after last dose drug	P21D	
P-5-10	7	Trial Follow-up	Follow-up	48 hrs after last dose drug	P21D	
5-10-P	7	Trial Follow-up	Follow-up	48 hrs after last dose drug	P21D	

Inconsistent numbers and durations of Elements Example: Figure 10

Trial with Inconsistent Numbers and Durations of Elements

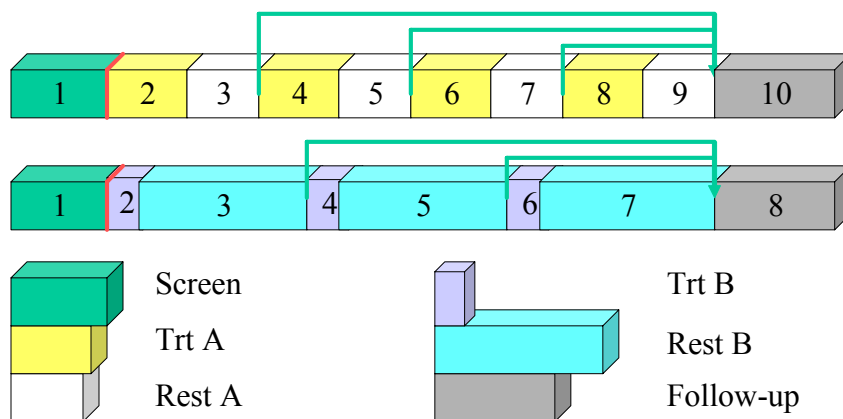


Figure 10

The trial in Figure 10 compares cancer therapies given as 3-week and 4-week cycles. The table below is a join of TA and TE datasets for this trial, sorted by TAETORD. Note that the number of Elements in Arm A is different from the number of Elements in Arm B and that the Start and End Rules for Elements with the same value of TAETORD are only loosely comparable for the 2nd through the 7th Elements. It is questionable whether it would be meaningful to compare data for the 2nd Elements through the 7th Elements, not meaningful to compare the 8th Elements, and there are no 9th and 10th Elements in Arm B for comparison

It would be reasonable to compare data for the Screen and Follow-up Elements. The existing Element variable is available for selecting data for this comparison, so it is not necessary to define Epochs to facilitate this comparison.

ARM	TAETORD	EPOCH	ELEMENT	TESTRL	TEDUR	TABRANCH	TATRANS
A	1		Screen	Informed consent	14D	Randomized to A	
B	1		Screen	Informed consent	14D	Randomized to B	
A	2		Trt A	First dose A	9D		
B	2		Trt B	First dose B	P7D		
A	3		Rest A	48 hrs after last dose A	12D		If disease progresses, go to 10 th Element
B	3		Rest B	48 hrs after last dose B	21D		If disease progresses, go to 8 th Element
A	4		Trt A	First dose A	9D		
B	4		Trt B	First dose B	P7D		
A	5		Rest A	48 hrs after last dose A	12D		If disease progresses, go to 10 th Element
B	5		Rest B	48 hrs after last dose B	21D		If disease progresses, go to 8 th Element

ARM	TAETORD	EPOCH	ELEMENT	TESTRL	TEDUR	TABRANCH	TATRANS
A	6		Trt A	First dose A	9D		
B	6		Trt B	First dose B	P7D		
A	7		Rest A	48 hrs after last dose A	12D		If disease progresses, go to 10 th Element
B	7		Rest B	48 hrs after last dose B	21D		
A	8		Trt A	First dose A	9D		
B	8		Follow-up	Decision not to treat further	90D		
A	9		Rest A	48 hrs after last dose A	12D		
A	10		Follow-up	Decision not to treat further	90D		

Inconsistent Element Durations Example: Figure 11

Trial with Inconsistent Element Durations

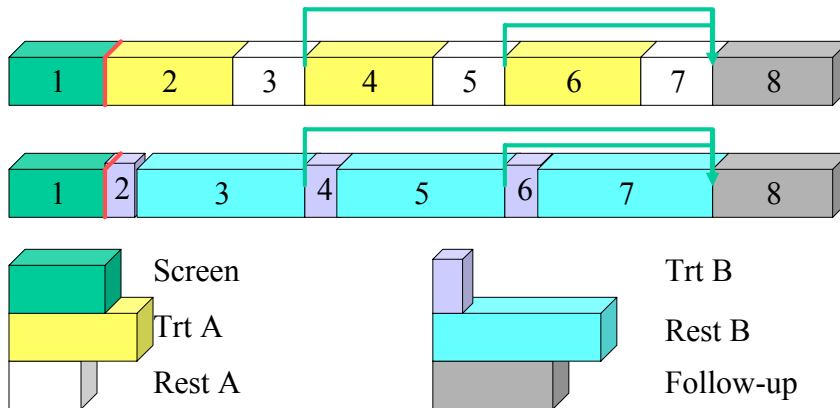


Figure 11

In the trial in Figure 11, the numbers of Elements are the same, and comparisons of the 2nd through 7th Elements of the two treatment Arms may be reasonable. The issue of whether Elements of different durations should be compared should be decided on the merits of the particular trial.

Dissimilar treatments: Figure 12

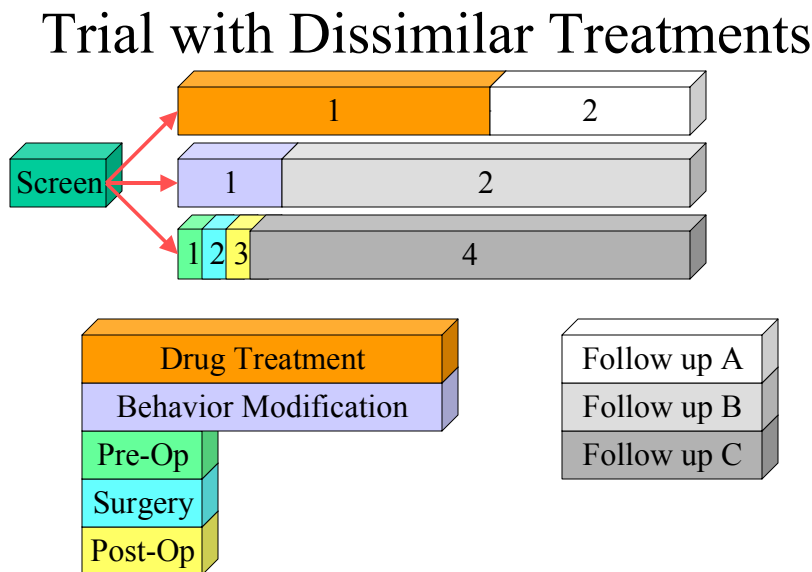


Figure 12

The trial shown in Figure 12 compares dissimilar treatments with different durations. The usefulness of comparisons between data from these different treatment Elements would depend on the particular trial. If such comparisons were desired, an alternate model that uses one Surgery treatment Element (combining Pre-Op, Surgery and Post-Op in one element), instead of the three Elements shown here, might be considered. See [Section 7.5.1](#) for additional discussion of 'lumping' versus 'splitting' in defining Elements.

7.4.5 Rules concept

The Branch and Transition columns shown in the example tables thus far are instances of a new concept being added to SDS in Version 3.1. A variable may have a Role of 'Rule.' The values of a Rule variable describe conditions for something happening. At the moment, values of Rule variables are text. At some point in the future, it is expected that these will become executable code. Other Rule variables are present in the Trial Elements and Trial Visits domains.

The value in the Branch rule gives conditions under which a subject was assigned to this Arm. A value in the Transition rule gives conditions under which a subject would move to an Element other than the next Element.

7.4.6 Recap of Trial Arms Variables

Variables	Purpose
STUDYID, DOMAIN	Identifiers
ARM, ARMCD	Name the Arm
TAETORD	Put Elements in order within the Arm
ETCD, ELEMENT	Name an Element within the Arm
TABRANCH	Mark a branch in the trial design at the end of the Element. A branch may be a randomization or another method of assignment to Arms

Variables	Purpose
TATRANS	A rule for where to go at the end of the Element. A blank represents the default rule, 'go to the next Element in sequence.'
EPOCH	Permissible. Names the Trial Epoch with which the Element is associated.

7.4.7 Truncated Arms

In trials with multiple branch points, such as that shown in Figure 2, some subjects may drop out of the study before they reach the second branch point. This raises the question how a value of ARM should be assigned to such a subject.

We recommended adopting names for the Arms of a multi-branching trial that reflect the various branches. For example, the four Arms in the trial shown in Figure 2 could be named A-Open, A-Rescue, B-Open, and B-Rescue. ARM values of A and B can be used for subjects who do not reach the second branch point.

Note that it may be necessary to create an Arm name for subjects who fail to reach the branch point even for a trial with only one Branch. For example, it might be necessary to submit data on subjects who drop out during the run-in Element in the trial in Figure 1.

7.5 TRIAL ELEMENTS

Trial Elements (TE) contains one record for each type of Element in TA. It is like a key or legend for the TA domain. Note that although the same value of ELEMENT may appear in more than one record in the TA dataset, it will appear in just one record in the TE dataset.

Examples: Figures 13, 14 and 15 show keys for Trial Elements added to the Trial Arms models shown previously in Figures 3, 5, and 8.

Trial Elements Corresponding to Figure 3: Figure 13

Arms and Elements, Parallel Trial

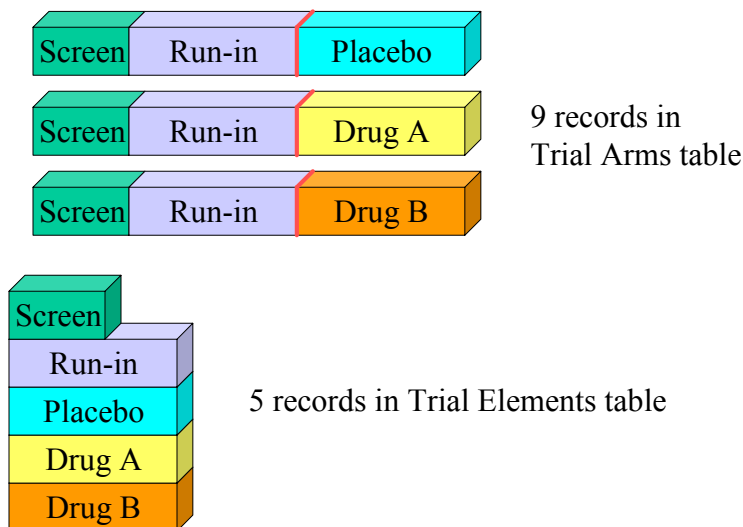


Figure 13

Trial Elements Corresponding to Figure 5: Figure 14

Arms and Elements, Crossover Trial

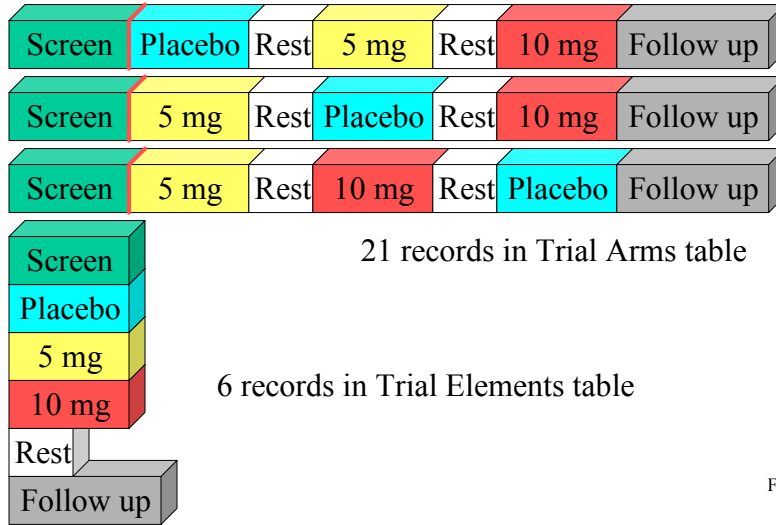


Figure 14

Trial Elements Corresponding to Figure 8: Figure 15

Arms and Elements, Oncology Trial

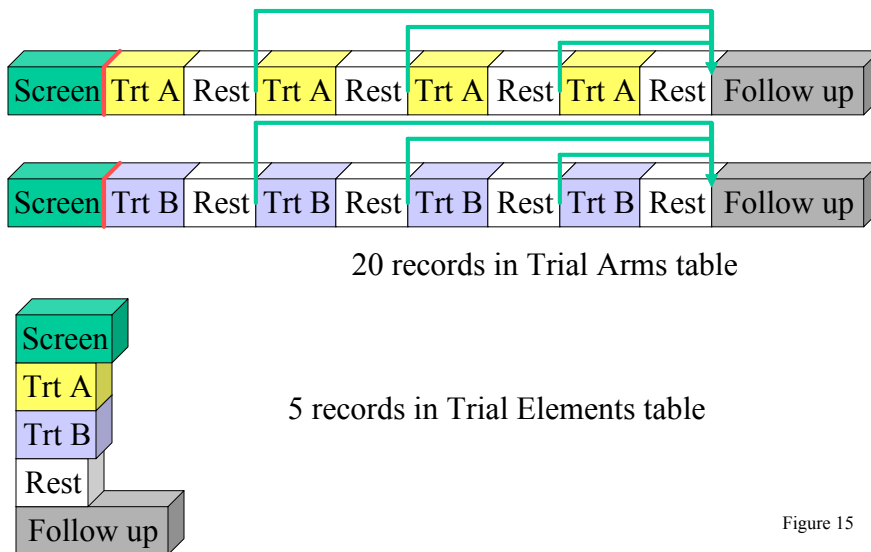


Figure 15

7.5.1 Identifying Trial Elements

An Element is a building block for creating Arms. Or, from another point of view, an Arm is composed of Elements (i.e., the trial design assigns subjects to Arms, which are comprised of steps called Elements). Trial Elements represent an interval of time in which certain activities are planned for the subjects participating in the trial. Both the activities and the time description are essential to the concept of the Element. The activities may be run-ins, wash-outs, drug treatments, or other interventions. The time description defines an interval of time, with a start and an end, during which the activities are planned to take place.

Assumption: There are no gaps between Elements. The instant one Element ends, the next Element begins. A subject spends no time 'between Elements.'

Since an Element is defined by both a period of time and what is happening to the subject during that time, 'Week 2 to Week 4' is not a valid Element. 'Two weeks of treatment with Drug A' is a valid Element. Elements may describe periods of non-treatment as well as periods of treatment. Common kinds of non-treatment Elements include wash-out, rest, and follow-up periods.

Deciding how finely to divide trial time when identifying trial Elements is a matter of judgment.

For example, a trial might include a dose titration, spelled out as one week each at a series of increasing doses until certain conditions are met. The trial design could be modeled in any of the following ways:

- using several one-week Elements at specific doses, followed by an Element of variable length at the chosen dose,
- as a titration Element of variable length followed by a constant dosing Element of variable length
- one Element with dosing determined by titration

The choice of model will depend on how the data will be analyzed and reported. If it is important to examine side effects or lab values at each individual dose, the first model is appropriate. If it is important only to identify the time to achieve titration, the second model might be appropriate. If the titration process is routine and of little interest, the third model might be adequate for the purposes of this trial.

Oncology trials with cycles of treatment and rest present another example where the needs of the trial will determine whether the trial should be modeled with two Elements per cycle, as in Figure 8 (previously shown), or with one Element per cycle, as in Figure 16 below.

Oncology Trial, One Element per Cycle: Figure 16

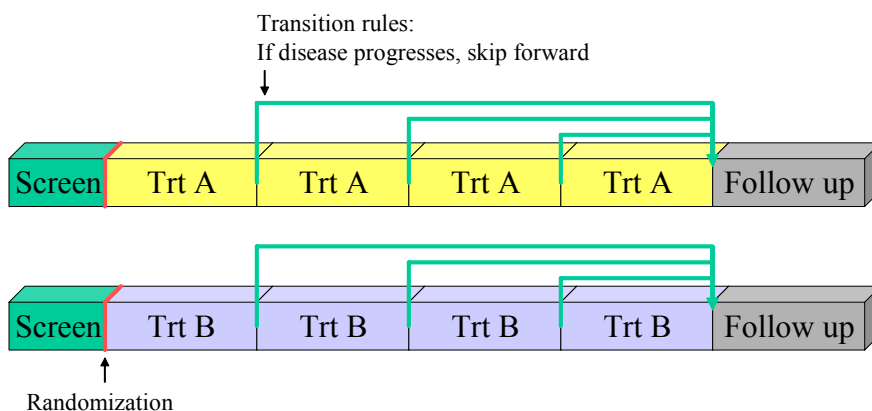


Figure 16

7.5.2 Developing the Trial Elements Table

An Element's name usually describes what happens to the subject during the Element. The time period associated with an Element is expressed in two Rules, the Element Start and Element End Rules. If the element is of fixed duration, then the Element Duration variable should also be populated.

Element Start Rules identify the time point at which a subject is considered to have entered an Element. Many analyses depend on placing data in the proper Element, so the Element Start Rules should be as clear as possible, and should be chosen with consideration for planned analyses.

Treatment Elements are usually considered to start with the beginning of the Element treatment (e.g., with the first dose of the relevant drug). The Start Rules for non-treatment Elements can be more difficult to define. They will often be defined relative to the end of the preceding treatment. For example, a wash-out period might be defined as starting 24 or 48 hours after the last dose of drug for the preceding treatment period. Defining a clear start point for the start of a non-treatment Element that follows another non-treatment Element can be difficult.

The Element Start Rule for a treatment Element may be thought of as 'active' and the Element Start Rule for a non-treatment Element as 'passive.' The start of a treatment Element will not occur until a dose is given, no matter how long that dose is delayed. Once the last dose is given, the start of the subsequent non-treatment Element is inevitable, as long as another dose is not given.

Element End Rules are rather different from Element Start Rules. The actual end of one Element is the beginning of the next Element. Thus the Element End Rule does not give the conditions under which an Element does end, but the conditions under which it should end. The most common type of Element End Rule is the rule for an Element of fixed duration. Such a rule might be expressed as '14 days after Element starts' or, perhaps more realistically, as '12-16 days after Element starts.' Element End Rules may also depend on other conditions. For instance, a typical criterion for ending a rest Element between oncology chemotherapy treatment Elements would be, '15 days after start of Element and after WBC values have recovered'.

There is not yet a standardized method for expressing rules in computer-executable form. However, for the most common type of Element End Rule, for elements of fixed duration, we can easily express element duration. Thus, a variable for Element Duration has been included in the model.

Example: Trial Elements table for the parallel trial shown in Figure 13

ELEMENT	TESTRL	TEENRL	TEDUR
Screen	Informed consent	1 week after start of element	P7D
Run-In	Eligibility confirmed	2 weeks after start of element	P14D
Placebo	First dose of placebo	2 weeks after start of element	P14D
Drug A	First dose of Drug A	2 weeks after start of element	P14D
Drug B	First dose of Drug B	2 weeks after start of element	P14D

Example: Trial Elements table for the crossover trial shown in Figure 14

ELEMENT	TESTRL	TEENRL	TEDUR
Screen	Informed consent	2 weeks after start of element	P14D
Placebo	First dose of placebo	2 weeks after start of element	P14D
5 mg	First dose of 5 mg drug	2 weeks after start of element	P14D
10 mg	First dose of 10 mg drug	2 weeks after start of element	P14D
Rest	48 hrs after last dose drug	1 week after start of element	P7D
Follow-up	48 hrs after last dose drug	3 weeks after start of element	P21D

The transition between one Element and the next can be complex, since it may involve meeting criteria for ending one Element, evaluating criteria for choosing which Element to go to next, and actually taking the action that initiates the new Element.

For example, in the trial shown in Figure 15, the Trial Elements dataset might be as follows:

ELEMENT	TESTRL	TEENRL	TEDUR
Screen	Consent	2 weeks after start of element	P14D
Trt A	First dose drug	5 days after start of element	P5D
Trt B	First dose drug	16 days after start of element and WBC recovers	
Rest	Last dose + 24 hrs	16 days after start of element and WBC recovers	
Follow-up	Decision not to treat further	4 weeks	P28D

A subject in the 3rd Element (Trt B) must meet the Element End criteria (16 days pass and WBC recovers), then be evaluated to see whether disease has progressed. If it has not progressed, they must still wait until the first dose of the 4th Element is administered before they can be considered to be in the 4th Element. During the whole time from meeting the Element End criteria to meeting the new Element Start criteria, the subject stays in the 'old' Element.

See [Section 7.7.2](#) for a discussion of unplanned Elements.

7.5.3 Recap of Trial Elements Variables

Variables	Purpose
STUDYID, DOMAIN	Identifiers
ETCD, ELEMENT	Name the Element
TESTRL	Rule describing the time point that defines the start of the Element
TEENRL, TEDUR	Rule describing when the Element should end and its duration (if the element is of fixed duration)

7.5.4 Distinguishing Elements from Epochs

Figure 9, which was used to introduce the concept of trial Epoch is also useful for showing how the concepts of Elements and Epoch can be confused. Note that the natural name for the first Epoch is 'Screen,' which is also the natural name of the Element that appears in each Arm in this Epoch. The same is true for the 7th Epoch, where the Epoch and the Elements in the Epoch are both named 'Follow-up.' Clearly, it will be easy to confuse Element and Epoch in these cases.

We recommend avoiding using the same names for Elements and Epochs. One possible solution is to prefix Epoch names with the word 'trial', on the grounds that an Epoch is a feature of the trial as a whole. In the example, 'Trial Screen' would be the Epoch, while 'Screen' would be the Element

In the treatment Epochs of this trial, the Epoch describes a period of time in the trial, First, Second, or Third Treatment, while the Elements describe particular treatments which may occur during different Epochs of the trial.

This trial has First and Second Rest Epochs, which are populated by the one Rest Element in this trial.

It may be useful to think of a trial design as a table like the following, with a row for each Arm and a column for each Epoch.

	Trial Screen	First Treatment	First Rest	Second Treatment	Second Rest	Third Treatment	Trial Follow-up
	1	2	3	4	5	6	7
P-5-10	Screen	Placebo	Rest	5 mg	Rest	10 mg	Follow-up
5-P-10	Screen	5 mg	Rest	Placebo	Rest	10 mg	Follow-up
5-10-P	Screen	5 mg	Rest	10 mg	Rest	Placebo	Follow-up

The bold column headers are the names of the Epochs. The secondary column headers (1, 2, etc.) show how each value of TAETORD in this trial corresponds to an Epoch. The bold row labels are the names of the Arms. The cells are filled with the names of the Elements in the trial Arms.

For some analysis purposes, one would want to group data by Element, but for other purposes, data would be grouped by Epoch or TAETORD. In this crossover trial, for instance, comparisons of incidences of side effects of the different treatments would group data by Element (Placebo vs. 5 mg vs. 10 mg). The analysis of PK results for this trial would include both treatment and period effects in the model. 'Treatment effect' derives from Element (Placebo vs. 5 mg vs. 10 mg), while 'Period effect' derives from the EPOCH (1st, 2nd, or 3rd Treatment), which is in turn associated with TAETORD (2nd vs. 4th vs. 6th Element).

7.6 TRIAL VISITS

Trial Visits (TV) describes the planned visits in a trial.

