Study Data Tabulation Model Implementation Guide:
Human Clinical Trials

Prepared by the
CDISC Submission Data Standards Team

Notes to Readers

- This is the approved implementation guide for Version 1 of the CDISC Study Data Tabulation Model

- This Implementation Guide comprises version 3.1 of the Submission Data Standards.

Revision History

Date	Version	Summary of Changes
2004-06-25	Document Version 1.0	Released version reflecting all changes identified during comment periods.
2004-07-14	Document Version 1.01	Corrects minor typos and errors in sections 4.1.4.2, 5.1.1.0.7, 5.1.2.0, 6.3.1, 6.3.3, 6.3.4, 6.3.6.

CONTENTS

1 Introduction

1.1 PURPOSE

This document describes the CDISC Version 3.1 (V3.1) 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 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 supersedes the Version 3.0 Submission Data Domain Models document, which was approved by Health Level Seven (HL7) as an informative document in June 2003.

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. The SDTM, which should be read before this Implementation Guide, describes a general conceptual model for representing clinical study data that is submitted to regulatory authorities. SDS V3.1 provides specific domain models, assumptions, business rules, and examples for preparing standard datasets that are based on the SDTM.

Tabulation datasets, which describe 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 is not currently intended to meet the needs supported by analysis datasets, which will continue to be submitted separately in addition to the tabulations.

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 SDS 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 (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 model 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 a new 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, DATASET 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, together with the SDTM, represents the first formal revision to the CDISC Submission Data Domain Models, Version 3.0 (V3). V3 represented a major change from the CDISC Version 2.0 (V2) domain models because it included a general model for representing all types of tabulation data (the general model has now been published separately as the SDTM). 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 were shared with industry at an FDA public meeting held on October 2, 2003, and feedback from the pilot was a primary input to V3.1. Another key input was a list of comments that had to be deferred for the June 9, 2003 publication, but which have now been addressed in this new version.

The most significant changes since the prior Version 3.0 are summarized below:

- The general underlying conceptual model described in V3 as the General Study Data Information Model, has now been published separately as the SDTM, a separate document from this implementation guide.
- Corrections and amendments applied in the SDTM have also been applied throughout the domain models, assumptions, and examples provided in this document.
- A new Trial Design component has been incorporated with examples to provide a standardized representation of study timing and treatment plans, and a way of tracking actual subject visits.
- A more thorough solution has been included for defining relationships between datasets, domains, and/or records.

- The representation of all date/time variables is now in ISO 8601 character format. The concept of a separate Date/Time Precision variable for each date/time variable has been eliminated in this version, because that purpose is met by the ISO 8601 standard.
- New domain variables have been added to represent additional timing descriptions, flags, and descriptive attributes of an observation (e.g., --SCAT, --DOSRGM, --NRIND).
- Some variables have been 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.
- The core variable concept has been expanded to categorize variables as required, expected, or permissible.
- The qualifier variable role has been sub-categorized into five more granular subcategories (See Section 2.1) to provide more insight into the use of variables.
- The SU, DS, and PE domains have been significantly redesigned.
- Additional business rules have been incorporated to address common data management situations, such as representing tests not done.
- Numerous changes to V3 variables, labels, formats, and notes have been made to reduce ambiguity and improve consistency.
- New sections have been included in this document to provide additional descriptive examples and additional guidance on proper usage.

Note that V3.1 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.

The first review version (Version 0.8) of the Implementation Guide was posted for comment (along with the SDTM) on the CDISC website and the HL7 ballot process in March and April of 2004. This final version incorporates changes that were identified as a result of the comment period. Rationale for these changes is described in the SDS V3.1 Comments and Responses document, which is available on the CDISC web site. The most significant changes from Review Version 0.8 are summarized below:

- Replaced the qualifier variable--SPREF with the identifier variable --SPID in all domains based on the 3 general classes
- Abridgement of certain variable labels (to conform with the SAS Transport 40 character limit)
- Removed the variable RACEOTH from the Demographics special purpose domain; this variable, when used, should be represented as an SCTESTCD in the Subject Characteristics domain
- Removed the variable AETRTEM from the AE domain; this variable, which requires an Evaluator, should be represented in Supplemental Qualifiers
- Removed the variable AESOTHC from the AE domain; this variable, when used, should be represented in Supplemental Qualifiers
- Renamed the variable AERELOTH from the AE domain to AERELNST to retain consistency of meaning of the SDS "OTH" naming fragment
- Renamed the variable AESOTH from the AE domain to AESMIE to retain consistency of meaning of the SDS "OTH" naming fragment
- Redefined and renamed the variable --TOXCAT from all domains based on the Findings class to --TOX to more accurately describe the intended concept
- Redefined the variable --TOXGR from all domains based on the Findings class to more accurately describe the intended concept
- Added the variable LBMETHOD to the Lab domain
- Redefined the timing variables --EPOCH, --STRF and --ENRF (table 2.2.5) to more accurately describe the intended concept
- Added variable TEDUR to the Trial Design model
- Renamed the variable TICRIT from the Trial Design Model to TIRL to more accurately describe the intended concept
- Redefined the general structure for expressing relationships among records in Section 8.0. This resulted in several variable name changes in tables 8.2.1 and 8.4.1 and the removal of table 8.3.1 (which is now addressed in table 8.2.1 under the new approach).

- Revised several assumptions and comments for the domain models
- Revised and amended several examples for consistency with these changes
- Included revised appendix 10.3 to include additional codes for EG and VS TESTCD values and SUPPQUAL
- Included new Appendix 10.5 to summarize lessons learned from the pilot.

Although this version is expected to be the first fully implementation-ready version of the CDISC Submission Data Standards directly referenced by the FDA for use on clinical studies involving human drug products, future improvements and enhancements may be made over time as sponsors develop more experience using 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 over time. CDISC also expects to produce further details on controlled terminology in the future, 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 Implementation Guide 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 STDM 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 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 challenging new concepts, readers may choose to defer this step until after they have mastered the essentials of creating standardized domains.
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 model can be submitted at any time through the CDISC Discussion Forum.