7.6.1 Identifying Trial Visits

The SDS definition of Visit is a clinical encounter. Assessments and procedures are planned at Visits.

Assumption: Visits have both a start and an end.

For many trials, the exact timing and duration of Visits are not important, but for other trials this detail is needed, so both start and end are included in the model.

Assumption: A Visit may start in one Element and end in another.

Example: Visits in clinical pharmacology trials are often planned to include a series of timed assessments (a profile) associated with a dose of drug. The dose is often the start of an Element. The planned Visit might start a hour before the dose (the end of the current Element) and end eight hours after the dose (the beginning of the new Element), to allow for pre-dose assessments and a series of post-dose assessments.

The term 'Visit' reflects the fact that data in out-patient studies is usually collected when the subject Visits the clinic. Diary data and other data collected outside the clinic may not fit the usual concept of a Visit, but the planned times of collection of such data may be described as 'Visits' in the Trial Visits dataset.

The description of planned 'Visits' may also require careful thought for studies of hospitalized subjects.

7.6.2 Developing the Trial Visits Table

For most trials, particularly blinded trials, the timing of Visits does not depend on Arm. However, planned Visits may be different for different Arms in the trial, so ARM and ARMCD are included in the Trial Visits dataset. If planned Visits do not depend on Arm, ARM and ARMCD can be left blank. If planned Visits do depend on Arm, records for planned Visits for each Arm should be submitted. The trial shown in Figure 12 would probably have different planned Visits for each Arm.

Visit Start Rules are different from Element Start Rules because they usually describe when a Visit should occur, rather than the moment at which a Visit is considered to start. There are usually gaps between Visits, periods of time that do not belong to any Visit, so it usually not necessary to identify the moment when one Visit stops and another starts. However, some trials of hospitalized subjects may divide time into Visits in a manner more like that used for Elements.

Visit Start Rules are usually expressed relative to the start or end of an Element, e.g., '1-2 hours before end of 3rd Element' or '8 weeks after end of 2nd Element.' Note that the Visit may or may not occur during the Element used as the reference for Visit Start Rule. For example, a trial with variable Elements might plan a Visit 6 months after the start of the first treatment period, regardless of the Element the subject is in at that time.

Visit End Rules are similar to Element End Rules, describing when a Visit should end. They may be expressed relative to the start or end of an Element, or relative to the start of the Visit.

Ranges may be used to describe the planned timing of Visits (e.g., 12-16 days after the start of 2nd Element), but this is different from the 'windows' that may be used in selecting data points to be included in an analysis associated with that Visit. For example, although Visit 2 was planned for 12-16 days after the start of treatment, data collected 10-18 days after the start of treatment might be included in a 'Visit 1' analysis. The two ranges serve different purposes.

Unplanned Visits are discussed further in [Section 7.8](#).

Trial Visits Example: Figure 17

Trial Visits

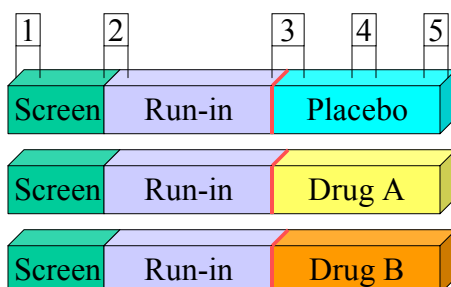


Figure 17

The trial shown in Figure 17 might have the following Trial Visits dataset:

VISITNUM	TVSTRL	TVENRL
1	Start of 1 st Element	1 hour after start of visit
2	30 minutes before end of 1 st Element	30 minutes after start of 2 nd Element
3	30 minute before end of 2 nd Element	1 hour after start of 3 rd Element
4	1 week after start of 3 rd Element	1 hour after start of visit
5	2 weeks after start of 3 rd Element	1 hour after start of visit

7.6.3 Recap of Trial Visits Variables

Variables	Purpose
STUDYID, DOMAIN	Identifiers
VISIT, VISITNUM, VISITDY	Name and order the visits
ARM, ARMCD	Blank if visit schedule does not depend on Arm. Names of the ARM if visit schedule does depend on Arm.
TVSTRL	Rule describing when the visit should start. Usually expressed relative to starts or ends of Elements described using TAETORD.
TVENRL	Rule describing when the visit should end. Usually expressed relative to the start of the Element, or to starts or ends of Elements described using TAETORD.

7.7 SUBJECT ELEMENTS

Subject Elements (SE) describes the sequence of Elements a subject passed through, with their start and stop date/times. Note there is no need for a Subject Arms dataset, because the Arm for each subject is represented in the Demography (DM) dataset.

The subject Element start date/times allow a comparison of the sequence of Elements a subject actually passed through to the planned sequence of Elements for the Arm to which they were assigned.

The date/times in Subject Elements also allows any date/time to be placed within one Element. This assignment of data to actual Elements is often used in analyses.

Subject Elements data involves judgment and interpretation.

7.7.1 Identifying Subject Elements

The Element Start Rules in the Trial Elements dataset should allow identification of the Elements the subject passed through and the derivation of start date/times for each subject Element. This derivation may be more or less difficult, depending on the trial design, on how well the Element Start Rules have been defined, and on how well the date/times associated with the transition from one Element to the next have been collected.

Note that this process of matching the subject's actual experience with planned Elements involves judgment, and may be open to interpretation. The programming used to derive subject Elements will incorporate decisions about how close a match a time period and its treatments have to be to the plan to be considered a 'match.' For instance if the plan included Elements with treatment doses of 5 mg and 10 mg, is a period of treatment with a dose of 7 mg a '5mg' Element? If an Element is planned to last 2 weeks but actually lasts 6 weeks, is this a match? It is usually better to associate subject experience with one of the planned Elements, than to describe, note or summarize deviations from the plan, but this is a matter of judgment.

7.7.2 Unplanned Elements

It is usually better to parse a subject's time in the trial into segments labeled with names from the list of planned Elements, but occasionally a subject departs so much from plan that a decision is made to label a part of the subject's experience as an unplanned Element.

Example: A crossover trial that involves three single doses of drug, one each week for three weeks. The period from one dose to the next is considered a treatment Element. A subject is enrolled in the trial and completes two of the doses. He departs for a month-long trip. On his return, he is given the third dose, then returns for follow-up a week later. It is decided that only the first week after the second dose should be considered as belonging to the second treatment Element. The remaining 3 weeks before the third dose is labeled an 'Unplanned' Element.

Example: A trial involves treatment with an IV drug that is prepared in the hospital pharmacy. By error, the subject receives only one tenth of the planned dose. The error is discovered, and the subject is treated again, with the correct dose. It is decided to label the period starting with the initiation of the erroneous infusion and ending the initiation of the correct infusion an 'unplanned' Element.

If an unplanned Element occurs, a record for the unplanned Element should be included in SE. For this record, ETCD cannot be populated with any of the values associated with planned Elements.

It is recommended that the SE record for the unplanned Element use 'UNPLAN' as the value for ETCD, that ELEMENT be left blank, and that a description of the unplanned Element be given in SEUPDES.

7.7.3 Deriving SE End Date/Times

Because there are no gaps between Elements, once subject Element start date/times have been derived for all Elements, the derivation of all but one of the subject end date/times is trivial. For all subject Elements except the last, the end date/time of the Element is equal to the start date time of the next Element.

It may be desirable to define RFENDTC (in the DM domain) as the end date/time for the end of the last subject element date/time.

7.7.4 Recap of Subject Elements Variables

Variables	Purpose
STUDYID, DOMAIN, USUBJID	Identifiers
ETCD, ELEMENT	Name this Element in the subject's course
SESTDTC, SEENDTC	Date/times of start and end of the Element for the subject
SEUPDES	Permissible variable. Blank except for unplanned Elements. Describes the unplanned Element.

7.7.5 Using Subject Elements Data to Place Subject Data within an Element or Epoch

SDS V3 contained deprecated Period, Phase, and Cycle variables. The date/times in SE allow any date/time to be placed in one subject Element, or before the first subject Element (before the trial) or after the last subject Element (after the trial). Thus, if the subject Elements dataset is submitted, the deprecated period, phase, and cycle variables are not needed.

An observation date/time may be compared to the date/times in the Subject Elements dataset in order to place that observation in the proper element. However, if the date/time of the observation is equal to the end date/time of one element and the start date/time of the following element, then it is not immediately clear whether the observation should be assigned to the element that is ending or to the element that is starting. Thus Sponsors should choose a convention for assigning observations to one element or the other, and document the convention. It will be particularly important to specify this convention if date/times are collected only to the nearest day, when "ties" of this kind will be fairly common.

The sequence of subject Elements may be compared to the planned sequence of Elements for the Arm to which the subject was assigned. If the sequences match (or if the actual sequence is a truncated version of the planned sequence, as when a subject withdraws early) then it makes sense to associate a value of TAETORD with each subject Element. For trials with Epochs, values of EPOCH may also be associated with subject Elements.

If a subject's planned and actual sequences of Elements do not match, then it may not make sense to associate values of TAETORD or EPOCH with the subject Elements, depending on the use that is being made of TAETORD or EPOCH.

7.8 SUBJECT VISITS

Subject Visits (SV) describes the visits a subject passed through, with their start and end date/times.

7.8.1 Identifying Subject Visits

Visit start and end date/times are usually collected directly. In many studies, only a single visit date is collected for each visit. This single date can be used to populate both start and end date/time.

Unplanned Visits commonly occur in clinical trials, and records for unplanned visits should be included in SV.

Mechanisms for labeling unplanned visits are in common use. SDS is not recommending particular conventions for populating VISIT, VISITNUM, and VISITDY for unplanned visits at this time.

7.8.2 Recap of Subject Visits Variables

Variables	Purpose
STUDYID, DOMAIN, USUBJID	Identifiers
VISIT, VISITNUM, VISITDY	Variables that name and order the visits
SVSTDTC, SVENDTC	Date/times of start and end of this visit for this subject
SVUPDES	Permissible variable. Blank except for unplanned Visits. Describes the unplanned Visit.

7.9 TRIAL INCLUSION/EXCLUSION CRITERIA

The Trial Inclusion Exclusion Domain (TI) is not subject oriented. It contains all the inclusion and exclusion criteria for the trial, and thus provides information that may not be present in the subject-level data on inclusion and exclusion criteria. The IE Domain (described in [Section 6.3.2](#)) contains records only for inclusion and exclusion criteria that a subject did not meet, and thus includes only criteria which one or more subjects failed to meet.

Inclusion and Exclusion Criteria are rules, so the TI domain includes both the familiar text version of a criterion, IETEST, and a Rule version, TIRL.

ti.xpt, Trial Inclusion/Exclusion Criteria, Version 3.1.1. August 26, 2005. One record per I/E criterion

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		CRF	Identifier	Unique identifier for a study within the submission.	Req
DOMAIN	Domain Abbreviation	Char	**TI	Derived	Identifier	Two-character abbreviation for the domain.	Req
IETESTCD	Inclusion/Exclusion Criterion Short Name	Char	*	Sponsor Defined	Topic	Short name IETEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in IETESTCD cannot be longer than 8 characters, nor can it start with a number (e.g., '1TEST'). IETESTCD cannot contain characters other than letters, numbers, or underscores. The name "IE" prefix is used to ensure consistency with the IE domain.	Req
IETEST	Inclusion/Exclusion Criterion	Char	*	Sponsor Defined / Protocol	Synonym Qualifier	Full text of the inclusion or exclusion criterion. The prefix "IE" is used to ensure consistency with the IE domain.	Req
IECAT	Inclusion//Exclusion Category	Char	*INCLUSION, EXCLUSION	Sponsor Defined	Grouping Qualifier	Used to categorization of the inclusion or exclusion criteria.	Req
TIRL	Inclusion/Exclusion Criterion Rule	Char		Sponsor Defined	Rule	Optional rule that expresses the criterion in computer-executable form.	Perm

7.10 TRIAL SUMMARY INFORMATION

The Trial Summary Model allows the sponsors to submit a summary of the trial in a structured format. The Trial Summary dataset contains values for each characteristic described for a trial. Trial Summary is used to record basic information about the study such as trial phase, protocol title, and trial objectives. The Trial Summary domain contains information about the *planned* trial characteristics such as length of trial, planned number of subjects, planned maximum age of subject and others. The Trial Summary (TS) domain does not contain subject level data or data that can be derived from subject data. Thus, for example, it includes the number of subjects planned for the trial but not the number of subjects enrolled in the trial.

ts.xpt, Trial Summary, Version 3.1.1, August 26, 2005. One record per trial summary parameter

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Derived	Identifier	Unique identifier for a study within the submission.	Req
DOMAIN	Domain Abbreviation	Char	**TS	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation.	Req
TSSEQ	Sequence Number	Num		Derived	Identifier	Sequence number given to ensure uniqueness within a dataset. Allows inclusion of multiple records for the same parm, and can be used to join related records.	Req
TSPARMCD	Trial Summary Parameter Short Name	Char	**	Derived	Topic	Short name for the Parameter described in TSPARM. The value in TSPARMCD cannot be longer than 8 characters, nor can it start with a number. TSPARMCD cannot contain characters other than letters, numbers, or underscores. Examples: DESIGN, MASK, COMPTRT	Req
TSPARM	Trial Summary Parameter	Char	**	Derived	Synonym Qualifier	Term for the Trial Summary Parameter. The value in TSPARM cannot be longer than 40 characters. Examples: Description Of Trial Design, Trial Blinding Schema, Comparative Name Of Treatment	Req
TSVAL	Parameter Value	Char	**	Derived	Result Qualifier	Value of TSPARM. Example: 'ASTHMA' when TSPARM value is 'Trial Indications'. TSVAL cannot be null – a value is required for the record to be valid. The first 200 characters of TSVAL will be in TSVAL, then next 200 in TSVAL1, and continuing as needed to TSVALn.	Req

* indicates variable may be subject to sponsor-controlled terminology; ** indicates variable may be subject to external controlled terminology.

7.10.1.1 Assumptions for Trial Summary (TS) domain model

1. The intent of the domain model is to provide a summary of the trial information. This is not subject level data.
2. TSVAL may have controlled terminology depending on the value of TSPARMCD, see [appendix 10.3.5](#) "Trial Summary" for more information.
3. Trial Summary is similar to the COMMENTS special-purpose domain. It allows for one value to span multiple variables (TSVAL-TSVALn) in order to accommodate values longer than 200 characters.¹
4. The TSPARMCD value 'PHASE' is not defined with any exactness since it can vary by drug class.
5. For LENGTH OF TRIAL is defined as the planned length of time for a subject's participation. For "LENGTH OF TRIAL", units for duration are not needed. See Section [4.1.4.3](#).
6. For some trials, there will be multiple records in the Trial Summary dataset for a single parameter such as "TYPE OF TRIAL" since a trial could be for both the Safety and Efficacy parameters. In this case, when TSPARMCD=TYPE there will be two records for TSVAL, one with the value of 'SAFETY' and the other with the value of 'EFFICACY'.

¹ Allowing for multiple variables per comment is an interim solution until the limitations posed by SAS® Version 5 transport files are eliminated.

7.11 HOW TO MODEL THE DESIGN OF A CLINICAL TRIAL

The following steps allow the modeler to move from familiar concepts, such as Arms, to less familiar concepts, such as Elements and Epochs. The actual process of modeling a trial may depart from these numbered steps. Some steps will overlap, there may be several iterations, and not all steps are relevant for all studies.

1. Start from the flow chart or schema diagram usually included in the trial protocol. This diagram will show how many arms the trial has, and the branch points, or decision points, where the arms diverge.
2. Write down the decision rule for each branching point in the diagram. Does the assignment of a subject to an arm depend on a randomization? On whether the subject responded to treatment? On some other criterion?
3. If the trial has multiple branching points, check whether all the branches that have been identified really lead to different Arms. The Arms will relate to the major comparisons the trial is designed to address. For some trials, there may be a group of somewhat different paths through the trial that are all considered to belong to a single Arm.
4. For each arm, identify the major time periods of treatment and non-treatment a subject assigned to that arm will go through. These are the Elements, or building blocks, of which the Arm is composed.
5. Define the starting point of each Element. Define the rule for how long the Element should last. Determine whether the element is of fixed duration.
6. Re-examine the sequences of Elements that make up the various Arms and consider alternative Element definitions. Would it be better to 'split' some Elements into smaller pieces or 'lump' some Elements into larger pieces? Such decisions will depend on the aims of the trial and plans for analysis.
7. Compare the various Arms. In most clinical trials, especially blinded trials, the pattern of Elements will be similar for all Arms, and it will make sense to define Trial Epochs. Assign names to these Epochs. During the conduct of a blinded trial, it will not be known which Arm a subject has been assigned to, or which treatment Elements they are experiencing, but the Epochs they are passing through will be known.
8. Identify the Visits planned for the trial. Define the planned start timings for each visit, expressed relative to the ordered sequences of Elements that make up the Arms. Define the rules for when each visit should end.
9. Identify the inclusion and exclusion criteria to be able to populate the TI domain.

8 Representing Data Relationships

Although each domain in a regulatory submission is defined to stand alone as a separate dataset, there are many occasions when it is necessary or desirable to represent relationships among datasets or records, such as relationships between separate domains, or between domains and other special-purpose datasets. In some cases, this is necessary because the fixed structure of the SDTM observation classes restricts the ability of sponsors to represent data that do not entirely fit within those structures. The SDTM accommodates five distinct types of relationships, which are described in more detail in subsequent sections.

- [Section 8.1](#) describes a relationship between a group of records for a given subject within the same domain.
- [Section 8.2](#) describes a relationship between independent records (usually in separate domains) for a subject, such as a concomitant medication taken to treat an adverse event.
- [Section 8.3](#) describes a relationship between two (or more) datasets where all the records of one (or more) dataset(s) have parent or counterpart record(s) in another dataset (or datasets).
- [Section 8.4](#) describes a dependent relationship between data that cannot be represented by a standard variable in the general class and a parent record (or records) within a domain.
- [Section 8.5](#) describes a dependent relationship between a comment and a parent record (or records) in other domains, such as a comment recorded with an adverse event.
- [Section 8.6](#) discusses the concept of related datasets and whether to place additional data in a domain or supplemental qualifier table

See [appendix 10.6](#) for a proposed approach for modeling findings data that may refer to interventions or events information.

All relationships make use of the standard domain identifiers, STUDYID, DOMAIN, and USUBJID. In addition, the variables IDVAR and IDVARVAL are essential for identifying the record-level merge/join keys. The specific set of identifiers necessary to properly identify each type of relationship is described in detail in the following sections.

The Sequence Number (--SEQ) variable uniquely identifies a record for a given USUBJID within a domain. --SEQ is required in all domains except DM. For example, if subject 1234-2003 has 25 adverse event records in the adverse event (AE) domain, AESEQ values for this subject should be the numbers 1 to 25, with numbering starting over with 1 for the next subject. Numbers may, however, be omitted from the sequence if the sponsor assigns the numbers early in the process and subsequently deletes some records.

The Reference Identifier (--REFID) variable can be used to capture a sponsor-defined or external identifier, such as an identifier provided in an electronic data transfer. Some examples are lab-specimen identifiers, ECG identifiers, or image filename identifier (sometimes used to reference external files such as medical images or waveform ECGs). --REFID is permissible in all domains, but never required. Values for --REFID are sponsor defined and can be any alphanumeric strings the sponsor chooses, consistent with their internal practices.

The Group Identifier (--GRPID) variable is explained below in [section 8.1](#).

8.1 RELATING GROUPS OF RECORDS WITHIN A DOMAIN

The optional grouping identifier variable --GRPID is permissible in all domains that are based on the general observation classes to identify relationships between records within a domain (such as Intervention records for a combination therapy). The relationship is defined by assigning the same unique (within USUBJID) character value to the --GRPID variable. All records for a subject in a domain are considered to be related when they have the same --GRPID value. The --GRPID values are only meaningful within the domain, and in the Relate Records (RELREC), Supplemental Qualifiers (SUPPQUAL), and Comments (COMMENTS) datasets, as described in subsequent sections. The values used for --GRPID can be any values the sponsor chooses; however, the philosophy for assigning values should be consistent across the submission to avoid confusion. For example, --GRPID could be any or all of the following:

- The therapy name for multiple exposure records associated with a combination therapy
- A sponsor-defined term used to group sets of questions together in the Questionnaire (QS) domain
- Any sponsor-defined constant value used to group multiple concomitant medications together to indicate they were all taken to treat the same condition.

Using --GRPID in the SDS Domains can be helpful to reduce the number of records in the RELREC, SUPPQUAL, and COMMENTS datasets when it is necessary to use those datasets to capture/identify relationships/associations for records or values to a 'group' of parent/peer records.

8.1.1 --GRPID Example

The following table illustrates how to use --GRPID in the Concomitant Medications (CM) domain to identify a combination therapy. Note that the example only includes the most relevant variables from CM for this example; the complete CM domain might include other variables not shown below. In this example, subject 1234 has reported two combination therapies, each consisting of three separate medications. Each component of a combination is given the same value for GRPID.

STUDYID	DOMAIN	USUBJID	CMSEQ	CMGRPID	CMTRT	CMDECOD	CMDOSE	CMDOSU	CMSTDTC	CMENDTC
1234	CM	1234	1	COMBO THPY 1	Verbatim Med A	Generic Med A	100	mg	2004-01-17	2004-01-19
1234	CM	1234	2	COMBO THPY 1	Verbatim Med B	Generic Med B	50	mg	2004-01-17	2004-01-19
1234	CM	1234	3	COMBO THPY 1	Verbatim Med C	Generic Med C	200	mg	2004-01-17	2004-01-19
1234	CM	1234	4	COMBO THPY 2	Verbatim Med D	Generic Med D	150	mg	2004-01-21	2004-01-22
1234	CM	1234	5	COMBO THPY 2	Verbatim Med E	Generic Med E	100	mg	2004-01-21	2004-01-22
1234	CM	1234	6	COMBO THPY 2	Verbatim Med F	Generic Med F	75	mg	2004-01-21	2004-01-22

8.2 RELATING PEER RECORDS IN SEPARATE DATASETS

The Related Records (RELREC) dataset is used to describe relationships between records in two (or more) datasets, such as an Event record and an Intervention record, or a Finding record and an Event record.

A relationship is defined by creating RELREC records for each of the records to be related and by assigning a unique character identifier value for the relationship. Each RELREC record contains keys that identify a record (or group of records) and the relationship identifier, which is stored in the RELID variable. The value of RELID can be any constant value chosen by the sponsor.

Records expressing a relationship are specified using the keys STUDYID, RDOMAIN, and USUBJID, which are populated with the values for STUDYID, DOMAIN, and USUBJID from the SDS domain records, along with an IDVAR and an IDVARVAL. Single records can be related by using --SEQ in IDVAR. Groups of records can be specified by using --GRPID in IDVAR and the appropriate --GRPID value in IDVARVAL. Using the optional grouping identifier variable --GRPID (see [section 8.1](#)) to group a set of related records in the domains can be a more efficient method of representing relationships in RELREC, such as when relating an adverse event (or events) to a “group” of concomitant medications taken to treat the adverse event(s).

For example, suppose adverse events are determined at the investigational site to be associated with the administration of a concomitant medication used in standard care. Since the adverse events and the concomitant medication are captured in separate domains, a separate record in the RELREC dataset must be used to describe the observed association among these records.

The RELREC dataset should be used only when the sponsor either:

1. Collects information about explicit relationships, such as concomitant medications taken as a result of an adverse event
2. Determines there is a scientific rationale or reporting need to identify relationships after data collection such as determining that an adverse event was due to a concomitant medication, or
3. Collects information of a nature that necessitates using multiple datasets, as described in [section 8.3](#).

This is not to imply that a sponsor should retrospectively identify all potential relationships in the RELREC dataset; instead, the purpose of the RELREC dataset is to give sponsors a standardized way to represent such relationships when a sponsor identifies a specific need to do so. The RELREC dataset should only be used to associate entire observations on a one-to-one or one-to-many level. Additional Qualifier variables that cannot be represented in existing variables in Findings, Interventions, or Events domains should be represented using the Supplemental Qualifiers (SUPQUAL) dataset, which is [described in Section 8.4](#).

8.2.1 RELREC Dataset

relrec.xpt, Related Records, Version 3.1.1, August 26, 2005. One record per record or domain relationship

Variable	Variable Label	Type	CDISC Notes
STUDYID	Study Identifier	Char	Study Identifier of the SDS domain record(s).
RDOMAIN	Related Domain Abbreviation	Char*	Domain Abbreviation of the SDS domain record(s).
USUBJID	Unique Subject Identifier	Char	Unique Subject Identifier of the SDS domain record(s).
IDVAR	Identifying Variable	Char**	Value of the identifying variable in the general class dataset that identifies the related record(s). Examples include --SEQ and --GRPID.
IDVARVAL	Identifying Variable Value	Char	Value of identifying variable described in IDVAR. If --SEQ is the variable being used to describe this record, then the value of --SEQ would be entered here.
RELTYPE	Relationship Type	Char**	Identifies the hierarchical level of the records in the relationship. Values should be either ONE or MANY. However, values are only necessary when identifying a relationship between datasets (as described in Section 8.3).
RELID	Relationship Identifier	Char	Unique value within USUBJID that identifies the relationship. All records for the same USUBJID that have the same RELID are considered 'related/associated.' RELID can be any value the sponsor chooses, and is only meaningful within the RELREC dataset to identify the related/associated Domain records.

* indicates variable may be subject to sponsor-controlled terminology

** indicates variable may be subject to external controlled terminology.

8.2.2 RELREC Dataset Examples

Example A:

Example A shows how to use the RELREC dataset when USUBJID 123456 had two lab tests (LBSEQ values of 47 and 48) and took two concomitant medications (CMSEQ values of 11 and 12) as the result of an adverse event (AESEQ value of 5). This example assumes that the lab values and the concomitant medications are not related to each other.

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
EFC1234	AE	123456	AESEQ	5		1
EFC1234	CM	123456	CMSEQ	11		1
EFC1234	CM	123456	CMSEQ	12		1
EFC1234	AE	123456	AESEQ	5		2
EFC1234	LB	123456	LBSEQ	47		2
EFC1234	LB	123456	LBSEQ	48		2

NOTE: The two lab tests could have the same LBGRPID if they were on the same specimen. In such a case, the two RELREC records with RDOMAIN=LB could be replaced with one RELREC record with IDVAR=LBGRPID and IDVARVAL having the LBGRPID value assigned in the LB domain.

Example B:

Example B is the same scenario as Example A; however, the concomitant medications and lab tests are associated with each other in addition to being associated with the adverse event.

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
EFC1234	AE	123456	AESEQ	5		1
EFC1234	CM	123456	CMSEQ	11		1
EFC1234	CM	123456	CMSEQ	12		1
EFC1234	LB	123456	LBSEQ	47		1
EFC1234	LB	123456	LBSEQ	48		1

Example C:

Example C is the same scenario as Example B, however, both the concomitant medications are necessary for this treatment and have been assigned a CMGRPID of 'COMBO 1' by the sponsor, allowing the elimination of a record in the RELREC dataset.

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
EFC1234	AE	123456	AESEQ	5		1
EFC1234	CM	123456	CMGRPID	COMBO1		1
EFC1234	LB	123456	LBSEQ	47		1
EFC1234	LB	123456	LBSEQ	48		1

8.3 RELATING DATASETS

The Related Records (RELREC) dataset can also be used to identify relationships between datasets (e.g., a one-to-many or parent-child relationship, as discussed in [Section 8.3.1](#)). The relationship is defined by including a single record for each related dataset that identifies the key(s) of the dataset that can be used to relate the respective records.

Relationships between datasets should only be recorded in the RELREC dataset when the sponsor has found it necessary to split information between datasets that are related, and that may need to be examined together for analysis or proper interpretation. It is not necessary to use the RELREC dataset to identify associations from data in the SUPPQUAL dataset or the COMMENTS dataset, as both these datasets include the key variable identifiers of the parent record that are necessary to make the association.

8.3.1 RELREC Dataset Relationship Example

This example shows how to use the RELREC dataset to represent related information that is submitted as two datasets which have a one-to-many relationship. In the example below all the records in one domain are being related to all of the records in the other so both USUBJID and IDVAR are null.

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
EFC1234	MB		MBSPID		ONE	A
EFC1234	MS		MSSPID		MANY	A

In the sponsor's operational database, these datasets may have existed as either separate datasets that were merged for analysis, or one dataset that may have included observations from more than one observation class (e.g., Events and Findings). The value in IDVAR should be the name of the key used to merge/join the two datasets. In the above example, the --SPID variable is used as the single key to identify the related observations. The values for the --SPID

variable in the two datasets are sponsor defined. Although other variables may also serve as a single merge key when the corresponding values for IDVAR are equal, --SPID or --REFID are typically used for this purpose.

Since IDVAR identifies the keys that can be used to merge/join all records between the datasets, the root values (i.e., --SPID in the above example) for IDVAR must be the same for both records with the same RELID. --SEQ cannot be used because --SEQ only has meaning within a dataset, not across datasets.

8.4 RELATING NON-STANDARD VARIABLE VALUES TO A PARENT DOMAIN

The Supplemental Qualifiers (SUPPQUAL) special purpose dataset model is used to capture non-standard variables and their association to parent records in domains, and allows capturing values for variables not presently included in the general-observation-class models. Because the SDTM does not allow the addition of new variables, it is necessary for sponsors to represent the metadata and data for each non-standard variable/value combination in a SUPPQUAL dataset. The SUPPQUAL dataset model is structured similarly to the RELREC dataset, in that it uses the same set of keys to identify related records. Each SUPPQUAL record also includes the name of the Qualifier variable being added (QNAM), the label for the variable (QLABEL), the actual value for each instance or record (QVAL), the origin (QORIG) of the value (whether it was collected via CRF, assigned or derived), and the Evaluator (QEVAL) to specify the role of the individual who assigned the value (such as ADJUDICATION COMMITTEE or SPONSOR).

One common case for using a SUPPQUAL dataset is to capture attributions. An attribution is typically an interpretation or subjective classification of one or more observations by a specific evaluator, such as a population flag that classifies a subject or their data according to their evaluability for efficacy analysis. Since it is possible that different attributions may be necessary in some cases, SUPPQUAL provides a mechanism for incorporating as many attributions as are necessary. It is recognized that sponsors may also need to use a SUPPQUAL dataset to submit additional non-standard variables that cannot be represented in the general classes. Standard names for certain expected values for QNAM and QLABEL are included in [Appendix 10.3.4](#).

The combined set of values for the first six columns (STUDYID...QNAM) should be unique for every record. That is, there should not be multiple records in a SUPPQUAL dataset for the same QNAM value, as it relates to IDVAR/IDVARVAL for a USUBJID. For example, if two individuals provide a determination on whether an Adverse Event is Treatment Emergent (e.g., the investigator and an independent adjudicator) then separate QNAM values should be used for each set of information, perhaps AETRTEMI and AETRTEMA. This is necessary to ensure that reviewers can join/merge/transpose the information back with the records in the original domain without risk of losing information.

When populating a SUPPQUAL dataset with population flags related to the Demographics domain (subject-level evaluability), there should be one record for each population flag for each subject. QVAL values for population flags should be Y or N, with no Null values. In the event that evaluability is based upon individual visits or CRF pages, additional population flags attached to other domains may be included in SUPPQUAL datasets.

Just as use of the optional grouping identifier variable --GRPID can be a more efficient method of representing relationships in RELREC, it can also be used in a SUPPQUAL dataset to identify individual qualifier values (SUPPQUAL records) related to multiple domain records that could be grouped, such as relating an attribution to a group of ECG measurements.

8.4.1 SUPPQUAL Dataset

suppqual.xpt, Supplemental Qualifiers, Version 3.1.1, August 26, 2005. One record per variable

Variable	Variable Label	Type	CDISC Notes	References
STUDYID	Study Identifier	Char	Study Identifier of the Parent record(s).	SDTM 2.2.4
RDOMAIN	Related Domain Abbreviation	Char*	Domain Abbreviation of the Parent record(s). Null for an attribution (e.g., ITT) that may apply for all records in all domains for a subject.	SDTM 2.2.4
USUBJID	Unique Subject Identifier	Char	Unique Subject Identifier of the Parent record(s).	SDTM 2.2.4
IDVAR	Identifying Variable	Char*	Identifying variable in the dataset that identifies the related record(s). Examples: --SEQ, --GRPID.	
IDVARVAL	Identifying Variable Value	Char	Value of identifying variable of the parent record(s).	
QNAM	Qualifier Variable Name	Char*	The short name of the Qualifier variable, which is used as a column name in a domain view with data from the parent domain. The value in QNAM cannot be longer than 8 characters, nor can it start with a number (e.g., '1TEST'). QNAM cannot contain characters other than letters, numbers, or underscores. This will often be the column name in the sponsor's operational dataset.	SDTMIG 4.1.5.2
QLABEL	Qualifier Variable Label	Char	This is the long name or label associated with QNAM. The value in QLABEL cannot be longer than 40 characters. This will often be the column label in the sponsor's original dataset.	
QVAL	Data Value	Char	Result of, response to, or value associated with QNAM. A value for this column is required; no records can be in SUPPQUAL with a null value for QVAL.	
QORIG	Origin	Char*	Since QVAL can represent a mixture of collected (on a CRF), derived, or assigned items, QORIG is used to indicate the origin of this data. Examples include CRF, ASSIGNED, or DERIVED.	
QEVAL	Evaluator	Char*	Used only for results that are subjective (e.g., assigned by a person or a group). Should be null for records that contain objectively collected or derived data. Some examples include ADJUDICATION COMMITTEE, STATISTICIAN, DATABASE ADMINISTRATOR, CLINICAL COORDINATOR, PRIMARY INVESTIGATOR, etc.	

* indicates variable may be subject to sponsor-controlled terminology;

** indicates variable may be subject to external controlled terminology.

Note all variables except QEVAL must be populated for each record. A record in a SUPPQUAL dataset relates back to its general observation class dataset parent record(s) via the key identified in the IDVAR variables. SUPPQUAL dataset records that are related to all subject records in all domains, such as the Intent To Treat (ITT) and Safety (SAFETY) subject-level population flags, will have both IDVAR and IDVARVAL set to null. This is because the only keys needed to identify the relationship/association to all subject records (for that subject) in all domains are STUDYID and USUBJID.

A SUPPQUAL dataset can contain both objective data (where values are collected or derived algorithmically) and subjective data (attributions where values are assigned by a person or committee). For objective data, the value in QEVAL will be null. For subjective data (when QORIG="ASSIGNED"), the value in QEVAL should reflect the role of the person or institution assigning the value (e.g., PRIMARY INVESTIGATOR or ADJUDICATION COMMITTEE).

All records in the SUPPQUAL datasets must have a value for QVAL. Transposing source variables with missing/null values may inadvertently generate SUPPQUAL records with null values for QVAL, causing the SUPPQUAL dataset to be extremely large. When this happens, the sponsor should delete the records where QVAL is null prior to submission.

8.4.2 Submitting Supplemental Qualifiers in Separate Datasets

Because sponsors often have a large number of supplemental qualifier variable values, the SDTM IG has defined a preferred approach to submit supplemental qualifiers by domain rather than placing all of the supplemental information within one dataset. Following this convention for the supplemental qualifiers produces a one-to-one correspondence between a domain dataset and its supplemental qualifier dataset. In such cases, the following rules should be met:

- The set of supplemental qualifiers for each domain should be included in a separate dataset with the name supp-- where -- denotes the source domain which the supplemental qualifiers relates back to. For example, population flags and other demographic qualifiers would be placed in suppdm.xpt.
- For users conforming to V3.1.1 and later versions of the SDTMIG, a SUPPQUAL dataset should NOT be submitted in addition to separate supp-- datasets for individual domains. In other words, separate supplemental qualifier domains cannot be used with some domains and SUPPQUAL for the other s).

8.4.3 SUPPQUAL Examples

The example below demonstrates how a set of SUPPQUAL datasets should be used to relate non-standard information to a parent domain.

In the first two rows of suppdm.xpt, population flags are defined as supplemental information to a subject's demography data. Since IDVAR and IDVARVAL are not populated, we know that the qualifying information is applicable to the subject as a whole, not to any specific record.

The next two rows of suppae.xpt add qualifying information to adverse event data. IDVAR defines the key variable that links this information to the AE data (AESEQ). IDVARVAL specifies the value of the key variable within the parent record that the SUPPQUAL record applies to. The remaining rows specify the supplemental variables' names (AESOTHC and AETRTEM), labels, values, origin (CRF or DERIVED), and who made the evaluation.

The final set of records in the suppq.s.xpt example show how to record the language used for a questionnaire. The parent domain (RDOMAIN) is QS, and the merge key (IDVAR) is QSCAT. QNAM holds the name of the supplemental qualifier variable being defined (QSLANG). The language recorded in QVAL applies to all of a subject's records where IDVAR (QSCAT) equals IDVARVAL. In this case, IDVARVAL has values for two questionnaires (SF36 and ADAS) for two separate subjects. QVAL identifies the questionnaire language version (Flemish or Farsi) for each subject.

suppdm.xpt: Supplemental Qualifiers for DM

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
1996001	DM	99-401			ITT	Intent to Treat	Y	DERIVED	SPONSOR
1996001	DM	99-401			PPROT	Per Protocol Evaluable	N	DERIVED	SPONSOR

suppae.xpt: Supplemental Qualifiers for AE

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
1996001	AE	99-401	AESEQ	1	AESOTHC	Desc of Other Med Impt Serious Event	Spontaneous Abortion	CRF	
1996001	AE	99-401	AESEQ	1	AETRTEM	Treatment Emergent	N	DERIVED	SPONSOR

suppq.s.xpt: Supplemental Qualifiers for QS

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
1996001	QS	99-401	QSCAT	SF36	QSLANG	Questionnaire Language	FLEMISH	CRF	
1996001	QS	99-401	QSCAT	ADAS	QSLANG	Questionnaire Language	FLEMISH	CRF	
1996001	QS	99-802	QSCAT	SF36	QSLANG	Questionnaire Language	FARSI	CRF	
1996001	QS	99-802	QSCAT	ADAS	QSLANG	Questionnaire Language	FARSI	CRF	

Examples of data that should not be in the SUPPQUAL dataset include:

- Subject-level objective data that fit in Subject Characteristics (SC)
- ECG Interpretation -- this should be added as an additional test called EGINTP in EGTESTCD, and given the same EGGRPID or EGREPID value for all records associated with that ECG.

8.5 RELATING COMMENTS TO A PARENT DOMAIN

The Comments special-purpose domain, which is described in [Section 5.1.2](#), is used to capture unstructured free text comments. It allows for the submission of comments related to a particular domain (e.g., adverse events) or those collected on separate general-comment pages not associated with a domain. Comments may be related to a Subject, a domain for a Subject, or to specific parent records in any domain. The Comments domain is structured similarly to the SUPQUAL dataset, in that it uses the same set of keys (STUDYID, RDOMAIN, USUBJID, IDVAR and IDVARVAL) to identify related records.

All comments are considered child records of data captured in domains. STUDYID, USUBJID, and DOMAIN (with the value "CO") must always be populated. RDOMAIN, IDVAR, and IDVARVAL should be populated as follows:

1. Comments related only to a subject in general would have RDOMAIN null, as the only key needed to identify the relationship/association to that subject is USUBJID.
2. Comments related only to a specific domain (and not to any specific record) for a subject would populate RDOMAIN with the code for the domain with which they are associated. IDVAR and IDVARVAL would be null.
3. Comments related to specific domain records for a subject would populate the RDOMAIN, IDVAR, and IDVARVAL variables with values from the parent record(s).

Again, --GRPID can be used as the value in IDVAR to identify comments with relationships to multiple domain records, such as a comment that applies to a 'group' of concomitant medications.

8.5.1 COMMENTS Example

In the following example, Subject 1234-2004 has five separate comments: one comment on an adverse events (AE) record, one on an exposure (EX) record, and one on a vital signs (VS) record; one on the PE CRF page, and another on a separate comments CRF page. Another comments example is provided in [Section 5.1.2](#).

STUDYID	DO MAIN	RDO MAIN	USUBJID	CO SEQ	IDVAR	IDVAR VAL	COREF	CODTC	COVAL	COEVAL
1234	CO	AE	1234-2004	1	AESEQ	7	PAGE 650		Same AE has occurred prior to treatment	PRINCIPAL INVESTIGATOR
1234	CO	EX	1234-2004	2	EXSEQ	17			Some delays in taking medications because of vacation	PRINCIPAL INVESTIGATOR
1234	CO	VS	1234-2004	3	VSSEQ	5			Value confirmed in range, but questionable based on other related measurements.	ADJUDICATION COMMITTEE
1234	CO	PE	1234-2004	4			PAGE 40	2003-11-08	Subject was very uncooperative during physical exam.	PRINCIPAL INVESTIGATOR
1234	CO		1234-2004	5				2004-01-14	Subject overall attitude has improved greatly during study.	PRINCIPAL INVESTIGATOR

8.6 HOW TO DETERMINE WHERE DATA BELONG IN THE SDTM

A domain is defined as a collection of observations on a particular topic. For example, Medical History is a collection of information related to events or conditions that happened before the clinical study.

Typically the term 'domain' has been synonymous with a dataset. However, there may be cases where data related to the same topic might, at times, need to be captured in more than one domain or observation class. In such cases, information that may have been historically presented in a single dataset may now need to be represented in more than one SDTM dataset. It is often necessary to refer to the protocol or statistical analysis plan to decide how to model new data domains not modeled in the SDTMIG and understand whether a collection of information logically belongs together in a single domain or dataset.

When determining how to model a new type of data, it is best to identify the domain topicality first and classify its primary observation class as an Event, Intervention, or Finding (see section 2.4). To do this, begin by evaluating the overall logical content of the information rather than analyzing the physical structure. Applying this principle to Medical History, for example, note that the collected information is related to previous conditions or events, regardless of whether dates were collected. Just because dates were not collected or available in specific cases does not change the fact that the primary topicality fits Events. Conversely, if the topic of the information is more related to specific measurements rather than just recording occurrences, it is typically modeled in the Findings class.

Once the domain topicality and its general observation class are identified, then the physical structure of the information should be evaluated to verify whether all collected information can be represented in a single observation class. Just because the information is collected /recorded on the same individual CRF pages does not necessarily mean that all the data collected/recorded on those individual CRF pages belongs in the same domain.

Information grouped together on CRFs for convenience but obviously having different topicalities (such as Concomitant Medications collected on SAE forms) should be mapped to multiple domains. If the sponsor feels that it's important for the reviewer to easily identify the relationship between the AE and CM records, then RELREC can be used (see Section 8.2.1).

Information grouped together because it is dependent on other information, such as findings collected as a result of an event, requires careful evaluation to determine whether all the information should be in a single domain, split into multiple datasets, or submitted in a supplemental qualifiers dataset. The following guidelines should be used in making this determination:

- Submit data in a supplemental qualifiers special-purpose dataset (see Section 8.4) if the information that does not fit consists only of additional variables that would not be classified as Identifiers, Topic, or Timing variables in a separate domain. In other words, the information consists only of what would be classified as additional Qualifiers to further describe an existing record, and would not be considered sufficient to exist as an entire 'record' of information by itself in any observation class. A general way of thinking about this is if the data would simply have been another column in the sponsor's original dataset.
- Submit data as a separate dataset if the information that does not fit could comprise a complete record in that dataset (it contains variables that would be classified as Identifiers, Topic, or Timing variables). Section 8.2 describes how information should be stored in a second dataset of another general observation class for the same Domain when the information has, at minimum a Topic Variable, its own Timing variables and (for Findings domains) at least one Result Qualifier.

General rules for creating a separate related dataset of common topicality with another are as follows:

1. Typically, additional datasets should be created for either
 - a. Interventions with associated Findings, or
 - b. Events with associated Findings
 - c. Two sets of Findings that do not fit logically in a single dataset.

2. The topicality of the information and the context determine the Parent (or master) dataset general observation class (Interventions, Events or Findings). In some cases, a Parent (master) dataset may be required, even if the records in that dataset have limited information
3. Both datasets should include a common identifier variable (e.g., --REFID) to capture the identification (name) of the record to be used as a merge/join key; however, it is recognized that in some cases it would be more appropriate to use --GRPID. Values for the --REFID or --GRPID variables would need to be the same for related records in the two (or more) datasets.
4. Both datasets should use a different two-character code for a value in the Domain variable, and as the variable prefix (MB and MS for example). The dataset names for the two datasets should be the two-character Domain codes chosen for the datasets (MB and MS respectively). Since it is not self-evident that a Domain is captured in multiple datasets using this combination of dataset names (a reviewer might not automatically know that MB and MS are related to each other), sponsors **must** populate the RELREC dataset (see [Section 8.2.1](#)) with rows that identify the dataset relationship and document the approach used in the Define file (metadata) and/or Annotated CRF. This approach is fully compatible with the current V3.x standard as long as the dataset relationship is clearly documented.

9 Implementation Examples

9.1 DEMOGRAPHICS EXAMPLE

This is an example of demographics that illustrates the use of the RACE and ETHNIC variables. Both RACE and ETHNIC are subjected to the rules of controlled terminology. Race (RACE) and ethnicity (ETHNIC) are independent of each other, for example, the first subject (USUBJID=1) is 'WHITE' for RACE and 'HISPANIC OR LATINO' for ETHNIC. Ethnicity can only have two values, either 'HISPANIC OR LATINO' or 'NOT HISPANIC OR LATINO'. In cases of mixed race, the RACE variable should be used for the primary race, and the SC domain should be used for additional non-primary race values. For example, in Row 4, the 4th subject (USUBJID=4), who is recorded as 'ASIAN' for RACE could have the value 'INDIAN' in SC where SCTESTCD equals 'RACEOTH.'