2 Fundamentals of the SDTM

2.1 OBSERVATIONS AND VARIABLES

The V3.1 submission data standards are based on the SDTM, 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*, has been 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, --SCAT, --GRPID, --SPEC, --LOT, and --NAM. The latter three grouping qualifiers can be used to tie a set of observations to a common source (i.e., specimen, drug lot, or laboratory name, respectively).
- *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.
- *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, and --LOINC which is an equivalent term for a --TEST and --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.
- *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, --ORNHI, and --ORNLO, all of which are variable qualifiers of --ORRES: and --DOSU, --DOSFRM, and --DOSFRQ, 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 start date, which captures the information, 'starting on Study Day 6', while an

---

example of a Variable 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 observations with a topic-specific commonality about a subject. 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 as 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. 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*

- 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 V3.1 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 Formats 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 **). CDISC intends to provide other publications providing additional guidance on use of controlled terminology in the future.

2.3 SPECIAL-PURPOSE DOMAINS

The CDISC V3.1 Submission Data Domain Models include two special-purpose domains with a specific structure – Demographics and Comments. 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. These domains, which are described in Section 5 of this document, cannot be extended with any additional qualifier or timing variables other than those specified. Additional non-extensible special-purpose domains are included in the Trial Design model described in Section 7 and in Section 8.

2.4 THE GENERAL DOMAIN CLASSES

Most other observations collected during the study 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 (usually for therapeutic purposes) 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.

The assumptions for the Domain models for representing information typically submitted from Clinical Trials are described in Section 4 of this document.

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. Two 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.

2.5 THE CDISC STANDARD DOMAIN MODELS

The following standard domains with their respective DOMAIN codes, have been defined by CDISC in this document:

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

Findings:
- ECG Tests - EG
- Inclusion/Exclusion Exceptions - IE
- Laboratory Tests - LB
- Questionnaires - QS
- Physical Examinations - PE
- Subject Characteristics - SC
- 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

Special-Purpose Relationship Datasets:
- Supplemental Qualifiers - SUPPQUAL
- 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. 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 to the V3.1 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 model, and the corresponding descriptions from the data definition document, but new variables 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 V3.1 model, sponsors should keep in mind that the V3/V3.1 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 V3.1. 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, or include incremental analysis variables in separate analysis files. 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 3.1 models, they should provide the rationale and description of the added variable(s) to the CDISC SDS Team so the impact on V3.1 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 in the CDISC Discussion Forum.

2.6   CREATING A NEW DOMAIN

This section describes the overall process of how to create a new CDISC SDTM domain. 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. First, 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 V3.1, such information can often 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. 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 SDS Version <Version Number e.g., 3.1>, <dataset structure (e.g., 'one record per subject' or 'one record per event'), >, Tabulation.

14. Submit a request to add specific new variables necessary to represent additional information in the general observation classes through the CDISC discussion facility.

As noted in Step 7 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 used for naming 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.

2.6.1 How to Include Variables in New Domains (Steps 4 through 6)
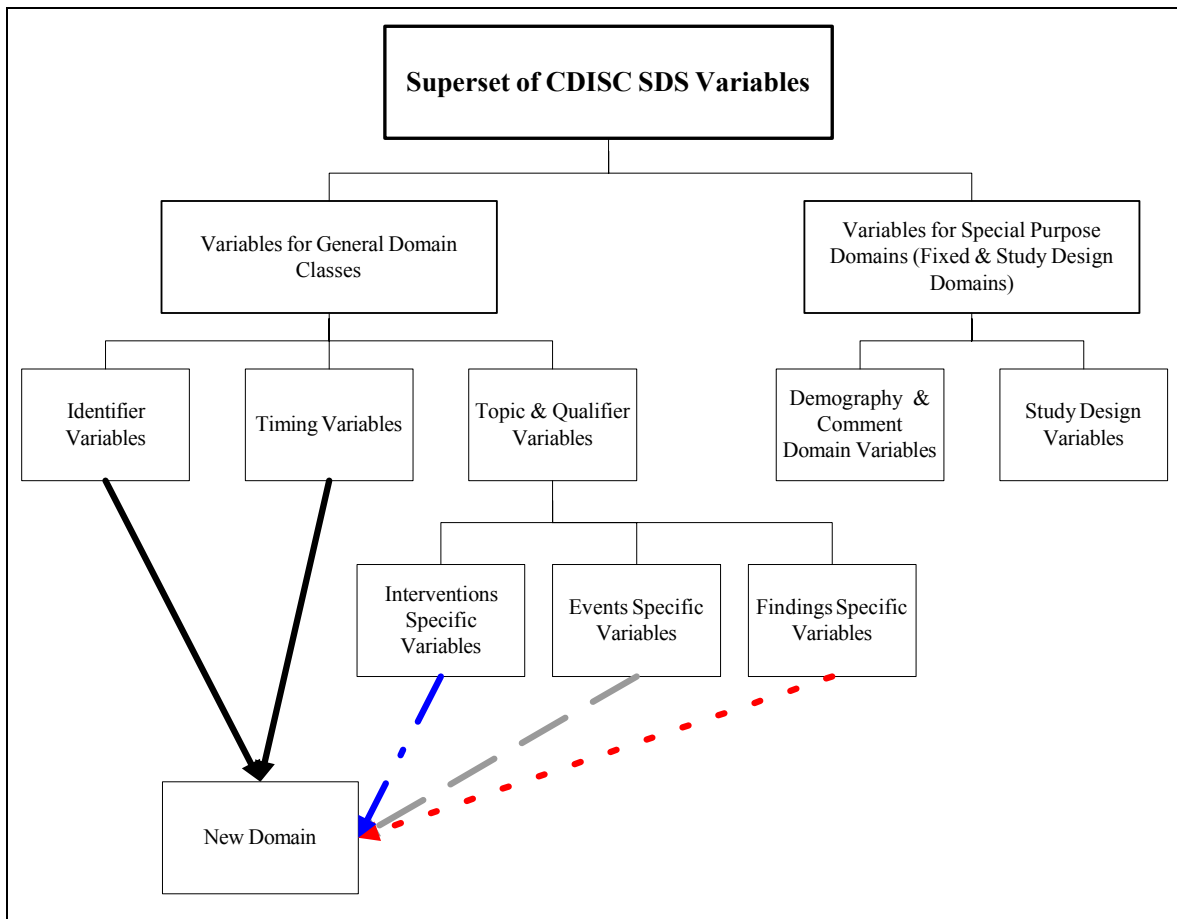
As illustrated in the figure above, the CDISC SDTM has an inclusive superset or a predefined list of variables. The SDTM variables are subdivided into two general variable classes: those for the 'general observation classes' and those for the 'special domains' (those with fixed structures and the trial 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 and 2 above.