Row	STUDYID	DOMAIN	USUBJID	RACE	ETHNIC
Row 1	ABC123	DM	1	WHITE	HISPANIC OR LATINO
Row 2	ABC123	DM	2	WHITE	NOT HISPANIC OR LATINO
Row 3	ABC123	DM	3	BLACK OR AFRICAN AMERICAN	NOT HISPANIC OR LATINO
Row 4	ABC123	DM	4	ASIAN	NOT HISPANIC OR LATINO
Row 5	ABC123	DM	5	AMERICAN INDIAN OR ALASKA NATIVE	NOT HISPANIC OR LATINO
Row 6	ABC123	DM	6	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDERS	NOT HISPANIC OR LATINO

9.2 INTERVENTIONS EXAMPLES

9.2.1 CM Example: Intermittent Use of Concomitant Medications

Sponsors collect the timing of concomitant medication use with varying specificity, depending on the pattern of use; the type, purpose, and importance of the medication; and the needs of the study. Often the same information can be conveyed with a range of dates and the daily frequency rather than presenting every unique instance of medication use. If appropriate, medications taken intermittently or sporadically over a time period may be reported with a date range and a frequency of 'PRN'.

The example below shows three subjects who took the same medication at the same times. For the first subject (USUBJID=1), each instance is recorded separately, and frequency (CMDOSFRQ) is ONCE. For the second subject (USUBJID=2), the second record (CMSEQ=2) shows that aspirin was taken twice on January 7th, so the frequency is BID. The frequency is also included for the other daily records to avoid confusion. The third subject's (USUBJID=3) records are

collapsed (this is shown as an example only, not as a recommendation) into a single entry that spans the relevant time period, with a frequency of PRN. This approach assumes that knowing exactly when aspirin was used is not important for evaluating this study's safety and efficacy.

Row	STUDYID	DOMAIN	USUBJID	CMSEQ	CMTRT	CMDOSFRQ	CMSTDTC	CMENDTC
Row 1	ABC123	CM	1	1	Aspirin	ONCE	2004-01-01	2004-01-01
Row 2	ABC123	CM	1	2	Aspirin	ONCE	2004-01-02	2004-01-02
Row 3	ABC123	CM	1	3	Aspirin	ONCE	2004-01-03	2004-01-03
Row 4	ABC123	CM	1	4	Aspirin	ONCE	2004-01-07	2004-01-07
Row 5	ABC123	CM	1	5	Aspirin	ONCE	2004-01-07	2004-01-07
Row 6	ABC123	CM	1	6	Aspirin	ONCE	2004-01-09	2004-01-09
Row 7	ABC123	CM	2	1	Aspirin	QD	2004-01-01	2004-01-03
Row 8	ABC123	CM	2	2	Aspirin	BID	2004-01-07	2004-01-07
Row 9	ABC123	CM	2	3	Aspirin	QD	2004-01-09	2004-01-09
Row 10	ABC123	CM	3	1	Aspirin	PRN	2004-01-01	2004-01-09

9.2.2 EX Examples:

Exposure Example 1:

This is an example of an Exposure dataset for a parallel design study. In this example, subjects were randomized to one of three treatment groups: Drug A 40 mg q.d., Drug A 20 mg q.d., or Drug B 150 mg bid. Drug A was in capsule form and Drug B was in tablet form. To maintain the blind, subjects took both capsules and tablets with only one form being active. The study included 8 weeks of treatment, with subjects remaining on the same treatment throughout the study. With respect to timing of doses, the sponsor only collected the start and stop dates of uninterrupted periods of treatment. Note below that Subject 12345003 missed taking study medications on Study Days 23 and 24.

	STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXDOSE	EXDOSU	EXDOSFRM
Row 1	12345	EX	12345001	1	DRUG A	40	MG	CAPSULE
Row 2	12345	EX	12345001	2	PLACEBO	0		TABLET
Row 3	12345	EX	12345002	1	DRUG A	20	MG	CAPSULE
Row 4	12345	EX	12345002	2	PLACEBO	0		TABLET
Row 5	12345	EX	12345003	1	PLACEBO	0		CAPSULE
Row 6	12345	EX	12345003	2	DRUG B	150	MG	TABLET
Row 7	12345	EX	12345003	3	PLACEBO	0		CAPSULE
Row 8	12345	EX	12345003	4	DRUG B	150	MG	TABLET

	EXDOSFRQ	EXDOSTO T	EXROUTE	EXSTDTC	EXENDTC	EXSTDY	EXENDY
Row 1 (cont)	QD	40	ORAL	2002-01-10	2002-03-08	1	58
Row 2 (cont)	BID	0	ORAL	2002-01-10	2002-03-08	1	58
Row 3 (cont)	QD	20	ORAL	2002-01-10	2002-03-07	1	57
Row 4 (cont)	BID	0	ORAL	2002-01-10	2002-03-07	1	57
Row 5 (cont)	QD	0	ORAL	2002-01-11	2002-02-01	1	22
Row 6 (cont)	BID	300	ORAL	2002-01-11	2002-02-01	1	22
Row 7 (cont)	QD	0	ORAL	2002-02-04	2002-03-06	25	55
Row 8 (cont)	BID	300	ORAL	2002-02-04	2002-03-06	25	55

Exposure Example 2:

This is an example of an Exposure dataset for a two-way crossover study comparing once daily oral administration of Drug A 20 mg with Drug B 30 mg. Study drug was taken for 3 consecutive mornings 30 minutes prior to a standardized breakfast. There was a 5-day washout period between treatments.

	STUDYID	DOMAIN	USUBJID	EXSEQ	EXGRPID	EXTRT	EXDOSE	EXDOSU	EXDOSFRM	EXDOSFRQ
Row 1	56789	EX	56789001	1	1	DRUG A	20	MG	CAPSULE	QD
Row 2	56789	EX	56789001	2	1	DRUG A	20	MG	CAPSULE	QD
Row 3	56789	EX	56789001	3	1	DRUG A	20	MG	CAPSULE	QD
Row 4	56789	EX	56789001	4	2	DRUG B	30	MG	TABLET, COATED	QD
Row 5	56789	EX	56789001	5	2	DRUG B	30	MG	TABLET, COATED	QD
Row 6	56789	EX	56789001	6	2	DRUG B	30	MG	TABLET, COATED	QD
Row 7	56789	EX	56789003	1	1	DRUG B	30	MG	TABLET, COATED	QD
Row 8	56789	EX	56789003	2	1	DRUG B	30	MG	TABLET, COATED	QD
Row 9	56789	EX	56789003	3	1	DRUG B	30	MG	TABLET, COATED	QD
Row 10	56789	EX	56789003	4	2	DRUG A	20	MG	CAPSULE	QD
Row 11	56789	EX	56789003	5	2	DRUG A	20	MG	CAPSULE	QD
Row 12	56789	EX	56789003	6	2	DRUG A	20	MG	CAPSULE	QD

	EXDOSTOT	EXROUTE	EXSTDTC	EXENDTC	EXSTDY	EXENDY	EXTPT
Row 1 (cont)	20	ORAL	2002-07-01T07:30	2002-07-01T07:30	1	1	30 MINUTES PRIOR TO STD BREAKFAST
Row 2 (cont)	20	ORAL	2002-07-02T07:30	2002-07-02T07:30	2	2	30 MINUTES PRIOR TO STD BREAKFAST
Row 3 (cont)	20	ORAL	2002-07-03T07:32	2002-07-03T07:32	3	3	30 MINUTES PRIOR TO STD BREAKFAST
Row 4 (cont)	30	ORAL	2002-07-09T07:30	2002-07-09T07:30	9	9	30 MINUTES PRIOR TO STD BREAKFAST
Row 5 (cont)	30	ORAL	2002-07-10T07:30	2002-07-10T07:30	10	10	30 MINUTES PRIOR TO STD BREAKFAST
Row 6 (cont)	30	ORAL	2002-07-11T07:34	2002-07-11T07:34	11	11	30 MINUTES PRIOR TO STD BREAKFAST
Row 7 (cont)	30	ORAL	2002-07-03T07:30	2002-07-03T07:30	1	1	30 MINUTES PRIOR TO STD BREAKFAST
Row 8 (cont)	30	ORAL	2002-07-04T07:24	2002-07-04T07:24	2	2	30 MINUTES PRIOR TO STD BREAKFAST
Row 9 (cont)	30	ORAL	2002-07-05T07:24	2002-07-05T07:24	3	3	30 MINUTES PRIOR TO STD BREAKFAST
Row 10 (cont)	20	ORAL	2002-07-11T07:30	2002-07-11T07:30	9	9	30 MINUTES PRIOR TO STD BREAKFAST
Row 11 (cont)	20	ORAL	2002-07-12T07:43	2002-07-12T07:43	10	10	30 MINUTES PRIOR TO STD BREAKFAST
Row 12 (cont)	20	ORAL	2002-07-13T07:38	2002-07-13T07:38	11	11	30 MINUTES PRIOR TO STD BREAKFAST

9.2.3 SU Example:

Example and Usage Notes for V3.1.1 Substance Use (SU) Domain Model

The use of the Interventions general observation class to model substance use data meets the needs of the complex, detailed dosing, frequency, and duration information requirements for analysis where collection of this information is key to the outcome of a clinical trial (e.g., smoking cessation).

This example is for clinical trials where detailed substance use dosing, frequency, and duration information is required (e.g., smoking cessation trials). In such cases, detailed information (e.g., smoking substance, total amount per day, start and stop dates) about the substance used is collected. This data could then be used in primary efficacy analyses to support therapy effectiveness claims. Figure 9.2.3.1 presents a CRF that captures data of this nature. The information in the gray bar/ellipses is the sponsor's original annotation.

Figure 9.2.3.1 - Sample Case Report Form for Smoking Status

Visit x	SMOKING	Smoking Status			
---------	----------------	-----------------------	--	--	--

Please identify **one** of the following smoking categories for the subject:

₁ SMOKCAT ₁ Not assessed NOTDONE

<input type="checkbox"/> ₁	Non-smoker (never smoked)	Start date	<input type="checkbox"/> ₁ SMOKSDT	Stop date	<input type="checkbox"/> ₁ SMOKEDT														
<input type="checkbox"/> ₂	Ex smoker	<table border="1" style="width: 100%; height: 20px; border-collapse: collapse;"> <tr> <td style="width: 20px;">d</td><td style="width: 20px;">d</td><td style="width: 20px;">m</td><td style="width: 20px;">m</td><td style="width: 20px;">y</td><td style="width: 20px;">y</td><td style="width: 20px;">y</td><td style="width: 20px;">y</td> </tr> </table>	d	d	m	m	y	y	y	y	<table border="1" style="width: 100%; height: 20px; border-collapse: collapse;"> <tr> <td style="width: 20px;">d</td><td style="width: 20px;">d</td><td style="width: 20px;">m</td><td style="width: 20px;">m</td><td style="width: 20px;">y</td><td style="width: 20px;">y</td><td style="width: 20px;">y</td><td style="width: 20px;">y</td> </tr> </table>	d	d	m	m	y	y	y	y
d	d	m	m	y	y	y	y												
d	d	m	m	y	y	y	y												
<input type="checkbox"/> ₃	Current smoker	<table border="1" style="width: 100%; height: 20px; border-collapse: collapse;"> <tr> <td style="width: 20px;">d</td><td style="width: 20px;">d</td><td style="width: 20px;">m</td><td style="width: 20px;">m</td><td style="width: 20px;">y</td><td style="width: 20px;">y</td><td style="width: 20px;">y</td><td style="width: 20px;">y</td> </tr> </table>	d	d	m	m	y	y	y	y									
d	d	m	m	y	y	y	y												

If Ex or Current smoker, complete detail section below, checking all that apply

SMOKTYPE

<input type="checkbox"/> ₁	Cigars	<table border="1" style="width: 100%; height: 20px; border-collapse: collapse;"> <tr> <td style="width: 20px;"></td><td style="width: 20px;"></td><td style="width: 20px;"></td> </tr> </table>				per day. Round to next highest whole number. Specify average daily amount smoked during the time the subject smoked.
<input type="checkbox"/> ₂	Pipe	<table border="1" style="width: 100%; height: 20px; border-collapse: collapse;"> <tr> <td style="width: 20px;"></td><td style="width: 20px;"></td><td style="width: 20px;"></td> </tr> </table>				bowls per day. Round to next highest whole number. Specify average daily amount smoked during the time the subject smoked.
<input type="checkbox"/> ₃	Cigarettes	<table border="1" style="width: 100%; height: 20px; border-collapse: collapse;"> <tr> <td style="width: 20px;"></td><td style="width: 20px;"></td><td style="width: 20px;"></td> </tr> </table>				per day. Round to next highest whole number. Specify average daily amount smoked during the time the subject smoked.

TOBACONO

Table 1 presents data that could be captured using the detailed smoking status CRF presented in Figure 9.2.3.1 using the Sponsor’s Operational database structure.

Row 1: Subject is a non-smoker.

Row 2: Subject is an ex-smoker. Detailed substance use information is provided.

Rows 3, 4: Subject is an ex-smoker of more than one type of tobacco. Detailed substance use information is provided.

Row 5: Subject is a current smoker. Detailed substance use information is provided.

Rows 6, 7: Subject came in for 2 visits, for the first visit subject was an ex-smoker. At the 2nd visit, subject began to smoke again. Detailed substance use information is provided.

Row 8: Subject did not have an assessment done.

Table 1: Data Captured Using the Detailed Smoking Status CRF in the Sponsor’s Operational Database

	STUDYID	USUBJID	VISIT	SEQ	SMOKCAT	SMOKTYPE	TOBACONO	SMOKSDT	SMOKEDT	NOTDONE
Row 1	1234	1234005	1	1	NON					
Row 2	1234	1234006	1	1	EX	CIGARS	1	1977	2003	
Row 3	1234	1234007	1	1	EX	PIPE	1	01-01-2000	04-02-2003	
Row 4	1234	1234007	1	2	EX	CIGARETTES	10	01-01-1990	04-02-1993	
Row 5	1234	1234008	1	1	CURRENT	CIGARETTES	20	1945		
Row 6	1234	1234009	1	1	EX	CIGARS	1	1965	1970	
Row 7	1234	1234009	2	2	CURRENT	CIGARETTES	20	01-01-2003		
Row 8	1234	1234010	1	1						1

Table 2 presents the data structured according to the SDTM. If this was a multi-page CRF, collecting multiple substance use information (for example, tobacco use and alcohol consumption) then SUCAT would have values of TOBACCO and ALCOHOL. As the primary physiological impact of tobacco use is due to nicotine content, the value of SUDECOD was set to NICOTINE. SUSTRF and SUENRF hold the values of SMOKCAT as they are interpreted based on the subject’s use status at sponsor-defined start and end reference timeframes.

Row 1: Because this subject did not consume tobacco products, the value of SRTRT was set to 'Tobacco'. In addition, the values of SUSTRF and SUENRF are set to NULL, and the value of SUOCCUR could be set to "N" because this subject did not use any tobacco products.

Row 8: Because the assessment for this subject was not done, the value of SUSTAT was set to NOT DONE.

Table 2: Data from Example 3 Modeled in SDS V3.x SU Domain

	STUDYID	USUBJID	SUSEQ	SUTRT	SUDECOD	SUCAT	SUSTAT	SUSTRF	SUENRF	SUDOSFRQ	SUDOSTOT
Row 1	1234	1234005	1	TOBACCO	NICOTINE	TOBACCO					
Row 2	1234	1234006	1	CIGARS	NICOTINE	TOBACCO		BEFORE	BEFORE	AVG/DAY	1
Row 3	1234	1234007	1	PIPE	NICOTINE	TOBACCO		BEFORE	BEFORE	AVG/DAY	1
Row 4	1234	1234007	2	CIGARETTES	NICOTINE	TOBACCO		BEFORE	BEFORE	AVG/DAY	10
Row 5	1234	1234008	1	CIGARETTES	NICOTINE	TOBACCO		BEFORE	ONGOING	AVG/DAY	20
Row 6	1234	1234009	1	CIGARS	NICOTINE	TOBACCO		BEFORE	BEFORE	AVG/DAY	1
Row 7	1234	1234009	2	CIGARETTES	NICOTINE	TOBACCO		DURING	ONGOING	AVG/DAY	20
Row 8	1234	1234010	1			TOBACCO	NOT DONE				

	VISITNUM	SUSTDTC	SUENDTC
Row 1 (cont)	1		
Row 2 (cont)	1	1977	2003
Row 3 (cont)	1	2000-01-01	2003-02-04
Row 4 (cont)	1	1990-01-01	1993-02-04
Row 5 (cont)	1	1945	
Row 6 (cont)	1	1965	1970
Row 7 (cont)	2	2003-01-01	
Row 8 (cont)	1		

Not all variables specified in the CDISC SU domain are present on the example CRF. Since they are not used, the variables are not present on the SU dataset.

9.3 EVENTS EXAMPLES

9.3.1 AE Example

In this example, an AE CRF collects AE terms as free text. There is a section at the bottom of the form for free-text comments. AEs are collected one time per study. In this example, first study drug was administered to the subject on 13Oct2003 at 12:00. 3 AEs were reported. AEs are coded using MedDRA and the sponsor's procedures include the possibility of modifying the reported term to aid in coding.

Notes:

1. Not all variables specified in the CDISC AE domain are present on the CRF (e.g., AEACNOTH, AERELNST, AEDUR). Since they are not used, the variables are not present in the AE dataset. However, all required and expected variables are present and populated.
2. Comments would be placed in the Comments domain (CO).
3. The CRF is structured so that serious category variables (e.g., AESDTH, AESHOSP) are checked only when an overall seriousness question is answered 'Y.'
4. If a subject had no adverse events reported, no records are included for that subject in the AE dataset.
5. If a subject has a value for AEENDTC then AEENRF is null.

Sample data that illustrates the above:

	STUDYID	DOMAIN	USUBJID	AESEQ	AETERM	AESTDTC	AEENDTC	AEMODIFY	AEDECOD
Row 1	ABC123	AE	123101	1	POUNDING HEADACHE	2003-10-12	2003-10-12	HEADACHE	HEADACHE
Row 2	ABC123	AE	123101	2	BACK PAIN FOR 6 HOURS	2003-10-13T13:05	2003-10-13T19:00	BACK PAIN	BACK PAIN
Row 3	ABC123	AE	123101	3	PULMONARY EMBOLISM	2003-10-21			PULMONARY EMBOLISM

	AEBODSYS	AESEV	AESER	AEACN	AEREL
Row 1 (cont)	NERVOUS SYSTEM DISORDERS	SEVERE	N	NOT APPLICABLE	DEFINITELY NOT RELATED
Row 2 (cont)	MUSCULOSKELETAL AND CONNECTIVE TISSUE DISORDERS	MODERATE	N	DOSE REDUCED	PROBABLY RELATED
Row 3 (cont)	VASCULAR DISORDERS	MODERATE	Y	DOSE REDUCED	PROBABLY NOT RELATED

	AEOUT	AESCONG	AESDISAB	AESDTH	AESHOSP	AESLIFE	AESMIE	AESTDY	AEENDY	AEENRF
Row 1 (cont)	RECOVERED/RESOLVED							-1	-1	
Row 2 (cont)	RECOVERED/RESOLVED							1	1	
Row 3 (cont)	RECOVERING/RESOLVING				Y	Y		9	.	AFTER

9.3.2 DS Examples

Disposition Example 1:

In this example, a DS CRF collects multiple disposition events at different time points in the study indicated by EPOCH. There are also several protocol milestones which are indicated by DSCAT = 'PROTOCOL MILESTONE'. DSTERM is populated with controlled terminology with the same value as DSDECOD except in the case when there is free text for DSTERM such as 'Subject moved'. In this case, the controlled terminology is only in DSDECOD (LOST TO FOLLOW-UP).

Notes

1. Not all variables specified in the CDISC DS domain are present on the CRF (e.g., DSSCAT, DSSTDY). Since they are not used; the variables are not present on the DS dataset. However, all required and expected variables are present and populated.
2. In the Events observation class, the date/time of the Disposition Event is the DSSTDTC variable. The date of collection of the disposition information is the DSDTC variable
3. EPOCH can be derived from VISITNUM. In the example below, when VISITNUM = 0, then EPOCH = SCREENING. EPOCH is populated when DSCAT is a disposition event and null when DSCAT is a protocol milestone.

There are multiple disposition events for each subject:

1. Subject 123101 has 3 records to indicate the completion of 3 stages of the study, which are screening, treatment phase and follow-up. Protocol milestone records for informed consent and randomization are also included.
2. Subject 123102 is a screen drop. Screen drops are identified by a DSDECOD that is not equal to 'COMPLETED' for the SCREENING stage. This is an example of the submission of the verbatim reason for discontinuation in DSTERM.
3. Subject 123103 completed the screening stage but did not complete the treatment stage.
4. Subject 123104 died on October 29, 2003 (see DSSTDTC) after the completion of treatment, but prior to the completion of follow-up. Note that the date of collection of the event information was on October 31, 2003 (DSDTC).
5. Subject 123105 discontinued study treatment due to an AE, but went on to complete the follow-up phase of the trial.

CDISC SDTM Implementation Guide (Version 3.1.1)

Sample data that illustrates the above:

	STUDYID	DOMAIN	USUBJID	DSSEQ	DSTERM	DSDECOD	DSCAT	VISITNUM	EPOCH	DSDTC	DSSTDTC
Row 1	ABC123	DS	123101	1	INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE	0		2003-09-21	2003-09-21
Row 2	ABC123	DS	123101	2	COMPLETED	COMPLETED	DISPOSITION EVENT	0	SCREENING	2003-09-29	2003-09-29
Row 3	ABC123	DS	123101	3	RANDOMIZED	RANDOMIZED	PROTOCOL MILESTONE	1		2003-09-30	2003-09-30
Row 4	ABC123	DS	123101	4	COMPLETED	COMPLETED	DISPOSITION EVENT	2	TREATMENT PHASE	2003-10-31	2003-10-31
Row 5	ABC123	DS	123101	5	COMPLETED	COMPLETED	DISPOSITION EVENT	3	FOLLOW-UP	2003-11-15	2003-11-15
Row 6	ABC123	DS	123102	1	INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE	0		2003-11-21	2003-11-21
Row 7	ABC123	DS	123102	2	SUBJECT DENIED MRI PROCEDURE	PROTOCOL VIOLATION	DISPOSITION EVENT	0	SCREENING	2003-11-22	2003-11-20
Row 8	ABC123	DS	123103	1	INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE	0		2003-09-15	2003-09-15
Row 9	ABC123	DS	123103	2	COMPLETED	COMPLETED	DISPOSITION EVENT	0	SCREENING	2003-09-22	2003-09-22
Row 10	ABC123	DS	123103	3	RANDOMIZED	RANDOMIZED	PROTOCOL MILESTONE	1		2003-09-30	2003-09-30
Row 11	ABC123	DS	123103	4	SUBJECT MOVED	LOST TO FOLLOW-UP	DISPOSITION EVENT	1	TREATMENT PHASE	2003-10-31	2003-10-31
Row 12	ABC123	DS	123104	1	INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE	0		2003-09-15	2003-09-15
Row 13	ABC123	DS	123104	2	COMPLETED	COMPLETED	DISPOSITION EVENT	0	SCREENING	2003-09-22	2003-09-22
Row 14	ABC123	DS	123104	3	RANDOMIZED	RANDOMIZED	PROTOCOL MILESTONE	1		2003-09-30	2003-09-30
Row 15	ABC123	DS	123104	4	COMPLETED	COMPLETED	DISPOSITION EVENT	2	TREATMENT PHASE	2003-10-15	2003-10-15
Row 16	ABC123	DS	123104	5	SUBJECT DIED	DEATH	DISPOSITION EVENT	1	FOLLOW-UP	2003-10-31	2003-10-29
Row 17	ABC123	DS	123105	1	INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE	0		2003-09-28	2003-09-28
Row 18	ABC123	DS	123105	2	COMPLETED	COMPLETED	DISPOSITION EVENT	0	SCREENING	2003-10-02	2003-10-02
Row 19	ABC123	DS	123105	3	RANDOMIZED	RANDOMIZED	PROTOCOL MILESTONE	1		2003-10-02	2003-10-02
Row 20	ABC123	DS	123105	4	ANEMIA	ADVERSE EVENT	DISPOSITION EVENT	2	TREATMENT PHASE	2003-10-17	2003-10-17
Row 21	ABC123	DS	123105	5	COMPLETED	COMPLETED	DISPOSITION EVENT	3	FOLLOW-UP	2003-11-02	2003-11-02

Disposition Example 2:

In this example, the sponsor has chosen to simply submit whether or not the subject completed the study. Examples are shown of 1 subject who completed the study and 2 who discontinued prior to completion. In this very straightforward situation, only required and expected variables have been included.

	STUDYID	DOMAIN	USUBJID	DSSEQ	DSTERM	DSDECOD	DSSTDTC
Row 1	ABC456	DS	456101	1	COMPLETED	COMPLETED	2003-09-21
Row 2	ABC456	DS	456102	1	SUBJECT TAKING STUDY MED ERRATICALLY	PROTOCOL VIOLATION	2003-09-29
Row 3	ABC456	DS	456103	1	LOST TO FOLLOW-UP	LOST TO FOLLOW-UP	2003-10-15

Disposition Example 3:

In this example, the sponsor collects whether or not the subject completes the treatment and follow-up phases as well as whether the subject's treatment assignment was unblinded. The date of the unblinding is represented in DSSTDTC. Note that maintaining the blind as per protocol is not considered to be an event since there is no change in the subject's state.

	STUDYID	DOMAIN	USUBJID	DSSEQ	DSCAT	EPOCH	DSTERM	DSDECOD	DSSTDTC
Row 1	ABC789	DS	789101	1	DISPOSITION EVENT	TREATMENT PHASE	COMPLETED	COMPLETED	2004-09-12
Row 2	ABC789	DS	789101	2	DISPOSITION EVENT	FOLLOW-UP	COMPLETED	COMPLETED	2004-12-20
Row 3	ABC789	DS	789102	1	DISPOSITION EVENT	TREATMENT PHASE	SKIN RASH	ADVERSE EVENT	2004-09-30
Row 4	ABC789	DS	789102	2	OTHER EVENTS	TREATMENT PHASE	SUBJECT HAD SEVERE RASH	TREATMENT UNBLINDED	2003-10-01
Row 5	ABC789	DS	789102	3	DISPOSITION EVENT	FOLLOW-UP	COMPLETED	COMPLETED	2004-12-28

9.3.3 MH Example:

In this example, a study collects medical history information using 3 separate CRFs. A general medical history CRF collects descriptions of events by body system (e.g., Endocrine, Metabolic) and asks whether the conditions were ongoing or not at study start (Rows 1-3). The reported events are coded using MedDRA. Another CRF asks whether the subject had any of a list of specific risk factors. The third CRF collects the primary diagnosis and onset date. Risk factors and primary diagnosis are not coded using MedDRA.

Notes

- a. Not all variables specified in the CDISC MH domain are present on the CRF (e.g., MHENDTC). Since they are not used, the variables are not present in the MH dataset shown. However, all required variables are present and populated.
- b. Comments, if provided, would be placed in the Comments domain (CO).
- c. MHCAT is used to specify the category of data collected (GENERAL MEDICAL HISTORY, RISK FACTORS, PRIMARY DIAGNOSIS).
- d. MHSCAT displays the body systems specified on the general medical history CRF.
- e. The “Ongoing at Study Start” question on the general medical history has been translated to MHENRF. If “Yes” is specified, MHENRF=“DURING/AFTER;” if “No” is specified, MHENRF=“BEFORE.”

Sample data that illustrates the above:

	STUDYID	DOMAIN	USUBJID	MHSEQ	MHTERM	MHDECOD	MHCAT	MHSCAT
Row 1	ABC123	MH	123101	1	ASTHMA	ASTHMA	GENERAL MEDICAL HISTORY	RESPIRATORY
Row 2	ABC123	MH	123101	2	FREQUENT HEADACHES	HEADACHE	GENERAL MEDICAL HISTORY	CNS
Row 3	ABC123	MH	123101	3	BROKEN LEG	BONE FRACTURE	GENERAL MEDICAL HISTORY	OTHER
Row 4	ABC123	MH	123101	5	HYPERTENSION		RISK FACTORS	
Row 5	ABC123	MH	123101	6	TIA		RISK FACTORS	
Row 6	ABC123	MH	123101	4	STROKE		PRIMARY DIAGNOSIS	

	MHBODSYS	MHSTDTC	MHENRF
Row 1 (cont)	RESPIRATORY SYSTEM DISORDERS		DURING/AFTER
Row 2 (cont)	CENTRAL AND PERIPHERAL NERVOUS SYSTEM DISORDERS		DURING/AFTER
Row 3 (cont)	MUSCULOSKELTAL SYSTEM DISORDERS		BEFORE
Row 4 (cont)			
Row 5 (cont)			
Row 6 (cont)		2004-09-17T07:30	

9.4 FINDINGS EXAMPLES

9.4.1 EG Examples

ECG Example 1:

This is an example of an ECG dataset for one subject at one visit where cycle measurements, findings, and an interpretation were delivered to the sponsor by an ECG provider. EGREFID is an internal or external ECG identifier that links to the external ECG Waveform Files. EGXFN is the file name for the external ECG Waveform file.

Rows 1-6 - Show how ECG cycle measurements are represented.

Rows 2-6 - Show the data in original units of measure in EGORRES and in the converted value in EGSTRESC and EGSTRESN

Rows 5-6 - Show QTCB and QTCF. The actual data should have a 'Y' in the EGDRVFL column since these results are derived by the sponsor in this example. See assumption [4.1.5.1](#).

Rows 7-10 - Show how ECG findings are represented.

Row 11 - Shows a way of representing comments that are important to the overall understanding of the ECG, but which are not truly findings or interpretations.

Row 12 - Shows how interpretations are represented.

CDISC SDTM Implementation Guide (Version 3.1.1)

	STUDYID	DOMAIN	USUBJID	EGSEQ	EGGRPID	EGREFID	EGTESTCD	EGTEST	EGXFN	EGNAM	EGPOS
Row 1	XYZ	EG	XYZ-US-701-002	1	1	334PT89	VRATE	Ventricular Rate	PQW436789-07	Test Lab	SUPINE
Row 2	XYZ	EG	XYZ-US-701-002	2	1	334PT89	PR	PR Interval	PQW436789-07	Test Lab	SUPINE
Row 3	XYZ	EG	XYZ-US-701-002	3	1	334PT89	QRS	QRS Interval	PQW436789-07	Test Lab	SUPINE
Row 4	XYZ	EG	XYZ-US-701-002	4	1	334PT89	QT	QT Interval	PQW436789-07	Test Lab	SUPINE
Row 5	XYZ	EG	XYZ-US-701-002	5	1	334PT89	QTCB	QTc - Bazett	PQW436789-07	Test Lab	SUPINE
Row 6	XYZ	EG	XYZ-US-701-002	6	1	334PT89	QTCF	QTc - Fridericia	PQW436789-07	Test Lab	SUPINE
Row 7	XYZ	EG	XYZ-US-701-002	7	1	334PT89	RHYRATE	Rhythm and Rate	PQW436789-07	Test Lab	SUPINE
Row 8	XYZ	EG	XYZ-US-701-002	8	1	334PT89	RHYRATE	Rhythm and Rate	PQW436789-07	Test Lab	SUPINE
Row 9	XYZ	EG	XYZ-US-701-002	9	1	334PT89	QTABN	QT Abnormalities	PQW436789-07	Test Lab	SUPINE
Row 10	XYZ	EG	XYZ-US-701-002	10	1	334PT89	VCABN	Ventricular Conduction Abnormalities		Test Lab	SUPINE
Row 11	XYZ	EG	XYZ-US-701-002	11	1	334PT89	TECHPROB	Technical Problems		Test Lab	SUPINE
Row 12	XYZ	EG	XYZ-US-701-002	12	1	334PT89	INTP	Interpretation			SUPINE

	VISIT	EGORRES	EGORRESU	EGSTRESC	EGSTRESN	EGSTRESU	EGEVAL	VISITNUM	EGDTC	EGDY
Row 1 (cont)	Screening 1	62	bpm	62	62	bpm		1	2003-04-15T11:58	-36
Row 2 (cont)	Screening 1	0.15	sec	150	150	msec		1	2003-04-15T11:58	-36
Row 3 (cont)	Screening 1	0.103	sec	103	103	msec		1	2003-04-15T11:58	-36
Row 4 (cont)	Screening 1	0.406	sec	406	406	msec		1	2003-04-15T11:58	-36
Row 5 (cont)	Screening 1	0.469	sec	469	469	msec		1	2003-04-15T11:58	-36
Row 6 (cont)	Screening 1	0.446	sec	446	446	msec		1	2003-04-15T11:58	-36
Row 7 (cont)	Screening 1	ATRIAL FIBRILLATION		ATRIAL FIBRILLATION				1	2003-04-15T11:58	-36
Row 8 (cont)	Screening 1	ATRIAL FLUTTER		ATRIAL FLUTTER				1	2003-04-15T11:58	-36
Row 9 (cont)	Screening 1	QT PROLONGATION		QT PROLONGATION				1	2003-04-15T11:58	-36
Row 10 (cont)	Screening 1	LEFT VENTRICULAR HYPERTROPHY		LEFT VENTRICULAR HYPERTROPHY				1	2003-04-15T11:58	-36
Row 11 (cont)	Screening 1	INCORRECT ELECTRODE PLACEMENT		INCORRECT ELECTRODE PLACEMENT				1	2003-04-15T11:58	-36
Row 12 (cont)	Screening 1	ABNORMAL, CLINICALLY INSIGNIFICANT		ABNORMAL, CLINICALLY INSIGNIFICANT			PRINCIPAL INVESTIGATOR	1	2003-04-15T11:58	-36

ECG Example 2:

Example 2 shows results for one subject across multiple visits where only the overall assessment was collected. In addition the ECG done at Visit 4 was read by the principal investigator and a cardiologist. In this example the EGGRPID is the same number and the EGSEQ increments by one.

Row 2 - The only one selected as Baseline.

Row 3 - Shows a date/time in ISO 8601 representation where both the date and time were collected.

	STUDYID	DOMAIN	USUBJID	EGSEQ	EGGRPID	EGTESTCD	EGTEST	EGPOS	EGORRES	EGSTRESC
Row 1	ABC	EG	ABC-99-CA-456	1	1	ECGINT	ECG Result	SUPINE	NORMAL	NORMAL
Row 2	ABC	EG	ABC-99-CA-456	2	2	ECGINT	ECG Result	SUPINE	ABNORMAL	ABNORMAL
Row 3	ABC	EG	ABC-99-CA-456	3	3	ECGINT	ECG Result	SUPINE	ABNORMAL	ABNORMAL
Row 4	ABC	EG	ABC-99-CA-456	4	4	ECGINT	ECG Result	SUPINE	ABNORMAL	ABNORMAL
Row 5	ABC	EG	ABC-99-CA-456	5	4	ECGINT	ECG Result	SUPINE	ABNORMAL	ABNORMAL

	EGSTRESN	EGBLFL	EGEVAL	VISIT	VISITNUM	VISITDY	EGDTC	EGDY
Row 1 (cont)			PI	SCREEN I	1	-2	2003-11-26	-2
Row 2 (cont)		Y	PI	SCREEN II	2	-1	2003-11-27	-1
Row 3 (cont)			PI	DAY 10	3	10	2003-12-07T09:02	10
Row 4 (cont)			PI	DAY 18	4	15	2003-12-12	15
Row 5 (cont)			Cardiologist	DAY 18	4	15	2003-12-12	15

9.4.2 IE Example

This example shows records for four subjects, each of whom had a single inclusion/exclusion exception. Subject XYZ-0007 met exclusion criteria number 17, but was included in the trial. The other three subjects each failed to meet inclusion criteria number 3, but were included in the trial.

	STUDYID	DOMAIN	USUBJID	IESEQ	IETESTCD	IETEST	IECAT	IESPID	IEORRES	IESTRESC
Row 1	XYZ	IE	XYZ-0007	1	EXCL17	Ventricular Rate	EXCLUSION	17	Y	Y
Row 2	XYZ	IE	XYZ-0023	1	INCL03	Acceptable mammogram from local radiologist?	INCLUSION	3	N	N
Row 3	XYZ	IE	XYZ-0047	1	INCL03	Acceptable mammogram from local radiologist?	INCLUSION	3	N	N
Row 4	XYZ	IE	XYZ-0096	1	INCL03	Acceptable mammogram from local radiologist?	INCLUSION	3	N	N

	VISIT	VISITNUM	VISITDY	IEDTC	IEDY
Row 1 (cont)	WEEK-8	1	-63	1999-01-10	-58
Row 2 (cont)	WEEK-8	1	-63	1999-01-06	-62
Row 3 (cont)	WEEK-8	1	-63	1999-01-12	-56
Row 4 (cont)	WEEK-8	1	-63	1999-01-13	-55

9.4.3 LB Examples

Example 1:

- Row 1** - Shows a value collected in one unit, but converted to selected standard unit.
- Rows 2, 3, 4** - Rows 2 and 3 show two records for Alkaline Phosphatase done at the same visit, one day apart. Row 4 shows how to create a derived record (average of the records 2 and 3) and flagged derived and as the one to use as baseline.
- Rows 6, 7** - Show a suggested use of the LBSCAT variable. It could be used to further classify types of tests within a laboratory panel (i.e., "DIFFERENTIAL").
- Row 9** - Shows the proper use of the LBSTAT variable to indicate "NOT DONE", where a reason was collected when a test was not done.
- Row 10** - Shows one subject's data collected at two different visits. For Visit 1 the subject had cholesterol measured. The normal range for this test is <200 mg/dL. The value <200 may not be used in the normal range variables LBORNRLHI or LBSTNRHI; however, a sponsor may decide, for example, to enter '0' into LBORNRLLO and '199' in LBORNRLHI. The sponsor must define the appropriate value for all of the normal range variables.

	STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBSCAT	LBORRES	LBORRESU	LBORNRLLO	LBORNRLHI
Row 1	ABC	LB	ABC-001-001	1	ALB	Albumin	Chemistry		34	g/L	35	50
Row 2	ABC	LB	ABC-001-001	2	ALKP	Alkaline Phosphatase	Chemistry		398	IU/L	40	160
Row 3	ABC	LB	ABC-001-001	3	ALKP	Alkaline Phosphatase	Chemistry		350	IU/L	40	160
Row 4	ABC	LB	ABC-001-001	4	ALKP	Alkaline Phosphatase	Chemistry			IU/L		
Row 5	ABC	LB	ABC-001-001	5	WBC	WBC Count	Hematology		5.9	10 ⁹ /L	4	11
Row 6	ABC	LB	ABC-001-001	6	LYMPH	Lymphocytes	Hematology	DIFFERENTIAL	6.7	%	25	40
Row 7	ABC	LB	ABC-001-001	7	NEUT	Neutrophils	Hematology	DIFFERENTIAL	5.1	10 ⁹ /L	2	8
Row 8	ABC	LB	ABC-001-001	8	PH	pH	Urinalysis		7.5		5.0	9.0
Row 9	ABC	LB	ABC-001-001	9	ALB	Albumin	Chemistry					
Row 10	ABC	LB	ABC-001-001	10	CHOL	Cholesterol	Chemistry		229	mg/dL	0	199
Row 11	ABC	LB	ABC-001-001	11	WBC	WBC Count	Hematology		5.9	10 ⁹ /L	4	11

	LBSTRESC	LBSTRESN	LBSTRESU	LBSTNRLO	LBSTNRHI	LBSTAT	LBREASND	LBBLFL	LBFAST	LBDRVFL	VISIT	VISITNUM	LBDTC
Row 1 (cont)	3.4	3.4	g/dL	3.5	5			Y	Y		BASELINE	1	1999-06-19
Row 2 (cont)	398	398	units/L	40	160				Y		BASELINE	1	1999-06-19
Row 3 (cont)	350	350	units/L	40	160				Y		BASELINE	1	1999-06-20

	LBSTRESC	LBSTRESN	LBSTRESU	LBSTNRLO	LBSTNRHI	LBSTAT	LBREASND	LBBLFL	LBFAST	LBDRVFL	VISIT	VISITNUM	LBDTC
Row 4 (cont)	374	374	units/L	40	160			Y	Y	Y	BASELINE	1	1999-06-19
Row 5 (cont)	5.9	5.9	10 ³ /uL	4	11			Y	Y		BASELINE	1	1999-06-19
Row 6 (cont)	0.4	0.4	10 ⁹ /L	1.2	3			Y	Y		BASELINE	1	1999-06-19
Row 7 (cont)	5.1	5.1	10 ⁹ /L	2	8			Y	Y		BASELINE	1	1999-06-19
Row 8 (cont)	7.5			5.00	9.00			Y	Y		BASELINE	1	1999-06-19
Row 9 (cont)						NOT DONE	INSUFFICIENT SAMPLE				VISIT 1	2	1999-07-21
Row10 (cont)	229	229	mg/dL	0	5.85						VISIT 1	2	1999-07-21
Row11 (cont)	5.9	5.9	10 ³ /uL	4	11				Y		VISIT 1	2	1999-07-21

Example 2:

This is an example of three sequential timed urine collections measuring glucose.

- Rows 1** - Shows an example of a pre-dose urine collection interval (from 4 hours prior to dosing until 15 minutes prior to dosing) with a negative value for LBELTM that reflects the end of the interval in reference to the fixed reference LBTPTRF, the date of which is recorded in LBRFTDTC. --RFTDTC is a new timing variable added to version 1.1 of the SDTM, and is incorporated in SDTM 3.1.1 to meet this purpose.
- Rows 2, 3** - Show an example of post-dose urine collection intervals with values for LBELTM that reflect the end of the intervals in reference to the fixed reference LBTPTRF, the date of which is recorded in LBRFTDTC.

	STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBORRES	LBORRESU	LBORNRL0	LBORNRI
Row 1	ABC	LB	ABC-001-001	1	GLUCOSE	Glucose	Urine	7	mg/dL	1	15
Row 2	ABC	LB	ABC-001-001	2	GLUCOSE	Glucose	Urine	11	mg/dL	1	15
Row 3	ABC	LB	ABC-001-001	3	GLUCOSE	Glucose	Urine	9	mg/dL	1	15

	LBSTRESC	LBSTRESN	LBSTRESU	LBSTNRLO	LBSTNRHI	VISIT	VISITNUM
Row 1 (cont)	0.38	0.38	mmol/L	0.1	0.8	BASELINE	1
Row 2 (cont)	0.61	0.61	mmol/L	0.1	0.8	BASELINE	1
Row 3 (cont)	0.5	0.5	mmol/L	0.1	0.8	BASELINE	1

	LBDC	LBENDC	LBTP	LBTPNUM	LBELTM	LBTPTRF	LBRFTDTC
Row 1 (cont)	1999-06-19T04:00	1999-06-19T07:45	Pre-dose	1	-P15M	Dosing	1999-06-19T08:00
Row 2 (cont)	1999-06-19T08:00	1999-06-19T16:00	0-8 hours after dosing	2	P8H	Dosing	1999-06-19T08:00
Row 3 (cont)	1999-06-19T16:00	1999-06-20T00:00	8-16 hours after dosing	3	P16H	Dosing	1999-06-19T08:00

9.4.4 PE Example

The example shows data for one subject collected at two different visits. In all of the records except 6 and 11 the data comes from the general physical examination. In this case PECAT is "General". Additional data collected on about an ophthalmologic examination is also added to this domain, refer to rows eight and thirteen.

- Row 1** - Shows how PESTRESC is populated if result is 'NORMAL'.
- Row 2** - Shows the proper use of the STAT variable to indicate "NOT DONE", and when PEREASND is used to indicate why a body system (PETEST) was not examined.
- Row 4-6** - Show how PESPID is used to show the sponsor-defined identifier, which in this case is a sequence number used for identifying abnormalities within a body system.
- Row 4-7** - Show how PESTRESC is populated if abnormality is dictionary coded.
- Rows 8, 13** - Show how the PECAT variable can be used to indicate a different type of physical examination. In this case, the ophthalmologic examination may have been collected in a separate dataset in the operational database.

	STUDYID	DOMAIN	USUBJID	PESEQ	PETESTCD	PETEST	PECAT	PELOC	PEBODSYS	PEORRES
Row 1	ABC	PE	ABC-001-001	1	HEAD	Head	GENERAL			NORMAL
Row 2	ABC	PE	ABC-001-001	2	RESP	Respiratory	GENERAL			
Row 3	ABC	PE	ABC-001-001	3	ENT	Ear/nose/throat	GENERAL			NORMAL
Row 4	ABC	PE	ABC-001-001	4	SKIN	Skin	GENERAL	FACE	SKIN AND	ACNE
Row 5	ABC	PE	ABC-001-001	5	SKIN	Skin	GENERAL	HANDS	SKIN AND	ALLERGIC REACTION
Row 6	ABC	PE	ABC-001-001	6	SKIN	Skin	GENERAL	LEFT ARM	SKIN AND	SKINRASH
Row 7	ABC	PE	ABC-001-001	7	CARDIO	Cardiovascular	GENERAL		INVESTIGATIONS	SYSTOLIC
Row 8	ABC	PE	ABC-001-001	8	FUNDOSCP	Fundoscopic	OPHTALMOLOGIC			NORMAL
Row 9	ABC	PE	ABC-001-001	9	RESP	Respiratory	GENERAL			NORMAL
Row 10	ABC	PE	ABC-001-001	10	ENT	Ear/nose/throat	GENERAL			NORMAL
Row 11	ABC	PE	ABC-001-001	11	NECK	Neck	GENERAL			NORMAL
Row 12	ABC	PE	ABC-001-001	12	CARDIO	Cardiovascular	GENERAL			NORMAL
Row 13	ABC	PE	ABC-001-001	13	FUNDOSCP	Fundoscopic	OPHTALMOLOGIC			NORMAL

	PESTRESC	PESEV	PESTAT	PEREASND	PESPID	VISIT	VISITNUM	VISITDY	PEDTC	PEDY
Row 1 (cont)	NORMAL				1	BASELINE	1	1	1999-06-06	-3
Row 2 (cont)			NOT DONE	INVESTIGATOR ERROR	1	BASELINE	1	1	1999-06-06	-3
Row 3 (cont)	NORMAL				1	BASELINE	1	1	1999-06-06	-3
Row 4 (cont)	ACNE NOS	MILD			1	BASELINE	1	1	1999-06-06	-3
Row 5 (cont)	DERMATITIS	MODERATE			2	BASELINE	1	1	1999-06-07	-2
Row 6 (cont)	RASH	SEVERE			3	BASELINE	1	1	1999-06-08	-1
Row 7 (cont)	CARDIAC				1	BASELINE	1	1	1999-06-06	-3
Row 8 (cont)	NORMAL				1	BASELINE	1	1	1999-06-06	-3
Row 9 (cont)	NORMAL				1	VISIT 1	2	45	1999-07-21	45
Row 10 (cont)	NORMAL				1	VISIT 1	2	45	1999-07-21	45
Row 11 (cont)	NORMAL				1	VISIT 1	2	45	1999-07-21	45
Row 12 (cont)	NORMAL				1	VISIT 1	2	45	1999-07-21	45
Row 13 (cont)	NORMAL				1	VISIT 1	2	45	1999-07-21	45

9.4.5 QS Examples

Example 1:

This is an example of data from a questionnaire from one subject at one visit with standard text answers that have an associated score. In this example the subject answered all of the questions in Rows 1-4 and Rows 7-9. The standard text (e.g., very good) translates to a score of 4.4. The value of 4.4 is populated in both QSSTRESN and QSSTRESC. Since this is the baseline data there is a flag in all records in QSBLFL. The values in Rows 5, 6, 10, and 11 are derived from previous records and are flagged with a Y in QSDRVFL. The example shows how the textual answer is handled in the QSORRES Variable, while the QSSTRESC and QSSTRESN contain the standardized score value.