Next the sponsor must go through a process of selecting the variables to include in the new domain (Steps 4 through 6). 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 V3.1 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

- Expected data types (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 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 otherwise required by regulatory authorities. 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 meaningful in the context of a specific domain. Columns for Expected variables are assumed to be present in each submitted datasets 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 submitted if 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 V3.1 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 6). 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', 'Role', and 'Comments'.

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.

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, CMSTDTC

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 use type per subject	Tabulation	STUDYID, USUBJID, SUTRT
AE	Adverse Events	AE.XPT	One record per 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, DSTERM, DSDTC
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 measurement per subject	Tabulation	STUDYID, USUBJID, VISITNUM, EGTESTCD, TPTNUM, EGSEQ
IE	Inclusion/Exclusion Exceptions	IE.XPT	One record per I/E criteria not met per subject	Tabulation	STUDYID, USUBJID, IETESTCD
LB	Laboratory Tests	LB.XPT	One record per lab test per subject	Tabulation	STUDYID, USUBJID, VISITNUM, TPTNUM ,LBTESTCD
PE	Physical Exam	PE.XPT	One record per subject or body system or body system finding	Tabulation	STUDYID, USUBJID, VISITNUM, PETESTCD, PESEQ
QS	Questionnaires	QS.XP	one record per question per subject	Tabulation	STUDYID, USUBJID, VISITNUM, TPTNUM, QSTESTCD, QSSEQ
SC	Subject Characteristics	SC.XPT	One record per subject characteristic	Tabulation	STUDYID, USUBJID, SCTESTCD SCSEQ
VS	Vital Signs	VS.XPT	One record per vital sign measurement per subject	Tabulation	STUDYID, USUBJID, VSTESTCD, VISITNUM, TPTNUM, VSSEQ
TE	Trial Elements	TE.XPT	One record per element	Tabulation	STUDYID, ETCD

Dataset	Description	Location	Structure	Purpose	Key Variables
TA	Trial Arms	TA.XPT	One record per element per arm	Tabulation	STUDYID, ARMCD, ETCD
TV	Trial Visits	TV.XPT	One record per 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
RELREC	Related Records	RELREC .XPT	One record per relationship	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, RELID
SUPPQUAL	Supplemental Qualifiers	SUPPQUAL .XPT	One record per qualifier value.	Tabulation	STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM

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 V3.1 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 in an appendix 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 how to submit 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 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

- 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.

Sponsors should always review the CDISC-standard labels and adjust only if necessary to properly convey the meaning of the submitted data. Sponsors should also supply the correct origins, roles, 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 occur 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.

4.1.1.2   Additional analysis variables should be added to separate 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.

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 V3.1, three categories of variables are specified in the 'Core' column in the domain models:

- A **Required** variable is any variable that is basic to the identification 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. Some but not all values for expected variables may be Null in a domain if unknown or not done.

- A **Permissible** variable may be used as appropriate, when collected or derived.
  - All timing variables are considered permissible.
  - Only qualifier variables specified in a domain model 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 SDS Implementation Guide.

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 Timing Variables DMDTC, and DMDY
   d.   The SEQ and GRPID variables in COMMENTS, which actually refer to a separate parent domain
   e.   All variables in RELREC, RELDS, 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. 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   The convention used for missing values (or values that were not collected for certain records in a Tabulation) 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 the records of 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 3 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 the SUPPQUAL 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 SUPPQUAL 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 which has been published externally (e.g., MedDRA, LOINC, maintained by CDISC in Appendix 10.4). The second consists of CT which 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 Terminology or Format column indicates the respective variable should be populated with a value from a sponsor-controlled list. When values appear in the Controlled Terms or Format column, these are expected to be use in submissions. CDISC will be providing future guidance on the use of controlled terminology in a separate document 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 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 the Define document.

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 should not be included in the Tabulation dataset, but 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 agreed-upon high-level term. 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 many domains. 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, is CT. 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 code 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 an '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 has been 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 template for ISO 8601 dates/times is as follows:

YYYY-MM-DDThh:mm:ss

where:

YYYY = four-digit year

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

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

T = separator to indicate 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)

Key aspects of the IS0 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 V3.1 SDS.

- 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. 20011226T00:00:01).

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, and suggested the concept of a precision variable ending in 'DTP'. This has been eliminated in V3.1. Precision, according to ISO 8601, 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. The use of a separate precision variable, as defined in Version 3.0, is no longer required.

The table below provides examples ISO 8601 representation of full and partial date/time values:

	Date and Time as Originally Recorded	Interval of Uncertainty	Nominal Date/Time (--DTC)
1	December 15, 2003 13:14:17	Complete date	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, concatenate the date/time values that describe the beginning and the end of the intervals of uncertainty and separate them with a slash 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	Between the first and the tenth of December, 2003	2003-12-01/2003-12-10
3	Sometime in the first half of 2003	2003-01-01/2003-06-30

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/Time[2]
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 only 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 should not in the submission dataset.

In V3.0, there were separate duration (--DUR) and duration unit (--DURU) variables. In V3.1, 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 the following format:

nYnMnDTnHnMnS or nW

Where 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.

Usage Notes:

- o There should be no spaces in the representation.

- o The 'T' designator must always precede Time components, but is not necessary if no time components are included.

- o Missing components can be omitted.

The table below provides examples of ISO 8601 compliant durations:

Duration as originally recorded	--DUR Value
2 Years	2Y
10 weeks	10W
3 Months 14 days	3M14D
2 Weeks 3 Days	2W3D
6 Months 17 Days 3 Hours	6M17DT3H
14 Days 7 Hours 57 Minutes	14DT7H57M
42 Minutes 18 Seconds	T42M18S

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 only collected once per subject. It is permissible in Disposition if more than one disposition 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 has 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 only 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.

--ENRF identifies the end of the 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). --ENRF should only 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.

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
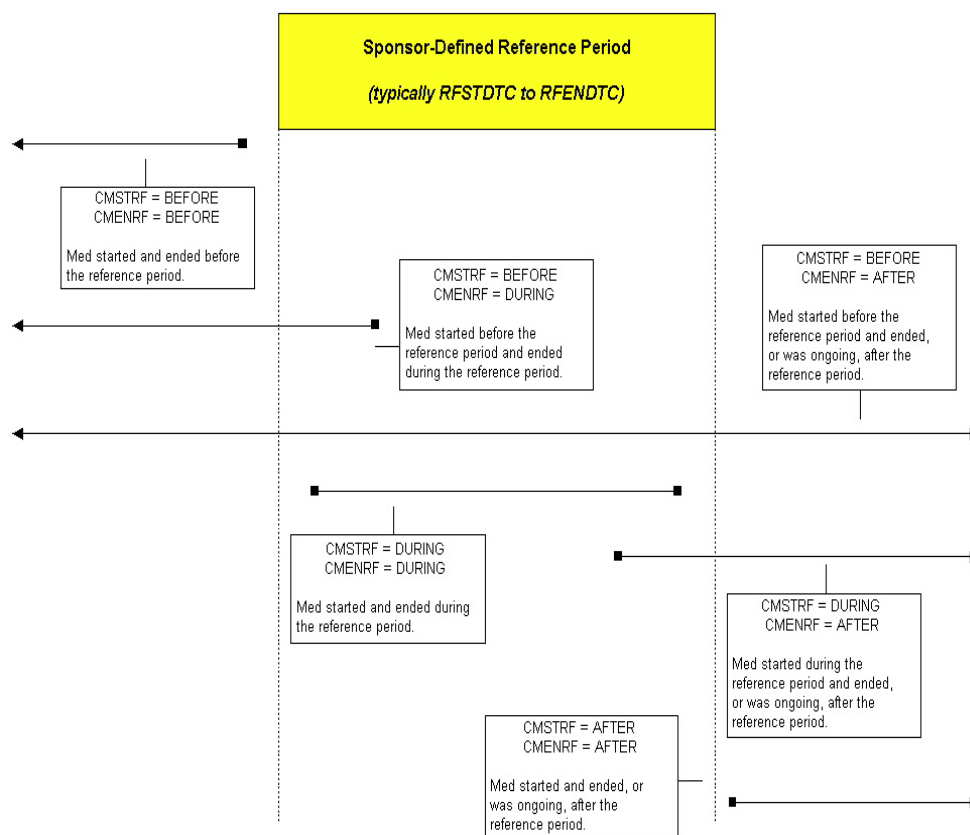
4.1.5 Other Assumptions

4.1.5.1   The --ORRES variables contain the result of the measurement or finding as originally received or collected. --ORRES is an expected variable, and should always be populated except for those instances when --STAT='NOT DONE', in which case --ORRES should be Null.

Results or findings in --ORRES that are numeric will be stored in the variables --STRESC in character format and --STRESN in numeric format.

When a record is derived (e.g., to represent an average or sum of collected values), it may not be possible to populate --ORRES if the collected values are either text or in different units, in which case --ORRES should be Null. Derived records are flagged with the --DRVFL variable.

Except for the above two instances, all results or findings will be converted to standard units or copied from --ORRES into --STRESC for storage in character format, regardless of whether they are character or numeric. If the results or findings contained in --STRESC (standardized or copied from --ORRES) are numeric, then --STRESN will contain the numeric results or findings from --STRESC, therefore there may be Null values as shown in the flowchart below.

- - ORRES (all original values)	------------>	- - STRESC (derive or copy all results)	------------>	- - STRESN (copy numeric results only)

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

When an entire examination (Laboratory draw, ECG, Vital Signs or Physical Examination) or test (glucose, PR interval, blood pressure, or hearing) is not done, a record in the dataset should indicate that the test is not done if the data was explicitly captured with a yes/no or done/not done question. If the data on the CRF is missing a record should not be created to indicate that the data was not collected.

The example explained below may be used when data is explicitly captured on the CRF for a test or result that is not done.

If the individual measurements are not performed, then the value of
- --TESTCD should be <Parameter Short Name>
- --TEST should be <Parameter Name>
- --ORRES should be NULL
- --STAT should be "NOT DONE"

For example, if an ECG parameter (e.g., Heart Rate) is NOT DONE then
- EGTESTCD = "HR"
- EGTEST = "Heart Rate"
- EGORRES=NULL
- EGSTAT="NOT DONE"

If an entire page / module is not done then the value of
- --TESTCD should be --ALL
- --TEST should be <Name of the Module>
- ---ORRES should be NULL
- --STAT should be "NOT DONE"

For example, if an entire ECG page / module is not performed then
- EGTESTCD = "EGALL"
- EGTEST= "ECG"
- EGORRES=NULL
- EGSTAT = "NOT DONE"

When the normal range for a test contains a > (greater than) or < (less than) sign, the value plus the sign cannot be added to the high or low normal range variable. The sponsor must define the appropriate value for all of the normal range variables that contain a > or < sign. For example the sponsor may decide to use one number lower than the value and zero when the < sign is part of the normal range value.

See examples below for various domains:

Laboratory Examples:
- Numeric values that have converted or copied (Examples 1 & 3).
- A character results that has been copied (Example 2).
- A result of 'TRACE' converted to a numeric value of 1 (Example 4).
- Value of 1+ converted to numeric 1 (Example 5).
- A result of 'BLQ' converted to a numeric value of zero (Example 6).
- Some original character results in LBORRES such as '<10,000', which are converted to numeric results per sponsor-defined algorithms (Example 7).