Rows 5, 6, 10, 11 - Show derived records. Notice how QSORRES is blank for derived records because there is no corresponding text value for the numeric value shown (see section 4.1.5.1).

	STUDYID	DOMAIN	USUBJID	QSSEQ	QSTESTCD	QSTEST	QSCAT	QSSCAT
Row 1	STUDYX	QS	P0001	1	GH1	Heath	SF36	GENERAL HEALTH
Row 2	STUDYX	QS	P0001	2	GH11A	Sick a litter easier	SF36	GENERAL HEALTH
Row 3	STUDYX	QS	P0001	3	GH11B	Healthy as anybody	SF36	GENERAL HEALTH
Row 4	STUDYX	QS	P0001	4	GH11C	Expect health to get worse	SF36	GENERAL HEALTH
Row 5	STUDYX	QS	P0001	5	GH	SF-36 General health perceptions	SF36	GENERAL HEALTH
Row 6	STUDYX	QS	P0001	6	GHINDEX	SF-36 General health perceptions (0-100)	SF36	GENERAL HEALTH
Row 7	STUDYX	QS	P0001	7	RP4A	Phys. Health-cut down time spent	SF36	ROLE-PHYSICAL
Row 8	STUDYX	QS	P0001	8	RP4B	Phys. Health-accomplished less	SF36	ROLE-PHYSICAL
Row 9	STUDYX	QS	P0001	9	RP4C	Phys. Health-limit kind of work	SF36	ROLE-PHYSICAL
Row 10	STUDYX	QS	P0001	10	RP	SF-36 Role-physical	SF36	ROLE-PHYSICAL
Row 11	STUDYX	QS	P0001	11	RPINDEX	SF-36 Role-physical (0-100)	SF36	ROLE-PHYSICAL

	QSORRES	QSSTRESN	QSSTRESC	QSBLFL	QSDRVFL	VISITNUM	VISIT	QSDTC	QSDY
Row 1 (cont)	VERY GOOD	4.4	4.4	Y		2	BASELINE	2001-03-28	-2
Row 2 (cont)	MOSTLY FALSE	4	4	Y		2	BASELINE	2001-03-28	-2
Row 3 (cont)	MOSTLY TRUE	4	4	Y		2	BASELINE	2001-03-28	-2
Row 4 (cont)	DEFINITELY FALSE	5	5	Y		2	BASELINE	2001-03-28	-2
Row 5 (cont)		21.4	21.4	Y	Y	2	BASELINE	2001-03-28	-2
Row 6 (cont)		82	82	Y	Y	2	BASELINE	2001-03-28	-2
Row 7 (cont)	NO	2	2	Y		2	BASELINE	2001-03-28	-2
Row 8 (cont)	NO	2	2	Y		2	BASELINE	2001-03-28	-2
Row 9 (cont)	NO	2	2	Y		2	BASELINE	2001-03-28	-2
Row 10 (cont)		8	8	Y	Y	2	BASELINE	2001-03-28	-2
Row 11 (cont)		100	100	Y	Y	2	BASELINE	2001-03-28	-2

Example 2:

This example show data from one subject collected at one visit for a questionnaire with standard text answers. These answers cannot be converted to numeric type, so only QSSTRESC is populated. The derived record, however, does have a derived value in QSSTRESN. Notice how QSTPTNUM is used to organize the same question being asked at various time points on the same date where no time was collected. In rows 1-3 QSTEST is ARM and QSTPTNUM is incremented 1,2,3. When QSTEST changes to BUTTER the number in QSTPTNUM start again at 1.

Row 11 - Shows derived records. Notice how QSORRES is blank for derived records because there is no corresponding text value for the numeric value shown (see section 4.1.5.1).

	STUDYID	DOMAIN	USUBJID	QSSEQ	QSTESTCD	QSTEST	QSCAT	QSSCAT	QSORRES	QSSTRESC
Row 1	STUDYX	QS	P0001	1	COG01T02	ARM	ADAS	WORD RECALL	NO	NO
Row 2	STUDYX	QS	P0001	2	COG01T02	ARM	ADAS	WORD RECALL	NO	NO
Row 3	STUDYX	QS	P0001	3	COG01T02	ARM	ADAS	WORD RECALL	NO	NO
Row 4	STUDYX	QS	P0001	4	COG01T03	BUTTER	ADAS	WORD RECALL	NO	NO
Row 5	STUDYX	QS	P0001	5	COG01T03	BUTTER	ADAS	WORD RECALL	NO	NO
Row 6	STUDYX	QS	P0001	6	COG01T03	BUTTER	ADAS	WORD RECALL	NO	NO
Row 7	STUDYX	QS	P0001	7	COG01T04	CABIN	ADAS	WORD RECALL	NO	NO
Row 8	STUDYX	QS	P0001	8	COG01T04	CABIN	ADAS	WORD RECALL	NO	NO
Row 9	STUDYX	QS	P0001	9	COG01T04	CABIN	ADAS	WORD RECALL	NO	NO
Row 10	STUDYX	QS	P0001	10	COG01T09	GRASS	ADAS	WORD RECALL	NO	NO
Row 11	STUDYX	QS	P0001	11	COG01X	WORD RECALL	ADAS	WORD RECALL		9

	QSSTRESN	QSBLFL	QSDRVFL	VISIT	VISITNUM	VISITYDY	QSDTC	QSDY	QSTPTNUM
Row 1 (cont)				SCREENING	1	-14	2001-03-20	-10	1
Row 2 (cont)				SCREENING	1	-14	2001-03-20	-10	2
Row 3 (cont)				SCREENING	1	-14	2001-03-20	-10	3
Row 4 (cont)				SCREENING	1	-14	2001-03-20	-10	1
Row 5 (cont)				SCREENING	1	-14	2001-03-20	-10	2
Row 6 (cont)				SCREENING	1	-14	2001-03-20	-10	3
Row 7 (cont)				SCREENING	1	-14	2001-03-20	-10	1
Row 8 (cont)				SCREENING	1	-14	2001-03-20	-10	2
Row 9 (cont)				SCREENING	1	-14	2001-03-20	-10	3
Row 10 (cont)				SCREENING	1	-14	2001-03-20	-10	1
Row 11 (cont)	9		Y	SCREENING	1	-14	2001-03-20	-10	99

9.4.6 SC Example

The example below shows data that is collected once per subject that does not change during the trial and does not fit into the demographics domain. For this example the eye color and initials were collected.

Rows 1, 4 illustrate use of an additional Race response. In this case, the value of RACE in DM would be 'OTHER'.

	STUDYID	DOMAIN	USUBJID	SCSEQ	SCTESTCD	SCTEST	SCORRES	SCSTRESC	VISIT	VISITNUM	SCDTC
Row 1	ABC	SC	ABC-001-001	1	RACEOTH	Race Other	JAPANESE	JAPANESE	Baseline	1	1999-06-19
Row 2	ABC	SC	ABC-001-001	2	EYECD	Eye Color	BROWN	BROWN	Baseline	1	1999-06-19
Row 3	ABC	SC	ABC-001-001	3	SUBJINIT	Subject Initials	HLT	HLT	Baseline	1	1999-06-19
Row 4	ABC	SC	ABC-001-002	1	RACEOTH	Race Other	ALASKA NATIVE	ALASKA NATIVE	Baseline	1	1999-03-19
Row 5	ABC	SC	ABC-001-002	2	EYECD	Eye Color	BLUE	BLUE	Baseline	1	1999-03-19
Row 6	ABC	SC	ABC-001-002	3	SUBJINIT	Subject Initials	BAM	BAM	Baseline	1	1999-03-19
Row 7	ABC	SC	ABC-001-003	2	EYECD	Eye Color	GREEN	GREEN	Baseline	1	1999-05-03
Row 8	ABC	SC	ABC-001-003	3	SUBJINIT	Subject Initials	ALM	ALM	Baseline	1	1999-05-03
Row 9	ABC	SC	ABC-002-001	2	EYECD	Eye Color	HAZEL	HAZEL	Baseline	1	1999-06-14
Row 10	ABC	SC	ABC-002-001	3	SUBJINIT	Subject Initials	CQH	CQH	Baseline	1	1999-06-14

9.4.7 VS Example

The example below shows one subject with two visits, Baseline and Visit 1, including examples of both collected and derived baseline measurements.

Rows 1-2, 4-5, 8, 9 -VSTPT and VSTPTNUM are populated since more than one measurement was taken at this visit.

Rows 3, 6 - Show an example of a derived value that was not considered to be an original result. In this case the sponsor derived the value in a different variable in the operational database. VSTPT and VSTPTNUM are not populated for these derived records.

Rows 8, 9 - Show two temperatures taken at the baseline visit. Row 9 has a "Y" in the VSBLFL to indicate it was used as the baseline measurement.

Row 14 - Shows a value collected in one unit, but converted to selected standard unit.

Row 15 - Shows the proper use of the STAT variable to indicate "NOT DONE" where a reason was collected when a test was not done.

CDISC SDTM Implementation Guide (Version 3.1.1)

	STUDYID	DOMAIN	USUBJID	VSSEQ	VSTESTCD	VSTEST	VSPOS	VSLOC	VSORRES	VSORRESU	VSSTRESC
Row 1	ABC	VS	ABC-001-001	1	SYSBP	Systolic Blood Pressure	SITTING	LEFT ARM	154	mmHg	154
Row 2	ABC	VS	ABC-001-001	2	SYSBP	Systolic Blood Pressure	SITTING	LEFT ARM	152	mmHg	152
Row 3	ABC	VS	ABC-001-001	3	SYSBP	Systolic Blood Pressure	SITTING	LEFT ARM			153
Row 4	ABC	VS	ABC-001-001	4	DIABP	Diastolic Blood Pressure	SITTING	LEFT ARM	44	mmHg	44
Row 5	ABC	VS	ABC-001-001	5	DIABP	Diastolic Blood Pressure	SITTING	LEFT ARM	48	mmHg	48
Row 6	ABC	VS	ABC-001-001	6	DIABP	Diastolic Blood Pressure	SITTING	LEFT ARM			46
Row 7	ABC	VS	ABC-001-001	7	PULSE	Pulse	SITTING	LEFT ARM	72	bpm	72
Row 8	ABC	VS	ABC-001-001	8	TEMP	Temperature		MOUTH	34.7	C	34.7
Row 9	ABC	VS	ABC-001-001	9	TEMP	Temperature		MOUTH	36.2	C	36.2
Row 10	ABC	VS	ABC-001-001	10	WEIGHT	Weight	STANDING		90.5	kg	90.5
Row 11	ABC	VS	ABC-001-001	11	HEIGHT	Height	STANDING		157	cm	157
Row 12	ABC	VS	ABC-001-001	12	SYSBP	Systolic Blood Pressure	SITTING	LEFT ARM	95	mmHg	95
Row 13	ABC	VS	ABC-001-001	13	DIABP	Diastolic Blood Pressure	SITTING	LEFT ARM	44	mmHg	44
Row 14	ABC	VS	ABC-001-001	14	TEMP	Temperature		MOUTH	97.16	F	36.2
Row 15	ABC	VS	ABC-001-001	15	WEIGHT	Weight					

	VSSTRESN	VSSTRESU	VSSTAT	VSREASND	VSBLFL	VSDRVFL	VISIT	VISITNUM	VSTPT	VSTPTNUM	VISITDY
Row 1 (cont)	154	mmHg					BASELINE	1	BASELINE 1	1	1
Row 2 (cont)	152	mmHg					BASELINE	1	BASELINE 2	2	1
Row 3 (cont)	153	mmHg			Y	Y	BASELINE	1			1
Row 4 (cont)	44	mmHg					BASELINE	1	BASELINE 1	1	1
Row 5 (cont)	48	mmHg					BASELINE	1	BASELINE 2	2	1
Row 6 (cont)	46	mmHg			Y	Y	BASELINE	1			1
Row 7 (cont)	72	bpm			Y		BASELINE	1			1
Row 8 (cont)	34.7	C					BASELINE	1	BASELINE 1	1	1
Row 9 (cont)	36.2	C			Y		BASELINE	1	BASELINE 2	2	1
Row 10 (cont)	90.5	kg			Y		BASELINE	1			1
Row 11 (cont)	157	cm			Y		BASELINE	1			1
Row 12 (cont)	95	mmHg					VISIT 1	2			35
Row 13 (cont)	44	mmHg					VISIT 1	2			35
Row 14 (cont)	36.2	C					VISIT 1	2			35
Row 15 (cont)			NOT DONE	Subject refused			VISIT 1	2			35

9.5 TRIAL DESIGN EXAMPLES

9.5.1 Defining Epochs - a How-to Example Using the Example Crossover Study

This example is based on the crossover study shown in Figure 4 of [Section 7.4.2](#).

Users may find it useful to construct a "study design matrix" like the following.

	Trial Screen	1st Treatment	1st Rest	2nd Treatment	2nd Rest	3rd Treatment	Trial Follow-up
P-5-10	Screen	Placebo	Rest	5 mg	Rest	10 mg	Follow-up
5-P-10	Screen	5 mg	Rest	Placebo	Rest	10 mg	Follow-up
5-10-P	Screen	5 mg	Rest	10 mg	Rest	Placebo	Follow-up

This table is not submitted as part of the SDTM. It, like the figure in [Section 7](#), is purely an aid to the user in working out the trial design datasets. Note that the row labels (in bold type) are the names of the Arms. The column headers (also in bold type) are the names of the trial Epochs. The contents of the cells are Elements (plain type). In many cases, the user will fill in the column headers first, then place the Elements in the cells of the table. In other cases, the user may find it easier to work out the Elements in each Arm, then to fill in the names of the Epochs.

The Trial Arms table contains a record for each Element in each Arm (row of the study design matrix). TAETORD gives the order of the Elements within the Arm. The study design matrix table does not explicitly record the branch – a diagram is better for displaying that feature of the trial design.

Example of a full Trial Arms dataset for a crossover study:

ARM	TAETORD	ELEMENT	TABRANCH
P-5-10	1	Screen	Randomized to Placebo - 5 mg - 10 mg
P-5-10	2	Placebo	
P-5-10	3	Rest	
P-5-10	4	5 mg	
P-5-10	5	Rest	
P-5-10	6	10 mg	
P-5-10	7	Follow-up	
5-P-10	1	Screen	Randomized to 5 mg/Placebo/10 mg
5-P-10	2	5 mg	
5-P-10	3	Rest	
5-P-10	4	Placebo	
5-P-10	5	Rest	
5-P-10	6	10 mg	
5-P-10	7	Follow-up	
5-10-P	1	Screen	Randomized to 5 mg/10 mg/Placebo
5-10-P	2	5 mg	
5-10-P	3	Rest	
5-10-P	4	10 mg	
5-10-P	5	Rest	
5-10-P	6	Placebo	
5-10-P	7	Follow-up	

The Trial Elements dataset has a record for each unique Element. The properties of the Elements are not recorded in the trial design matrix, but the trial design matrix can be used to check that all Elements have been included in the Trial Elements dataset. The following information may be helpful in working out the start and end rules of the Elements:

- TESTRL answers the question, "What happens that marks the start of the Element?" or, "How do you know the subject has entered this Element?"
- TEENRL answers the question, "When should the subject leave this Element?" Note that this rule concerns when the subject should leave the Element. The subject actually leaves the Element when they enter the next Element.
- TEDUR is used when the element end rule says the subject should leave the Element after a fixed period of time. TEDUR records this fixed period of time as a duration using ISO 8601 notation. If the element end rule is of some different form, for example if it involves satisfying certain conditions, then TEDUR is left blank.

Example of the Trial Elements dataset for the same crossover study:

ELEMENT	TESTRL	TEENRL	TEDUR
Screen	Informed consent	2 weeks after start of element	P14D
Placebo	First dose of placebo	2 weeks after start of element	P14D
5 mg	First dose of 5 mg drug	2 weeks after start of element	P14D
10 mg	First dose of 10 mg drug	2 weeks after start of element	P14D
Rest	48 hrs after last dose drug	1 week after start of element	P7D
Follow-up	48 hrs after last dose drug	3 weeks after start of element	P21D

The Trial Elements dataset contains information on timing (the start and end rules), while the Trial Arms dataset contains information on how the Elements are assembled to make the Arms. Users may find it useful to join the two datasets (merge on Element) to display all this information in one table.

The table below shows such a join of Trial Arms and Trial Elements. A column for EPOCH has been included in this table. Sorting this dataset by TAETORD, then ARM, makes it easy to check for consistency of the start and end rules of the Elements within an Epoch, across different Arms. Note that in this example, for each group of records with the same value of TAETORD and EPOCH, the Start Rules are the same or similar. For the Treatment Epochs, the start rules are "similar" in that they differ only in the specific dose (Placebo, 5 mg or 10 mg) that starts the element. The end rules for the Treatment Epochs are the same. For all the other Epochs in this trial, all the Elements within the Epoch are identical, so the start and end rules are also identical.

If a group of Elements that have been labeled as belonging to the same Epoch do not have similar start and end rules, then they probably do not really constitute a valid Epoch.

For a blinded trial, it may be useful to think about a "blinded start rule" or "epoch start rule." While the trial is still blinded, participants will not really know whether they have entered a particular treatment Element, only that they have entered a particular Epoch. For example, the element start rules for the 1st Trt Epoch specify receiving a particular dose (placebo, 5 mg or 10 mg), but during the trial, participants will know only that they have received the dose for that Epoch. Un-blinding the study reveals which treatment the subject received, and therefore which Element they passed through.

Join of Trial Arms and Trial Elements datasets (merged by Element), sorted by TAETORD, then ARM:

ARM	TAETORD	ELEMENT	TASTRL	TAENRL	TADUR	EPOCH	TABRANCH
5-P-10	1	Screen	Informed consent	2 weeks after start of element	P14D	Trial Screen	Randomized to 5 mg/Placebo/10 mg
P-5-10	1	Screen	Informed consent	2 weeks after start of element	P14D	Trial Screen	Randomized to Placebo/5 mg/10 mg
5-10-P	1	Screen	Informed consent	2 weeks after start of element	P14D	Trial Screen	Randomized to 5 mg/10 mg/Placebo
5-P-10	2	5 mg	First dose of 5 mg drug	2 weeks after start of element	P14D	1 st Trt	
P-5-10	2	Placebo	First dose of placebo	2 weeks after start of element	P14D	1 st Trt	
5-10-P	2	5 mg	First dose of 5 mg drug	2 weeks after start of element	P14D	1 st Trt	
5-P-10	3	Rest	48 hrs after last dose drug	1 week after start of element	P7D	1 st Rest	
P-5-10	3	Rest	48 hrs after last dose drug	1 week after start of element	P7D	1 st Rest	
5-10-P	3	Rest	48 hrs after last dose drug	1 week after start of element	P7D	1 st Rest	
5-P-10	4	Placebo	First dose of placebo	2 weeks after start of element	P14D	2 nd Trt	
P-5-10	4	5 mg	First dose of 5 mg drug	2 weeks after start of element	P14D	2 nd Trt	
5-10-P	4	10 mg	First dose of 10 mg drug	2 weeks after start of element	P14D	2 nd Trt	
5-P-10	5	Rest	48 hrs after last dose drug	1 week after start of element	P7D	2 nd Rest	
P-5-10	5	Rest	48 hrs after last dose drug	1 week after start of element	P7D	2 nd Rest	
5-10-P	5	Rest	48 hrs after last dose drug	1 week after start of element	P7D	2 nd Rest	
5-P-10	6	10 mg	First dose of 10 mg drug	2 weeks after start of element	P14D	3 rd Trt	
P-5-10	6	10 mg	First dose of 10 mg drug	2 weeks after start of element	P14D	3 rd Trt	
5-10-P	6	Placebo	First dose of placebo	2 weeks after start of element	P14D	3 rd Trt	
5-P-10	7	Follow-up	48 hrs after last dose drug	3 weeks after start of element	P21D	Trial FU	
P-5-10	7	Follow-up	48 hrs after last dose drug	3 weeks after start of element	P21D	Trial FU	
5-10-P	7	Follow-up	48 hrs after last dose drug	3 weeks after start of element	P21D	Trial FU	

9.5.2 Trial Summary Example

Trial Summary Example1:

Rows 1 and 2: This trial is of two different types so there are 2 records for the type of trial

Rows 1,2,5,6,7 and 12 use controlled terminology for TSVAL

	STUDYID	DOMAIN	TSSEQ	TSPARMCD	TSPARM	TSVAL
Row 1	XYZ	TS	1	TYPE	Type Of Trial	EFFICACY
Row 2	XYZ	TS	2	TYPE	Type Of Trial	SAFETY
Row 3	XYZ	TS	1	TITLE	Trial Title	A 24 Week Study of Daily Oral Investigational Drug vs. Placebo in Subjects with Asthma
Row 4	XYZ	TS	1	OBJPRIM	Primary Objective Of Trial	Reduce the incidence of exacerbations of asthma
Row 5	XYZ	TS	1	RANDOM	Trial Is Randomized	Y
Row 6	XYZ	TS	1	MASK	Trial Blinding Schema	DOUBLE BLIND
Row 7	XYZ	TS	1	DESIGN	Description of Trial Design	PARALLEL
Row 8	XYZ	TS	1	TRT	Reported Name Of Test Product	Investigational New Drug
Row 9	XYZ	TS	1	INDIC	Trial Indications	Asthma
Row 10	XYZ	TS	1	AGEMIN	Planned Minimum Age Of Subjects	18
Row 11	XYZ	TS	1	AGEMAX	Planned Maximum Age Of Subjects	70
Row 12	XYZ	TS	1	SEX	Sex Of Participants	BOTH
Row 13	XYZ	TS	1	COMPTRT	Comparative Treatment Name	Placebo

10 Appendices

10.1 CDISC SDS TEAM

Role	Name	Company
Team Lead	Wayne Kubick	Lincoln Technologies, CDISC
Team Co-Leads	Fred Wood	Procter and Gamble
	Mary Lenzen	Pfizer
	Tom Guinter	Octagon Research
Technical Manager	Julie Evans	CDISC
Sub-team Leads	Randall Austin	GlaxoSmithKline
	Chris Tolk	Bayer Healthcare
	Madhavi Vemuri	Johnson and Johnson PRD
	Diane Wold	GlaxoSmithKline
Team Participants	Karen Alexander	Boehringer-Ingelheim
	Gayle Amato	Bayer Healthcare
	Jane Boone	Webclin
	Gary Cunningham	AstraZeneca
	Kathleen Giordano	Merck
	Dan Godoy	AstraZeneca
	Andreas Gromen	Schering AG
	Joyce Hernandez	IBM
	Sandy Lei	Johnson and Johnson PRD
	Musa Nsereko	Cephalon
	Bill Qubeck	Pfizer
	Robert Sarate	Merck
	Sarah Shiue	Merck
	Gail Stoner	Centocor
	Mike Walega	Covance
	Gary Walker	Quintiles
	Yuguang Zhao	Sanofi-Aventis
FDA Participants	Randy Levin	FDA
	Norman Stockbridge	FDA
	Steve Wilson	FDA

10.2 GLOSSARY OF TERMS

The following abbreviations and terms are used in this document. Additional definitions can be found in the CDISC Glossary available at <http://www.cdisc.org/glossary/index.html>.