- A result is missing because the observation was 'NOT DONE', as reflected in the --STAT variable; neither LBORRES nor LBSTRESC are populated (Example 8).
- A result is derived from multiple records such as an average of baseline measurements for a baseline value (Example 9).
- The entire examination was not done (Example 10)

	LBTESTCD	LBSTAT	LBORRES	LBORR ESU	LBSTRE SC	LBSTRE SN	LBSTRE SU	LBDRV FL
1	GLUCOSE		6.0	mg/dL	60.0	60.0	mg/L	
2	UBAC		MODERAT E		MODER ATE			
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						

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

	EGTEST CD	EGSTAT	EGORRES	EGORRESU	EGSTRESC	EGSTRESN	EGSTRESU
1	QRS		362	SEC	362	362	SEC
2	QT		221	MSEC	.221	.221	SEC
3	Rhythm and Rate		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 (Examples 1 and 2).
- A result is missing because the observation was 'NOT DONE', as reflected in the VSSTAT variable; neither VSORRES nor VSSTRESC are populated (Example 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. (Examples 4, 5, and 6).
- The entire examination was not done (Example 7)

	VSTEST CD	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	DISBP		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 (Example1).
- A result is missing because the observation was 'NOT DONE', as reflected in the QSSTAT variable; neither QSORRES nor QSSTRESC are populated (Example 3).
- The entire examination was not done (Example 4)

Example	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 or RELDS datasets 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 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. 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	IDVAR VAL	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. Since observations recorded in the Findings class may be provided by either the Investigator or others, these domains include the --EVAL qualifier. 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 "1". 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, June 25, 2004. 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.	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. .	Req	SDSIG 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.	Req	SDSIG 4.1.4.1
SITEID	Study Site Identifier	Char		CRF or Derived	Record Qualifier	Unique identifier for the 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	
BRTHDTC	Date/Time of Birth	Char	ISO 8601	CRF or Derived	Result Qualifier	Date/time of birth of the subject.	Perm	SDSIG 4.1.4.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
AGE	Age in AGEU at Reference Date/Time	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, or DAYS	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	**HISPANIC, NON-HISPANIC	CRF	Result Qualifier	Ethnicity of subject.	Perm	
ARMCD	Planned Arm Code	Char	*	CRF or Derived	Result Qualifier	Short, 8-character version of ARM, used for 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	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	SDSIG 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	SDSIG 4.1.4.4

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

5.1.1.0   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 be present; investigator will be used to group data. This should be done consistently for the submission and the meaning of the variable made clear in the metadata definition. Investigator name may also be included as a support variable (INVNAM).

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 3.1 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 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/2353fnl.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 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 above.

5.1.2 Comments Domain Model — CO

The COMMENTS special-purpose domain, described in Section 8.5, 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 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]

CO.xpt -- Version 3.1. One record per comment per subject, June 25, 2004, Tabulation

---

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

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	Domain Abbreviation of the parent record(s) to which the comment refers, or on which the comment was collected. Use CO for general comments that do not relate to another domain.	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 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	Identifier Variable Name	Char	**	Derived	Identifier	Identifier variable in the dataset that identifies the related record(s). Examples --SEQ, or --GRPID. Null for comments collected on separate CRFs.	Perm	
IDVARVAL	Identifier Variable	Char		Derived	Identifier	Value of identifier 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. VITALS, or a combination of both e.g. 650-VITALS.	Perm	
CODTC	Date/Time of Comment	Char	ISO 8601	CRF	Timing	Use if collected.	Perm	SDSIG 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, PRINCIPAL INVESTIGATOR.	Perm	

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

5.1.2.0   Assumptions for Comments (CM) 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[2]; 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 IDVAR are the same as for the Supplemental Qualifiers (SUPPQUAL) special purpose dataset, 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 its use is required to unambiguously 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 5 and 6 (GRPID and SEQ) 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 both GRPID and SEQ are NULL.

STUDYID	DOMAIN	RDOMAIN	USUBJID	COSEQ	IDVAR	IDVARVAL	COREF	COVAL	COVAL1	COVAL2	COEVAL
1234	CO	AE	AB-99	1	AESEQ	7	650	200*A	200*B	100*C	PRINCIPAL INVESTIGATOR
1234	CO	EX	AB-99	2	EXGRPID	17	320-355	200*X	200*Y		PRINCIPAL INVESTIGATOR
1234	CO	VS	AB-99	3	EXGRPID	5		200*V	2*W		PRINCIPAL INVESTIGATOR
1234	CO		AB-99	4			800	150*P			PRINCIPAL INVESTIGATOR

2 As a result of the SDS V3.0 pilot, it was determined that the comments collected on the individual domains such as AECOM should be moved into the comments domain since they are also subject to the 200-character limitation.

6 Domain Models Based on the General Classes

6.1 INTERVENTIONS

6.1.1 Concomitant Medications — CM

CM.xpt, Concomitant Meds — Interventions, Version 3.1, June 25, 2004. 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	SDSIG 2.1; SDTM 2.2.4
CMSPID	Sponsor ID	Char		Sponsor Defined	Record Qualifier	Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database.Example: page number.	Perm	SDTM 2.2.4