CDISC	Clinical Data Interchange Standards Consortium
CRT	Case Report Tabulation
Dataset	A collection of structured data in a single file
Domain	A collection of observations with a topic-specific commonality about a subject
FDA	Food and Drug Administration
HL7	Health Level 7
SDS	Submission Data Standards (Team)
SDTMIG	"Study Data Tabulation Model Implementation Guide for Human Clinical Trials (this document)"
SEND	Standard for Exchange of Non-clinical Data
SDSIG	SDS Implementation Guide V. 3.1, now referred to as SDTMIG.
SDTM	Study Data Tabulation Model
V3.x	Refers to Version 3.1 of the SDSIG and all subsequent versions of the SDTMIG

10.3 STANDARDIZED AND RESERVED CODES

This appendix includes a preliminary list of standard codes that have been defined by the CDISC SDS Team for use with the SDTM. It includes a set of reserved domain codes, test codes for two findings domains, and an initial set of name codes for use in the Supplemental Qualifiers (SUPPQUAL) special-purpose dataset.

Since a robust set of controlled terminology is so essential for effective use of the SDTM and other CDISC models, CDISC has assigned a new "Controlled Terminology" team. Beginning in March 2005, CDISC will be publishing sets of proposals for controlled terminology in separate documentation packages that will be available through the CDISC website for review and comment. The set of controlled terminology will be expanded over time and will eventually be available through online search tools provided by the National Cancer Institute of the United States National Institutes of Health.

10.3.1 Reserved Domain Codes

The following domain codes have been reserved for use with the domain topicalities listed. CDISC will be preparing additional domain models to describe many of these over time.

Code	Domain	Class	Description
AE	Adverse Events	Events	See 6.2.1.1, assumption 1.
AU	Autopsy	Findings	Additional data about deaths, specific to findings from autopsies.
BM	Bone Mineral Density (BMD) Data	Findings	Findings resulting from evaluations of bone mineral density.
BR	Biopsy	Findings	Findings resulting from biopsies.
CM	Concomitant Meds	Interventions	See 6.1.1.1, assumption 1.
CO	Comments	Special Purpose	See 5.1.2.
DA	Drug Accountability	Findings	Data regarding the accountability of study drug, such as information on the receipt, dispensing, return, and packaging.
DC	Disease Characteristics	Findings	Data related to the characteristics of a specific disease state under study for a subject.
DM	Demographics	Special Purpose	Demographics includes a set of essential standard variables that describe each subject in a clinical study. It is the parent domain for all other observations for human clinical subjects. See SDTM 2.2.6.
DS	Disposition	Events	See 6.2.2.1, assumption 1.
DV	Protocol Deviations	Events	See separate DV domain model, assumption 1.
EE	EEG	Findings	Findings resulting from electroencephalogram (EEG) tests.
EG	ECG	Findings	Findings related to the collection of ECG data, including position of the subject, method of evaluation, all cycle measurements and all findings from the ECG including an overall interpretation if collected or derived.
EX	Exposure	Interventions	See 6.1.2.1, assumption 1.
HU	Healthcare Resource Utilization	Findings	Healthcare resource utilization data such as subject visits to physicians, hospitalizations, and nursing home stays.
IE	Inclusion / Exclusion	Findings	See 6.3.2.1, assumption 1.

Code	Domain	Class	Description
IM	Imaging	Findings	Findings resulting from diagnostic medical imaging tests such as X-ray or MRI.
LB	Laboratory Data	Findings	Laboratory test findings including, but is not limited to hematology, clinical chemistry and urinalysis data. Does not include microbiology or PK data, which are stored in separate domains.
MB	Microbiology Specimens	Findings	Microbiology Specimen findings, including gram stain results, and organisms found.
MH	Medical History	Events	See 6.2.3.1, assumption 1
ML	Meal Data	Interventions	Information regarding the subject's meal consumption, such as fluid intake, amounts, form (solid or liquid state), frequency, etc., typically used for PK analysis.
MS	Microbiology Susceptibility	Findings	Microbiology Susceptibility Test results, plus results of any other organism-related tests.
OM	Organ Measurements	Findings	Findings from organ measurement evaluations.
PC	PK Concentration	Findings	Concentrations of drugs/metabolites in fluids or tissues as a function of time.
PE	Physical Exam	Findings	Findings collected during a physical examination of the subject. It has findings that are discovered that are related to body systems. Does not include vital signs measurements, which are stored in the VS domain.
PP	PK Parameters	Findings	Pharmacokinetic parameters derived from pharmacokinetic concentration-time (PC) data.
PG	Pharmacogenomics	Findings	Pharmacogenomics findings initially focusing on Genotype and Gene Expression data.
QS	Questionnaires	Findings	See 6.3.5.1, assumption 1
SC	Subject Characteristics	Findings	See 6.3.6.1, assumption 1
SE	Subject Elements	Study Design	See SDTM 3.3.
SG	Surgery	Interventions	Surgical information.
SK	Skin Test	Findings	Findings from diagnostic skin tests.
SL	Sleep (Polysomnography) Data	Findings	Findings from diagnostic sleep tests (polysomnography).
	Signs and Symptoms	Findings/Events	Evidence of disease or condition, including objective signs and subjective symptoms that are typically observed by a physician or described to the investigator by the subject.
ST	Stress (Exercise) Test Data	Findings	Findings from exercise stress tests.
SU	Substance Use	Interventions	See 6.1.3.1, assumption 1
SV	Subject Visits	Trial Design	See SDTM 3.3.
TA	Trial Arms	Trial Design	See SDTM 3.2.
TE	Trial Elements	Trial Design	See SDTM 3.2.
TI	Trial Inclusion/Exclusion Criteria	Trial Design	See SDTM 3.4.
TS	Trial Summary	Trial Design	See SDTM 3.5.
TV	Trial Visits	Trial Design	See SDTM 3.2.
VS	Vital Signs	Findings	Measurements including but not limited to blood pressure, temperature, respiration, body surface area, BMI, height and weight.

10.3.2 Electrocardiogram Test Codes (for measured or calculated parameters)

The following table contains the test codes suggested by CDISC for use in Electrocardiogram data domains.

EGTESTCD	EGTEST
HR	Heart Rate
INTP	ECG Interpretation
PR	PR Interval
QRS	QRS Interval
QT	QT Interval
QTCB	QTc Interval Bazett
QTCF	QTc Interval Fridericia
RR	RR Interval
VRATE	Ventricular Rate

10.3.3 Vital Signs Test Codes

The following table contains the test codes suggested by CDISC for use in Vital Signs domains.

VSTESTCD	VSTEST
BMI	Body Mass Index
BODYFAT	Body Fat
BSA	Body Surface Area
DIABP	Diastolic Blood Pressure
FRMSIZE	Frame Size
HEIGHT	Height
HR	Heart Rate
MAP	Mean Arterial Pressure
PULSE	Pulse Rate
RESP	Respiratory Rate
SYSBP	Systolic Blood Pressure
TEMP	Temperature
WEIGHT	Weight

10.3.4 Supplemental Qualifiers Name Codes

The following table contains an initial set of standard name codes suggested by CDISC for use in the Supplemental Qualifiers (SUPQUAL) special-purpose dataset.

QNAM	QLABEL
AESOSP	Other Medically Important SAE (includes the descriptive text, not a flag)
AETRTEM	Treatment Emergent Flag
COMPLT	Completers Population Flag
FULLSET	Full Analysis Set Flag
ITT	Intent to Treat Population Flag
PPROT	Per Protocol Set Flag
SAFETY	Safety Population Flag

10.3.5 Trial Summary Codes

The following table contains **proposed** controlled terminology for use in the Trial Summary domain. Although this is the most current set of terminology available as of this writing, it is likely to grow or change as it undergoes further review by the CDISC Controlled Terminology team.

TSPARMCD	TSPARM	TSVAL
ADDON	TEST PRODUCT IS ADDED ON TO EXISTING TREATMENT	Y N
AEDICT	ADVERSE EVENT DICTIONARY	MedDRA WHOART SPONSOR DEFINED
AGESPAN	AGE SPAN	IN UTERO PRETERM NEWBORN INFANTS NEWBORN (0-27 DAYS) INFANT AND TODDLER (28 DAYS - 23 MONTHS) CHILDREN (2-11 YEARS) ADOLESCENT (12-17 YEARS) ADULT (18-65) ELDERLY (> 65)
AGEMAX	PLANNED MAXIMUM AGE OF SUBJECTS	No controlled terminology.
AGEMIN	PLANNED MINIMUM AGE OF SUBJECTS	No controlled terminology.
BLIND	TRIAL BLINDING SCHEMA	OPEN LABEL SINGLE BLIND DOUBLE BLIND
COMPTRT	COMPARATIVE TREATMENT NAME	No controlled terminology.
CONTROL	TYPE OF CONTROL	PLACEBO ACTIVE NONE
DIAGGRP	DIAGNOSIS GROUP	HEALTHY SUBJECTS
DOSE	TEST PRODUCT DOSE PER ADMINISTRATION	Use controlled terminology for interventions (DOSE).
DOSFRQ	TEST PRODUCT DOSING FREQUENCY	Use controlled terminology for interventions (DOSFRQ).
DOSU	TEST PRODUCT DOSE UNITS	Use controlled terminology for interventions (DOSU).
DRUGDICT	DRUG DICTIONARY	WHODRUG
INDIC	TRIAL INDICATIONS	No controlled terminology.
INDICTYP	TRIAL INDICATION TYPE	DIAGNOSIS CURE MITIGATION [Note: used narrowly to mean mitigate adverse effect of another treatment] TREATMENT PREVENTION

TSPARMCD	TSPARM	TSVAL
LENGTH	LENGTH OF TRIAL	No controlled terminology.
MHDICT	MEDICAL HISTORY DICTIONARY	MedDRA SPONSOR DEFINED
OBJPRIM	TRIAL PRIMARY OBJECTIVE	No controlled terminology however, the objective should be described in terms of the desired statement in labeling.
OBJSEC	TRIAL SECONDARY OBJECTIVE	No controlled terminology however the objective should be described in terms of the desired statement in labeling.
PHASE	TRIAL PHASE	1 1-2 2 2A 2B 2-3 3 3A 3B 3-4 4 5 NA
PLANSUB	PLANNED NUMBER OF SUBJECTS	No controlled terminology.
RANDOM	TRIAL IS RANDOMIZED	Y N
ROUTE	TEST PRODUCT ROUTE OF ADMINISTRATION	Use controlled terminology for interventions (ROUTE).
SEXPOP	SEX OF PARTICIPANTS	M F BOTH
SPONSOR	SPONSORING ORGANIZATION	No controlled terminology.
TITLE	TRIAL TITLE	No controlled terminology.
TRT	REPORTED NAME OF TEST PRODUCT	No controlled terminology.
TYPE	TYPE OF TRIAL	SAFETY EFFICACY BIO-EQUIVALENCE BIO-AVAILABILITY CONFIRMATORY EXPLORATORY PHARMACOECONOMIC PHARMACOGENOMICS PHARMACOKINETICS PHARMACODYNAMICS

10.4 CDISC VARIABLE-NAMING FRAGMENTS

The CDISC SDS group has defined a standard list of fragments to use as a guide when naming variables in SUPPQUAL (QNAM) or assigning TESTCD values that could conceivably be treated as variables in a horizontal listing derived from a V3.x dataset. In some cases, more than one fragment is used for a given keyword. This is necessary when a shorter fragment must be used for a TESTCD or QNAM that incorporates several keywords that must be combined while still meeting the 8-character variable naming limit of SAS transport files. When using fragments, the general rule is to use the fragment(s) that best conveys the meaning of the variable within the 8-character limit; thus, the longer fragment should be used when space allows. If the combination of fragments still exceeds 8 characters, a character should be dropped where most appropriate (while avoiding naming conflicts if possible) to fit within the 8-character limit.

In other cases the same fragment may be used for more than one meaning, but these would not normally overlap for the same variable.

Keyword(s)	Fragment
ACTION	ACN
ADJUSTMENT	ADJ
ANALYSIS DATASET	AD
BASELINE	BL
BIRTH	BRTH
BODY	BOD
CANCER	CAN
CATEGORY	CAT
CHARACTER	C
CONDITION	CND
CLASS	CLAS
CODE	CD
COMMENT	COM
CONCOMITANT	CON
CONGENITAL	CONG
DATE TIME - CHARACTER	DTC
DAY	DY
DEATH	DTH
DECODE	DECOD
DERIVED	DRV
DESCRIPTION	DESC
DISABILITY	DISAB
DOSE, DOSAGE	DOS, DOSE
DURATION	DUR
ELAPSED	EL
ELEMENT	ET
EMERGENT	EM
END	END, EN

Keyword(s)	Fragment
ETHNICITY	ETHNIC
EXTERNAL	X
EVALUATOR	EVAL
EVALUATION	EVL
FASTING	FAST
FILENAME	FN
FLAG	FL
FORMULATION, FORM	FRM
FREQUENCY	FREQ
GRADE	GR
GROUP	GRP
HIGHER LIMIT	HI
HOSPITALIZATION	HOSP
IDENTIFIER	ID
INDICATION	INDC
INDICATOR	IND
INTERVAL	INT
INTERPRETATION	INTP
INVESTIGATOR	INV
LIFE-THREATENING	LIFE
LOCATION	LOC
LOINC CODE	LOINC
LOWER LIMIT	LO
MEDICALLY-IMPORTANT EVENT	MIE
NAME	NAM
NON-STUDY THERAPY	NST
NORMAL RANGE	NR
NOT DONE	ND

Keyword(s)	Fragment
NUMBER	NUM
NUMERIC	N
ONGOING	ONGO
ORDER	ORD
ORIGIN	ORIG
ORIGINAL	OR
OTHER	OTH, O
OUTCOME	OUT
OVERDOSE	OD
PARAMETER	PARM
PATTERN	PATT
POPULATION	POP
POSITION	POS
QUALIFIER	QUAL
REASON	REAS
REFERENCE	REF, RF
REGIMEN	RGM
RELATED	REL, R
RELATIONSHIP	REL
RESULT	RES
RULE	RL
SEQUENCE	SEQ
SERIOUS	S, SER
SEVERITY	SEV

Keyword(s)	Fragment
SPECIMEN	SPEC, SPC
SPONSOR	SP
STANDARD	ST, STD
START	ST
STATUS	STAT
SUBCATEGORY	SCAT
SUBJECT	SUBJ
SUPPLEMENTAL	SUPP
SYSTEM	SYS
TEXT	TXT
TIME	TM
TIMEPOINT	TPT
TOTAL	TOT
TOXICITY	TOX
TRANSITION	TRANS
TREATMENT	TRT
UNIT	U
UNIQUE	U
UNPLANNED	UP
VARIABLE	VAR
VALUE	VAL
VEHICLE	V

10.5 LESSONS LEARNED FROM THE PILOT

After reviewing the data structures submitted for the pilot, the CDISC SDS Team overwhelmingly agreed that there were more commonalities than differences in the data structures submitted by the different companies. However, the team also concluded that the Version 3 concepts require additional documentation and training to minimize possible different interpretations as well as to highlight the model subtleties. Thus the principal conclusion from the pilot was the need for a much more detailed and complete implementation guide. The CDISC SDS Team presented these lessons at an FDA Public Meeting on October 2, 2003.

The three main objectives of the pilot were to ‘stress’ the conceptual models by including as many different types of therapeutic areas and data as possible, to test new model components (e.g., Trial Design), and to summarize sponsor and FDA experiences.

A broad variety of data was submitted from eight different therapeutic classes and four different trial designs. The therapeutic classes included Oncology, Metabolic, Cardiovascular, Anti-infective, Hormone Replacement Therapy, Central Nervous System, Respiratory, and Endocrine. The study designs included 1) Double-blind, parallel group, 2) Double-blind followed by open label, 3) Multiple treatment phases/arms, and 4) Infusion and oral dosing. All of the predefined Version 3.0 CDISC SDS domain models were tested as well as more than twelve new domains (e.g., Microbiology, Questionnaire, Study Drug Compliance, Mammogram, Tumor Identification and Measurement, Disease Response, Health Care Utilization). Although most of the participants used their legacy database, one participant tested the conversion of Version 2.0 datasets to Version 3.0.

The testing of new model components included the inclusion of new variables, concepts and datasets. An analysis of the variables used in the pilot was shared with the CDISC SDS Team. This analysis highlighted the variables actually used in the submission datasets, the new Version 3.0 variables, plus those that were not described within the specifications. It was through this analysis that the team deliberated over the cost/benefits of the inclusion of new variables or changing existing ones. Furthermore, the team discovered that new variables such as --GRPID, --SEQ, --SPID, --CAT, and --SCAT were beneficial but sponsors would need additional guidance to ensure consistent application.

One especially significant conclusion was the reexamination of the date-time precision variable combination. The team concluded that the ISO 8601 date/time specification was more efficient and complete and leveraged a preexisting industry standard compared to the Version 3.0 variable pairing. Therefore, after much debate the team endorsed and included the ISO 8601 date format into version 3.1. Additionally, new data domains and concepts were tested, for example the trial design datasets. The findings and feedback from sponsors as well as from FDA increased the importance of including these concepts in the Version 3.1 release.

The last major objective was to capture both sponsor participant and FDA experiences with the pilot. On the sponsor side, the team looked into the ease and consistency of use and categorization of data into the observation classes (Findings, Events and Interventions) as well as the identification of challenging areas. On the FDA side, the team was interested in reviewer experiences with the data and applications used in the pilot. A number of challenges were highlighted, such as re-conceptualizing how the electronic submission data is organized. For example, data that was traditionally collected within one CRF module may now go into multiple V3 domains and data that was collected on multiple CRF modules may go into one V3 domain. The second challenge faced by the participants was the physical mapping between horizontal data structures to a vertical or more ‘normalized’ paradigm. Like the industry at large, pilot participants’ needed to be educated on new terms (e.g., epochs, elements and arms) that are used to replace common terms (e.g., periods, phases, and treatment groups) and concepts. Lastly, it was not always clear cut where data should be placed if the data did not easily fit into a specific observation class; this was in part due to the fact that the Version 3 specifications were still in a state of flux during the pilot period.

In conclusion, the number-one learning from this pilot was that additional guidance and specifications were needed to reduce inconsistencies and enhance comprehension of the models. Specifically, a detailed implementation guide was necessary to more clearly communicate the specifications and the rules, as well as to provide additional guidance through examples is necessary. Also, the team learned that the vertical nature of the Findings datasets highlights the importance of and the need for specific controlled terminology to be able to provide record level metadata (e.g., via define.xml). Implementing Version 3 for the safety domains based on Version 2 was

straightforward. Lastly, complex data (e.g., Microbiology, where multiple relationships (one to many to many) exist within and between variables and records) and study designs (e.g., Oncology) could be represented in Version 3. The SDTM is not meant to mirror data as collected on CRFs; that is, while CRFs will continue to be optimized for data collection, the SDTM models are optimized for electronic data exchange with the FDA. The V3 structures seemed to handle all of the many variations in data provided from different sponsors, and seemed to provide a very effective foundation for a common data standard.

10.6 REVISION HISTORY

As stated above, this document, when approved, together with the SDTM, represents the most recent version of the CDISC Submission Data Domain Models, previously known as Version 3.1. The V3.x Study Data Tabulation Model Implementation Guide for Human Clinical Trials was the first implementation-ready version of the CDISC Submission Data Domain Models.

10.6.1 Changes from CDISC Version 2 Models to Version 3

V3 represented a major change from the CDISC Version 2.0 (V2) domain models for two primary reasons. First, it included a general model for representing all types of tabulation data (the general model has now been published separately as the SDTM). Second, the general vertical record structures used in V3 were more normalized than the more horizontal V2 record structures. Standard horizontal listings of data, such as those described in the V2 horizontal representations of ECGs and Vitals by visit, will be produced by FDA standard review tools.

V3 was initially released for comment in March 2003, and finalized as an approved HL7 informative document that addressed all prior comments on June 9, 2003. V3 was then tested in Summer 2003 by individuals from nine sponsor companies who participated in an FDA pilot project. The results of the pilot (which are discussed in [Section 10.5](#)) were shared with industry at an FDA public meeting held on October 2, 2003, and feedback from the pilot was a primary input to V3.x. Another key input was a list of comments that had to be deferred for the June 9, 2003 publication, but which were addressed in V3.1 and later versions.

10.6.2 Changes from CDISC V3 to V3.1

- 1) The general underlying conceptual model described in V3 as the General Study Data Information Model, was published separately as the SDTM, a separate document from this implementation guide.
- 2) Corrections and amendments applied in the SDTM were also applied throughout the domain models, assumptions, and examples provided in this document.
- 3) A new Trial Design component was incorporated with examples to provide a standardized representation of study timing and treatment plans, and a way of representing actual subject visits.
- 4) A more thorough solution was included for defining relationships between datasets, domains, and/or records.
- 5) The representation of all date/time variables was changed to ISO 8601 character format. The concept of a separate Date/Time Precision variable for each date/time variable was eliminated in this version, because that purpose is met by the ISO 8601 standard.
- 6) New domain variables were added to represent additional timing descriptions, flags, and descriptive attributes of an observation (e.g., --SCAT, --DOSRGM, --NRIND).
- 7) Some variables were removed from the domain models (e.g., --INTP, --DESC, --BLRESC, --BLRESN), either because they were deprecated in the prior version or were inconsistent with the intent of the model.
- 8) The core variable concept was expanded to categorize variables as required, expected, or permissible.
- 9) The Qualifier variable role was sub-categorized into five more granular subcategories (See [Section 2.1](#)) to provide more detail on the use of variables.
- 10) The SU, DS, and PE domains were significantly redesigned.
- 11) Additional business rules were incorporated to address common data management situations, such as representing tests not done.
- 12) Numerous changes to V3 variables, labels, formats, and notes were made to reduce ambiguity and improve consistency.
- 13) New sections were included in the Implementation Guide to provide additional descriptive examples and additional guidance on proper usage.