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	SDSIG 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	SDSIG 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	SDSIG 4.1.3.5
CMCAT	Category for Medication	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records. Examples: ANTI-CANCER MEDICATION, or GENERAL CONMED.	Perm	SDSIG 2.1
CMSCAT	Subcategory for Medication	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of a measurement or examination. Examples: CHEMOTHERAPY, RADIOTHERAPY, HORMONAL THERAPY, ALTERNATIVE THERAPY.	Perm	SDSIG 2.1
CMOCCUR	CM Occurrence	Char	**Y, N or Null	CRF or Sponsor Defined	Result Qualifier	Used when the use of specific medications is solicited. Values are Null for spontaneously solicited medications.	Perm	
CMSTAT	Concomitant Medication Status	Char	**NOT DONE	CRF or Sponsor Defined	Result Qualifier	The status indicates that the question was not asked.	Perm	SDSIG 4.1.5.1.
CMREASND	Reason Concomitant 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	SDSIG 4.1.5.3
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	SDSIG 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	Synonym Qualifier	Use only when the dictionary used codes to a single class.	Perm	SDSIG 4.1.3.5
CMDOSE	Dose per Administration	Num		CRF or Sponsor Defined	Result Qualifier	Amount of CMTRT taken.	Perm	
CMDOSTXT	Dose Description	Char		CRF or Derived	Result 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	Result Qualifier	Dose form for CMTRT. Examples: TABLET, LOTION, or INJECTABLE.	Perm	
CMDOSFRQ	Dosing Frequency Per Interval	Char	*	CRF or Sponsor Defined	Result 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	Result Qualifier	Total daily dose of CMTRT using the units in CMDOSU.	Perm	
CMDOSRGM	Intended Dose Regimen	Char		CRF or Derived	Result 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	Result Qualifier	Route of administration for CMTRT. Examples: ORAL, INTRAVENOUS.	Perm	
CMSTDTC	Start Date/Time of Medication	Char	ISO 8601	CRF	Timing		Perm	SDSIG 4.1.4.1
CMENDTC	End Date/Time of Medication	Char	ISO 8601	CRF	Timing		Perm	SDSIG 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	SDSIG 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	SDSIG 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	*	Derived	Timing	Identifies the start of the medication 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 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. CMSTRF should only 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	
CMENRF	End Relative to Reference Period	Char	**	Derived	Timing	Identifies the end of the medication 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 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. Medications that are ongoing at the end of the reference period should have a value of AFTER for this variable. CMENRF should only 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	

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

6.1.1.0   Assumptions for Concomitant Medications (CM) domain model

1. CM CRT 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).
    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. 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 in a single domain to support relationships within the domain and between domains.

4. Information regarding the Concomitant Medications/Therapy
    a. CMOCCUR should be used when a CRF includes a pre-specified list of Concomitant Medications/Therapies. Then indicate in CMOCCUR whether a medication/therapy was used.

5. Timing variables

    a. CMSTREF should indicate the start of the medication as being BEFORE, DURING, or AFTER the sponsor-defined reference period. This should only be populated when a medication start date is not collected. The sponsor-defined reference period is a continuous period of time defined by a discrete starting point (RFSTDTC) and a discrete ending point (RFENDTC). Sponsors should define the reference period in the study metadata.
    b. CMENREF should indicate the end of the medication as being BEFORE, DURING, or AFTER the sponsor-defined reference period. This should be used only when an end date is not collected. The sponsor-defined reference period should be indicated using the RFSTDTC and RFENDTC variables defined in the Demography domain. For medications that were ongoing at the end of the reference period, CMENRF would have the value AFTER. Sponsors should define the reference period in the study metadata.

6.1.2 Exposure — EX

EX.xpt, Exposure — Interventions, Version 3.1, June 25, 2004. 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	SDSIG 2.1; SDTM 2.2.4
EXSPID	Sponsor Reference Number	Char		Sponsor Defined	Record Qualifier	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
EXTRT	Name of Actual Treatment	Char		CRF or Randomization File or Derived	Topic	Name of the intervention treatment — usually the verbatim name of the treatment, drug, medicine or therapy given during the dosing period for the observation.	Req	SDSIG 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	SDSIG 2.1
EXSCAT	Subcategory for Treatment	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of treatment.	Perm	SDSIG 2.1
EXDOSE	Dose per Administration	Num		CRF or Sponsor Defined	Result 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	Result 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	Result Qualifier	Dose form for EXTRT. Examples: TABLET, LOTION, or INJECTABLE.	Perm	
EXDOSFRQ	Dosing Frequency Per Interval	Char	*	CRF or Sponsor Defined	Result 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	Result Qualifier	Total daily dose of EXTRT using the units in EXDOSU.	Perm	
EXDOSRGM	Intended Dose Regimen	Char		CRF or Derived	Result 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	Result 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	
TAETORD	Order of Element within Arm	Num		Sponsor Defined / Protocol	Timing	Number that gives the order of the Element within the arm.	Perm	
EXSTDTC	Start Date/Time of Treatment	Char	ISO 8601	CRF or Derived	Timing		Req	SDSIG 4.1.4.1
EXENDTC	End Date/Time of Treatment	Char	ISO 8601	CRF or Derived	Timing		Exp	SDSIG 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	SDSIG 4.1.4.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
EXENDY	Study Day of End of Treatment	Num		Derived	Timing	Study day of end of treatment relative to the sponsor-defined RFSTDTC.	Perm	SDSIG 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 EXTPTREF. 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 (EXTPTREF). This variable is useful where there are repetitive measures. Not a clock time.	Perm	
EXTPTREF	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.

6.1.2.0   Assumptions for Exposure (EX) domain model

1. EX CRT Definition

a)   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
EXCAT and EXSCAT may be used when appropriate to categorized treatments into categories and subcategories. For example, if a study contains several active comparator medications, EXCAT may be set to 'ACTIVE COMPARATOR.' Such categorization will not be useful in most studies, so these variables are permissible but not expected.

6.1.3 Substance Use - SU

SU.xpt, Substance Use — Interventions, Version 3.1, June 25, 2004. 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	SDSIG 2.1; SDTM 2.2.4
SUSPID	Sponsor Reference Number	Char		Sponsor Defined	Record Qualifier	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
SUTRT	Name of Substance	Char		CRF	Topic	Substance name. Examples: Cigarettes, Coffee.	Req	SDSIG 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	SDSIG 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	SDSIG 4.1.3.5

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
SUCAT	Category for Substance Use	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records. Examples: TOBACCO, ALCOHOL, or CAFFEINE.	Perm	SDSIG 2.1
SUSCAT	Subcategory for Substance Use	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of substance use. Examples: CIGARS, CIGARETTES, BEER, WINE	Perm	SDSIG 2.1
SUOCCUR	SU Occurrence	Char	**Y, N or Null	CRF or Sponsor Defined	Result Qualifier	Used when the use of specific substances is solicited. Values are Null for spontaneously reported substances.	Perm	
SUSTAT	Substance Use Status	Char	**NOT DONE	CRF	Result 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	SDSIG 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	SDSIG 4.1.3.5
SUCLASCD	Substance Use Class Code	Char	*	Derived	Variable Qualifier	May apply when coding substance abuse cases.	Perm	SDSIG 4.1.3.1
SUDOSE	Substance Use Consumption	Num		CRF or Sponsor Defined	Result Qualifier	Amount of SUTRT consumed.	Perm	
SUDOSTXT	Substance Use Consumption Text	Char		CRF or Derived	Result 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	Result Qualifier	Dose form for SUTRT. Examples: INJECTABLE, LIQUID, or POWDER.	Perm	
SUDOSFRQ	Use Frequency Per Interval	Char	*	CRF or Sponsor Defined	Result 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	Result Qualifier	Total daily use of SUTRT using the units in SUDOSU.	Perm	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
SUROUTE	Route of Administration	Char	*	Sponsor Defined	Result Qualifier	Route of administration for CMTRT. 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	SDSIG 7.8; SDSIG 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	SDSIG 7.8; SDSIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDSIG 7.8; SDSIG 4.1.4.5
SUSTDTC	Start Date/Time of Substance Use	Char	ISO 8601	CRF or Derived	Timing		Perm	SDSIG 4.1.4.1
SUENDTC	End Date/Time of Substance Use	Char	ISO 8601	CRF or Derived	Timing		Perm	SDSIG 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	SDSIG 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	SDSIG 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	

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

6.1.3.0   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 CRT 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 (SUTERM) 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 SUs 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 within SU. It should not be used in place of SUCAT or SUSCAT. It is not necessary to populate this variable

5. Presence or Absence of Substance Use
    If specific substances are solicited (i.e., a pre-specified list of terms is collected), SUOCCUR should be used to indicate whether each substance was consumed.

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, June 25.2004. One record per 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 ID	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	SDSIG 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	SDSIG 4.1.3.6

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
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	SDSIG 4.1.3.5; SDSIG 4.1.3.6
AECAT	Category for Adverse Event	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records.	Perm	SDSIG 4.1.2.6
AESCAT	Subcategory for Adverse Event	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of adverse event.	Perm	SDSIG 4.1.2.6
AEOCCUR	Adverse Event Occurrence	Char	**Y, N or Null	CRF or Sponsor Defined	Result Qualifier	Used when the occurrence of specific adverse events is solicited. Values are Null for spontaneously reported events.	Perm	
AEBODSYS	Body System or Organ Class	Char	**	CRF or Derived	Result Qualifier	Body system or organ class (Primary SOC) that is involved in an event or measurement from the standard hierarchy (e.g., MedDRA).	Exp	SDSIG 4.1.3.5
AELOC	Location of the Reaction	Char	*	CRF or Derived	Record Qualifier	Describes anatomical location relevant for the event. (e.g., Example: 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 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	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
AEREL	Causality	Char	*	CRF	Result Qualifier	Records the investigator's opinion as to the causality of the event to the treatment. Examples: DEFINITELY NOT RELATED, PROBABLY NOT RELATED, POSSIBLY RELATED, PROBABLY RELATED and DEFINITELY RELATED	Exp	
AERELNST	Relationship to Non-Study Treatment	Char		CRF	Result Qualifier	Records the investigator's opinion as to whether the event may have been to due to a treatment other than study drug. Reported as free text. Example: "More likely related to aspirin use	Perm	
AEPATT	Pattern of Event	Char	*	CRF	Record Qualifier	Used to indicate the pattern of the event over time. Examples: INTERMITTENT, CONTINUOUS, SINGLE EVENT.	Perm	
AEOUT	Outcome of Adverse Event	Char	*	CRF	Result Qualifier	Description of the outcome of an event. Examples include ICHE2B 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	Result Qualifier	Was the event associated with the development of cancer?	Perm	
AESCONG	Congenital Anomaly or Birth Defect	Char	**Y, N or Null	CRF	Result Qualifier	Was the event associated with congenital anomaly or birth defect?	Perm	
AESDISAB	Persist or Signif Disability/Incapacity	Char	**Y, N or Null	CRF	Result Qualifier	Did the event result in persistent or significant disability/incapacity?	Perm	
AESDTH	Results in Death	Char	**Y, N or Null	CRF or Derived	Result Qualifier	Did the event result in death?	Perm	
AESHOSP	Requires or Prolongs Hospitalization	Char	*Y, N or Null	CRF	Result Qualifier	Did the event require or prolong hospitalization?	Perm	
AESLIFE	Is Life Threatening	Char	**Y, N or Null	CRF	Result Qualifier	Was the event life threatening?	Perm	
AESOD	Occurred with Overdose	Char	**Y, N or Null	CRF	Result Qualifier	Did the event occur with an overdose?	Perm	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
AESMIE	Other Medically Important Serious Event	Char	**Y, N or Null	CRF	Result 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 which scale and version are used.	Perm	
AESTDTC	Start Date/Time of Adverse Event	Char	ISO 8601	CRF or Derived	Timing		Exp	SDSIG 4.1.4.1; SDSIG 4.1.4.2
AEENDTC	End Date/Time of Adverse Event	Char	ISO 8601	CRF or Derived	Timing		Exp	SDSIG 4.1.4.1; SDSIG 4.1.4.2
AESTDY	Study Day of Start of Event	Num		Derived	Timing	Study day of start of adverse event relative to the sponsor-defined RFSTDTC.	Perm	SDSIG 4.1.4.4
AEENDY	Study Day of End of Event	Num		Derived	Timing	Study day of end of event relative to the sponsor-defined RFSTDTC.	Perm	SDSIG 4.1.4.4
AEDUR	Duration of 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: 1DT2H.	Perm	SDSIG 4.1.4.3
AEENRF	End Relative to Reference Period	Char	**	Derived	Timing	Identifies the end of the event 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 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. AEENRF should only 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 AEENRF.	Perm	SDSIG 4.1.4.7

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

6.2.1.0 Assumptions for Adverse Events (AE) domain model

1. AE CRT Definition
   a. CRF data that capture adverse events either as free text or a pre-specified list of terms should be included within the AE CRT.