10.6.3 Changes from CDISC SDSIG V3.1 to SDTMIG V3.1.1 Draft

Major Changes		
Change	General	Changed abbreviation from SDSIG to SDTMIG. Updated references to SDTM v1.1. Introduced abbreviation of V3.x to refer to V3.1 IG and subsequent versions.
Change	2.5	Any valid variable included in an SDTM general class may now be used in a domain based on that class.
Add	4.1.4.8	New assumption on the use of date/time variables for findings collected over an interval.
Add	Separate Document	New DV Protocol Deviations domain model and example
Add	Separate Document	New DA Drug Accountability domain and example
Add	Separate Document	New Pharmacokinetics Concentrations domain model and example
Add	Separate Document	New Pharmacokinetics Parameters domain model and example
Add	7.10	New TS Trial Summary domain model and examples. Also added TS code to Appendix 10.3.1.
Change	8	Substantial revisions throughout this Section to clarify proper usage.
Add	8.4.3	Inserted new option for creating separate SUPPQUAL datasets (supp--) per domain.

Minor Changes		
Correct	1.3	Moved list of changes to appendix 10.6. Revised text to accommodate v. 3.11.
Correct	2.1	Corrected spelling of --ORNRLO and --ORNRHI.
Correct	2.5	Update list of domains with DV, DA, MB, PC, PP, TS
Correct	2.6	Clarified step 11 to emphasize that labels should only be changed in new domains. Added statement about posting suggested new codes to CDISC web site.
Correct	2.6.1	Corrected step number references in last 2 paragraphs.
Correct	3.2	Updated documentation to clarify that the key order in metadata should reflect the natural sort order of the domain when viewed, since all physical keys are fixed within the SDTM (based on STUDYID, USUBJID, --SEQ). Clarified that roles are optional for standard domains.
Correct	3.2.1	Changed dataset names to lowercase for consistency with FDA eCTD guidance requirements. Corrected Structure column for consistency with domain models.
Add	4.1.1	Added reference to CDISC Define.xml specification.
Correct	4.1.2.2	Remove reference to RELDS and COMMENTS.
Correct	4.1.2.4	Clarified when the use of mixed case text is allowable.
Correct	4.1.3.4	Clarified that analysis variables may be included as supplemental qualifiers.
Correct	4.1.3.5	Removed references to PE
Correct	4.1.4.1	Corrected format of fourth bullet example by adding hyphen separators to year. Fixed link

Minor Changes		
Correct	4.1.4.2	Reworded first sentence. Updated footnote to use correct expression syntax (input(DTC, datetime18))
	4.1.4.3	Corrected examples and expanded documentation on use of ISO8601 for durations
Correct	4.1.4.5	Updated to allow use in other domains.
Correct	4.1.4.7	Updated diagram and description to explain how to address timepoints that either start before the study with an unknown end or start during a study with an unknown end.
Correct	4.1.5.1	Clarified use of --ORRES in PE; minor text corrections to clarify. Used correct CT for SYSBP.
Correct	4.1.5.2	Clarified that long text should be split for readability when > 200 characters.
Correct	4.1.5.4	Clarified use of EVAL and that the name of the attribution should increment by 1 for each additional attribution on the same topic
Correct	5.1.1	Specified use of ISO 3166 3-character country code. Added example of USUBJID convention.
Correct/Add	5.1.1.1	Clarified assumption 1 on necessity of including SITEID. Clarified assumption 7 on how to represent additional race information. Added assumption 8 on screen failures.
Correct	5.1.2	Corrected reference to COVAL in paragraph 1. Corrected CDISC notes for DOMAIN, RDOMAIN, COREF and COVAL. Added reference to 8.5.
Correct	5.1.2.1	Corrected references to IDVAR in assumption 1; correct example.
Correct	6.1.1	Correct label, role and notes for CMSPID for consistency with other domain models. Corrected label for CMMODIFY for consistency with SDTM. Corrected role for CMCLASCD, CMDOSU, CMDOSFRM, CMDOSFRQ, CMROUTE to Variable Qualifier. Edited notes and added reference and new term to CMSTREF and CMENREF. Corrected CDISC Notes for CMOCCUR.
Add/Remove	6.1.1.1	Added new assumption 5 to discuss record structure. Removed old assumption 5 (which was redundant with Notes and assumption 4.1.4.7).
Correct	6.1.2	Corrected role of EXSPID for consistency with SDTM and other domain models. Correct role for EXDOSU, EXDOSFRM, EXDOSFRQ, EXROUTE to Variable Qualifier. Updated CDISC notes for EXSTDTC and EXENDTC.
Correct	6.1.3	Corrected role of SUSPID for consistency with SDTM and other domain models. Corrected role for SUCLASCD, SUDOSU, SUDOSFRM, SUDOSFRQ, SURROUTE to Variable Qualifier. Correct note to SUSTAT. Corrected Origin for SURROUTE for consistency.
Correct	6.2.1	Corrected labels for AESTDY, AEENDY, AEDUR for consistency with SDTM. Corrected notes for AETOXGR and AEDUR. Corrected note and CT for AEENRF.
Correct	6.2.1.1	Corrected text in 5a to allow use of both AEENDTC and AEENRF..
Correct	6.2.2	Corrected notes for DSREFID, DSSPID, DSSTDTC and DSSTDY. Corrected role for EPOCH to Timing. Corrected Label for DSTERM and DSDECOD
Correct	6.2.3	Corrected labels for MHSEQ, MHTERM, MHCAT, MHSCAT, MHDTC for consistency. Corrected notes to MHTERM, MHDECOD, MHSCAT, MHDTC, MHENDTC. Changed Core to Permissible for MHSTDTC. Corrected note to MHENRF to allow use with MHENDTC.
Correct	6.2.3.1	Updated Note 4d for clarity.

Minor Changes		
Correct	6.3.1	Added variables EGNRIND, EGMETHOD, EGTPTNUM to model. Corrected notes for EGGRPID, EGNRIND, EGSCAT, EGLOINC, EGEVAL. Corrected references for EGORRESU, EGSTRESN, EGSTRESU, EGDRVFL, EGELTM. Changed Core for EGPOS to Permissible. Added example in note to EGELTM.
Correct	6.3.1.1	Corrected Assumption 1.
Correct	6.3.2	Corrected Type of IESPID to Char for consistency. Corrected Core for IERRES and IESTRESC to Required. Corrected Core for IEDTC to Permissible. Updated reference for IEDY. Corrected note to IETEST.
Correct	6.3.2.1	Added new assumption 3 to clarify relationship to TI domain.
Correct	6.3.3	Added variables LBSTRESU, LBNAM, and VISITNUM (all omitted in error). Corrected references for LBORRESU, LBREASND, LBDRVFL, LBELTM. Changed Role of LBNRIND to Expected. Corrected Note to LBNRIND. Changed labels and notes to use "Reference Range" instead of "Normal Range". Added example to note for LBELTM.
Add	6.3.3.1	Added new assumption on use of reference range fields.
Correct	6.3.4	Added PEORRESU, PESTRESN, PESTRESU, PEBLFL, PEEVAL, PESEV to domain model. Corrected note for PETEST. Corrected type of PESPID to Char. Changed role of VISITNUM and PEDTC to Expected. Corrected reference for PEREASND. Moved PESPID to top of domain for consistency.
Correct	6.3.5	Corrected note for QSCAT, Changed Core for VISITNUM to Expected, Corrected reference for QSDRVFL, QSELTM. Added new assumptions.
Add	6.3.5.1	Added new assumptions 1-3 to better explain use of QS.
Correct	6.3.6	Added variables SCSTAT, SCREASND, SCDY for consistency. Corrected Role of SCSTRESC, SCSTRESN, SCSTAT, SCREASND to Result Qualifier.
Add	6.3.6.1	Added new assumption on use of SC vs. DM and SUPQUAL.
Correct	6.3.7	Added VSNRIND variable. Updated Core of VSSPID, VSPOS and VSNRIND to Permissible, VSTEST to Required, VSSTRESC and VSSTRESN to Result Qualifier, VSNRIND, VSSTAT and VSLOC to Permissible, VSBLFL and VSDTC to Expected. Corrected references for VSCAT, VSORRES, VSORRESU, VSSTRESC, VSSTAT, VSLOINC, VSBLFL, VSDRVFL, VSELTM. Corrected note for VSNRIND and VSDRVFL..
Add	7.1	Added references to Epochs.
Correct	7.2.1	Corrected structure in header. Corrected notes for ETCD, ELEMENT and TEENRL. Changed role of ELEMENT to Synonym Qualifier. Changed Core of TEENRL to Perm.
Correct	7.2.2	Corrected structure in header. Corrected labels for ARM and ARMCD (removed "Treatment Group"). Changed EPOCH Role to Timing. Corrected notes for ARMCD and ELEMENT..
Correct	7.2.3	Corrected structure in header. Corrected TVENRL Origin to Sponsor Defined/ Protocol. Corrected label for ARMCD and VISITNUM. Corrected note for ARMCD and ARM. Changed Core of VISIT to Perm.
Correct	7.3.1	Corrected DOMAIN code to SE. Corrected structure in header. Corrected note to ELEMENT.
Correct	7.3.2	Corrected structure in header. Removed "Treatment" from note to SVENDTC
Correct	7.4.2	Corrected ELEMENT values in table to match figure.
Replace	7.4.4	Replaced second example with corrected version that adds EPOCH. Added clarification to last paragraph. Corrected table values for ELEMENT for consistency with figures.
Correct	7.5	Corrected number of records in figure to 20.

Minor Changes		
Correct	7.7.5	Clarified wording of second paragraph.
Correct	7.9	Corrected Roles of IETEST to Synonym Qualifier and IECAT to Grouping Qualifier. Corrected notes for IETESTCD, IETEST, IECAT.
Correct	8.0	Removed references to lesions in bullet 3 and made minor text edits.
Correct	8.1.1	Clarified text in paragraph 1.
Correct	8.2	Substantial text corrections to entire section for clarification.
Correct	8.2.2	Corrected USUBJID reference in Example A
Correct	8.3	Inserted clarifying test in paragraph 2.
Correct	8.3.1	Replaced example of MB for split domains. Substantially revised text for clarification.
Add	8.4	Added new paragraphs 3 and 4 to clarify use of QNAM and QVAL. Substantially revised text for clarification.
Correct	8.4.1	Substantial text revisions for clarification. Corrected notes for QNAM and QLABEL. Added new last paragraph.
Add	8.4.2	Substantially revised text for clarification. Added new paragraph at end to emphasize that null values for QVAL are not to be used. Adjusted values in example; changed PPE to PPROT. Added new foreign language example.
Correct	8.5	Substantially revised text for clarification.
Correct	8.5.1	Corrected COREF values in example.
Add	8.6	Substantially revised text for clarification. Added new paragraphs 3-5 and revised other sentences to clarify principle of domain topicality.
Add	8.6.1	Inserted new Stress Test multiple-dataset example.
Correct	9.1	Removed last row in example
Remove	9.3.3	Deleted statement beginning with "If terms are specified in the CRF" below note e.
Replace	9.4.1	Updated EG Example with minor corrections.
Replace	9.4.2	Updated IE Example with minor corrections.
Replace	9.4.3	Updated LB Example; added second example to show a urine collection, including a negative reference period.
Replace	9.4.4	Updated PE Example with PESEV and minor corrections
Add	9.4.5	Corrected QS example. Added explanations.
Replace	9.4.6	Updated SC Example with minor corrections.
Replace	9.4.7	Updated VS Example with minor corrections.
Replace	9.5.1	Replaced Trial Design examples with new examples to better communicate core concepts; added new TS example.
Add	9.5.2	Added new TS example.
Remove	10.3.1	Removed reserved codes for DR, DY, NE, PF, RF, SZ, VX as these may no longer be necessary as separate domains. Changed code for Biopsy from BX to BR.
Add	10.3.5	Added controlled terms for new TS domain.
Correct	10.3.3	Added new code for BODYFAT
Correct	10.3.4	Added new code for SCRNFAT.
Correct	10.4	Corrected several fragment names.
Correct	10.5	Revised last paragraph. Other minor text revisions.

10.6.4 Changes from CDISC SDTMIG V3.1.1 Draft to V3.1.1 Final

Major Changes		
Add	2.4	Included additional information about choosing a general class and ability to use any qualifier from same general class.
Change	4.1.1.5	Corrected last bullet to allow use of any qualifier from same general class. Also clarified that a null column is acceptable for expected variables not collected.
Add	4.1.4.1 4.1.4.2	Reworded for improved clarity. Added examples of new ISO 8601-compliant representation for partial date/times with omitted components.
Change	4.1.4.7	Removed restriction that --STRF and --ENRF can only be used if start and/or end dates are missing. Also removed restrictions from domain models using those variables (CM, SU, AE, MH).
Add	4.1.4.9	New assumption on representing date results in findings.
Correct	4.1.5.1	Substantially reworded assumption for clarity. Corrected EGTESTCD for RHYRATE in ECG example.
Change	Sections 5 & 6	Changed the header for all domain models to use lowercase xpt filenames for consistency with eCTD specifications.
Change	5.1	Revise notes and Core for RFSTDTC, RFENDTC and revise Assumption 8 to allow for representation of screen failures with ARMCD='SCRNFAIL' rather than using SUPPQUAL.
Change	Section 6	Changed label for all domain instances of --SPID from "Sponsor ID" to "Sponsor-Defined Identifier" so as not to be confused with the ID of a sponsor company. Reclassified --STAT as a Record Qualifier.
Change	Section 6	Changed domain models for CM, SU, AE, MH to address possible use of new SuppQual variable --PRESP, reconsider use of --OCCUR, and emphasize importance of including data that represents actual interventions or events.
Change	6.3.5.1	Revised QS Assumption 1 to clarify domain contents.
Add	7.3.1	Added new variable SESEQ to uniquely identify each row and set the sequence of elements followed by a subject.
Change	8.4	The use of a single SUPPQUAL dataset is not recommended for users of V3.1.1 and later versions. The use of separate datasets has been found to be much more convenient for sponsors and is necessary for compatibility with the CDISC ODM model and define.xml.
Change	8.6	Changes to text on determining where to put data in the SDTM and removal of example, which is under discussion among the SDS Team.
Change	9.2.3	Removed example 1 from prior SU example, since proper use of the SUOCCUR variable is being reconsidered by the SDS Team; Revised example 2 (now1).
Change	9.3.3	Modified example to remove references to MHOCUR variable, since use of this variable is being reconsidered by the SDS Team.
Add	10.3.1	Added description for each domain code; adjusted list of reserved codes.
Add	10.7	New section 10.7 on Warranties, Limitation of Liability, Disclaimers.

Minor Changes		
Add	1.1	Expanded definition of tabulation datasets. Included reference for FDA study data specifications.
Add	1.3	Modified bullet 6 to reference new appendix 10.6.
Correct	2.1	Removed --SPEC, --LOT and --NAM as examples of Grouping Qualifiers, moved to Record Qualifiers. Referred to use of Result Qualifiers in DM.
Add	2.2	Additional explanatory text in paragraph 1.
Correct	6.1.1	Corrected label for CMREASND to conform with 40 character limit.
Correct	6.1.3	Added SUSTRF and SUENRF; modified label of SUTRT. Reclassified former Result Qualifiers as Record Qualifiers.
Correct	6.2.1	Corrected examples in CDISC Note for AEREL. Added notes and ‘***’ to AETOXGR and reclassified former Result Qualifiers as Record Qualifiers.
Correct	6.2.1.1	Reworded assumption 1 to clarify definition. Revised Assumption 4 to reduce ambiguity and discuss possible future use of AEPRESP and AEOCCUR.
Correct	6.2.2	Corrected example list of values for DSDECOD.
Correct	6.2.2.1	Reworded assumption 1 to clarify definition.
Correct	6.2.3	Changed Core to Perm for MHBODSYS and MHDECOD. Corrected Origin for MHBODSYS to Derived. Deleted 2 nd note for MHBODSYS reclassified former Result Qualifiers as Record Qualifiers.
Correct	6.2.3.1	Reworded assumption 1 to clarify definition. Reworded assumption 3a to clarify use of MHCAT and MHSCAT.
Add	6.3.1	Added note to EGLOINC; corrected label for ESPID, corrected note to EGORES.
Correct	6.3.2	Corrected labels to IETESTCD, IETEST and IECAT for consistency with SDTM TI labels. Corrected labels and notes for IEORES and IESTRESC.
Correct/Add	6.3.3	Corrected examples in CDISC NOTE for LBMETHOD. Added LBENDTC to model.
Correct	6.3.3.1	Corrected wording of assumption 1 for LB.
Correct	6.3.4	Corrected header to PE table.
Correct	6.3.5.1	Removed "QSORRES" from assumption 4.
Correct	6.3.6	Moved SCSPID after SCGRPID for consistency with other domains.
Correct	6.3.6.1	Removed last 2 sentences from Assumption 2, which described an example that belonged in supdm rather than suppsc.
Correct	7.2.1	Added “ISO 8601” to Format for TEDUR; Corrected note for ETCD (ETCD note also corrected in other trial design domains)
Correct	7.2.2	Corrected note for ARMCD (ARMCD note also corrected in other trial design domains)
Correct	7.3.2	Corrected note to VISITDY (to remove reference to use of decimals).
Correct	7.4.4	Changed examples for TEDUR to ISO8601 format.
Correct	7.5.2	Changed examples for TEDUR to ISO8601 format.
Correct	7.9	Corrected labels to IETESTCD, IETEST and IECAT for consistency with SDTM TI labels.
Correct	7.10	Corrected Notes to TSPARM and TSVAL; corrected core for TSVAL
Correct	8.3.1	Updated example to use MB and MS microbiology domains.
Correct	8.4.2	Clarified wording of bullet 2.
Correct	8.6.1	Removed incorrect example; a better example will be provided in the next version of the SDTMIG.

Minor Changes		
Add	9.3.3	Revised example, eliminating MHOCCUR example and adding new note and new column to illustrate use of MHENRF.
Correct	9.4.1	Corrected values of EGEVAL in Example 1.
Correct	9.4.4	Corrected values of USUBJID in PE Example.
Correct	9.4.6	Removed SCDY from example (not likely to be used in SC). Changed SCTESTCD values from "INITIALS" to "SUBJINIT" for consistency with domain examples.
Correct	9.4.7	Corrected explanation of example and row references.
Correct	9.5.1	Changed examples for TADUR and TEDUR to ISO8601 format
Add	10.3.1	Added SI Infection Site Measurements to domain codes; removed SS.
Delete	10.3.2	Removed "FINDING" from list of controlled terms for EG.
Correct	10.3.3	Minor changes to VSTEST for TEMP and RESP
Correct	10.3.4	Changed AGEGRP to AGESPAN, AGEGRP, CONTROL, DRUGDICT, OBJPRIM, OBJSEC; correction to code "3-4" in PHASE; added new values to TYPE. Changed MASK to BLIND; changed MEDDICT to MHDICT. Removed TSPARM = DESIGN.
Correct	10.3.5.1	Corrected list of Trial Summary parameters; a more complete list of parameters will be defined separately by the CDISC Terminology team.
Correct	10.4	Corrected list of fragments (Domain codes are no longer included in this fragment list, but appear only in 10.3.1).

10.7 REPRESENTATIONS AND WARRANTIES; LIMITATIONS OF LIABILITY, AND DISCLAIMERS

CDISC Patent Disclaimers. It is possible that implementation of and compliance with this standard may require use of subject matter covered by patent rights. By publication of this standard, no position is taken with respect to the existence or validity of any claim or of any patent rights in connection therewith. CDISC, including the CDISC Board of Directors, shall not be responsible for identifying patent claims for which a license may be required in order to implement this standard or for conducting inquiries into the legal validity or scope of those patents or patent claims that are brought to its attention.

Representations and Warranties. Each Participant in the development of this standard shall be deemed to represent, warrant, and covenant, at the time of a Contribution by such Participant (or by its Representative), that to the best of its knowledge and ability: (a) it holds or has the right to grant all relevant licenses to any of its Contributions in all jurisdictions or territories in which it holds relevant intellectual property rights; (b) there are no limits to the Participant's ability to make the grants, acknowledgments, and agreements herein; and (c) the Contribution does not subject any Contribution, Draft Standard, Final Standard, or implementations thereof, in whole or in part, to licensing obligations with additional restrictions or requirements inconsistent with those set forth in this Policy, or that would require any such Contribution, Final Standard, or implementation, in whole or in part, to be either: (i) disclosed or distributed in source code form; (ii) licensed for the purpose of making derivative works (other than as set forth in Section 4.2 of the CDISC Intellectual Property Policy ("the Policy")); or (iii) distributed at no charge, except as set forth in Sections 3, 5.1, and 4.2 of the Policy. If a Participant has knowledge that a Contribution made by any Participant or any other party may subject any Contribution, Draft Standard, Final Standard, or implementation, in whole or in part, to one or more of the licensing obligations listed in Section 9.3, such Participant shall give prompt notice of the same to the CDISC President who shall promptly notify all Participants.

No Other Warranties/Disclaimers. ALL PARTICIPANTS ACKNOWLEDGE THAT, EXCEPT AS PROVIDED UNDER SECTION 9.3 OF THE CDISC INTELLECTUAL PROPERTY POLICY, ALL DRAFT STANDARDS AND FINAL STANDARDS, AND ALL CONTRIBUTIONS TO FINAL STANDARDS AND DRAFT STANDARDS, ARE PROVIDED "AS IS" WITH NO WARRANTIES WHATSOEVER, WHETHER EXPRESS, IMPLIED, STATUTORY, OR OTHERWISE, AND THE PARTICIPANTS, REPRESENTATIVES, THE CDISC PRESIDENT, THE CDISC BOARD OF DIRECTORS, AND CDISC EXPRESSLY DISCLAIM ANY WARRANTY OF MERCHANTABILITY, NONINFRINGEMENT, FITNESS FOR ANY PARTICULAR OR INTENDED PURPOSE, OR ANY OTHER WARRANTY OTHERWISE ARISING OUT OF ANY PROPOSAL, FINAL STANDARDS OR DRAFT STANDARDS, OR CONTRIBUTION.

Limitation of Liability. IN NO EVENT WILL CDISC OR ANY OF ITS CONSTITUENT PARTS (INCLUDING, BUT NOT LIMITED TO, THE CDISC BOARD OF DIRECTORS, THE CDISC PRESIDENT, CDISC STAFF, AND CDISC MEMBERS) BE LIABLE TO ANY OTHER PERSON OR ENTITY FOR ANY LOSS OF PROFITS, LOSS OF USE, DIRECT, INDIRECT, INCIDENTAL, CONSEQUENTIAL, OR SPECIAL DAMAGES, WHETHER UNDER CONTRACT, TORT, WARRANTY, OR OTHERWISE, ARISING IN ANY WAY OUT OF THIS POLICY OR ANY RELATED AGREEMENT, WHETHER OR NOT SUCH PARTY HAD ADVANCE NOTICE OF THE POSSIBILITY OF SUCH DAMAGES.

Note: The CDISC Intellectual Property Policy can be found at http://www.cdisc.org/about/bylaws_pdfs/CDISCIPPolicy-FINAL.pdf.