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 AEs 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 within AE. It should not be used in place of AECAT or AESCAT.

4. Presence or Absence of Events
   Adverse events are generally collected in two different ways, either by recording free text or using a pre-specified list of terms.
   a. When adverse events are collected with the recording of free text, a record may be entered 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. Records should be included in the submission dataset only for adverse events that have actually occurred.
   b. If specific adverse events or a pre-specified list of terms are collected, AEOCCUR should be used to indicate whether each event occurred.
   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.

5. Timing Variables
   a. AEENDTC or AEENRF should not both 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.

Final

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 SUPPQUAL 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).

6.2.2   Disposition — DS

DS.xpt, Disposition — Events Version 3.1, June 25, 2004. 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	SDSIG 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 ID	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	Disposition Term	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	SDSIG 4.1.3.6

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
DSDECOD	Controlled Terminology for DSTERM	Char	*	CRF or Derived	Synonym Qualifier	Controlled terminology for the name of disposition event or protocol milestone. Examples: COMPLETED, INFORMED CONSENT OBTAINED, INSUFFICIENT CLINICAL RESPONSE, ADVERSE EVENT, DEATH, PROTOCOL VIOLATION, LOST TO FOLLOW-UP, DOES NOT MEET ENTRANCE CRITERIA, RANDOMIZED, OTHER.	Req	SDSIG 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	SDSIG 2.1
DSSCAT	Subcategory for Disposition Event	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of disposition event.	Perm	SDSIG 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	SDSIG 7.8; SDSIG 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	SDSIG 7.8; SDSIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDSIG 7.8; SDSIG 4.1.4.5
EPOCH	Stage of Disposition Event	Char	*	CRF or Derived	Record Qualifier	EPOCH is used when DSCAT = 'DISPOSITION EVENT'. Examples: TREATMENT PHASE, SCREENING, FOLLOW-UP.	Perm	SDSIG 7.4.4
DSDTC	Date/Time of Collection	Char	ISO 8601	CRF or Derived	Timing	Use if collected.	Perm	SDSIG 4.1.4.1
DSSTDTC	Start Date/Time of Disposition Event	Char	ISO 8601	CRF or Derived	Timing	Use if collected.	Exp	SDSIG 4.1.4.1
DSSTDY	Study Day of Start of Disposition Event	Num		Derived	Timing	Study day of start of disposition event relative to the sponsor-defined RFSTDTC.	Perm	SDSIG 4.1.4.4

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

6.2.2.0 Assumptions for Disposition (DS) domain model

1. DS CRT Definition
    a. CRF data that captures disposition status and events that are protocol milestones. Disposition status documents whether or not the subject completed a phase or segment of the trial (e.g., screening, treatment, follow-up). Protocol milestones are planned points in time such as randomization date and date of informed consent. 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 is 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, June 25, 2004. 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	SDSIG 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 ID	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	SDSIG 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	SDSIG 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 Define Sponsor Comments column.	Exp	SDSIG 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	SDSIG 2.1

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	SDSIG 2.1
MHOCCUR	Medical History Occurrence	Char	**Y, N or Null	CRF or Sponsor Defined	Result Qualifier	Used when the occurrence of medical history events is solicited. Values are Null for spontaneously reported events	Perm	
MHSTAT	Medical History Status	Char	**NOT DONE	CRF or Sponsor Defined	Result Qualifier	The status indicates that the question was not asked.	Perm	SDSIG 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	SDSIG 4.1.5.3
MHBODSYS	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.	Exp	SDSIG 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	SDSIG 7.8; SDSIG 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	SDSIG 7.8; SDSIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDSIG 7.8; SDSIG 4.1.4.5
MHDTC	Date/Time of Collection	Char	ISO 8601	CRF or Derived	Timing		Perm	SDSIG 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	SDSIG 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	SDSIG 4.1.4.4

June 25, 2004   Final

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
MHENRF	End Relative to Reference Period	Char	**	Derived	Timing	Identifies the end of the event 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 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. MHENRF should only 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 MHENRF.	Perm	SDSIG 4.1.4.7

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

6.2.3.0   Assumptions for Medical History (MH) domain model

1. MH CRT Definition
   a. CRF data that captures subject medical history information, both conditions and events, that are independent of planned study evaluations and occur prior to the clinical trial should be included within the MH CRT.
   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. MHDECOD is the preferred term derived by the sponsor from the coding dictionary. It is expected that the reported term (MHTERM) 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).
   d. MHBODSYS 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. 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 within MH. It should not be used in place of MHCAT or MHSCAT. 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. 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.
   b. If specific medical history events are solicited (i.e., a pre-specified list of terms is collected), MHOCCUR should be used to indicate whether each event occurred.
   c. When the response for a medical history question on a CRF was not collected, then the value of MHSTAT = 'NOT DONE' and MHOCCUR will be Null.
   d. 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.
   e. MHSTAT should be used to indicate that medical history was not done. The reason that medical history was not done may be documented in MHREASND.

5. Timing Variables
   a. Either MHENDTC or MHENRF should be populated, but not both.

Final

6.3 FINDINGS

6.3.1 ECG Test Results — EG

EG.xpt, ECG — Findings, Version 3.1, July 14, 2004. 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 link together a block of related records for a subject in a single domain.	Perm	SDSIG 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 ID	Char		Sponsor Defined or Derived	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	SDSIG 2.1
EGSCAT	Subcategory for ECG	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of the ECG.	Perm	SDSIG 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.	Exp	
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 measurement, or ATRIAL FIBRILLATION or QT PROLONGATION when EGTEST is an interpretation.	Exp	SDSIG 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	SDSIG 4.1.5.1
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	SDSIG 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	SDSIG 4.1.5.1
EGSTRESU	Standard Units	Char	*	CRF or Derived	Variable Qualifier	Standardized unit used for EGSTRESC or EGSTRESN.	Perm	SDSIG 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	Result 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	SDSIG 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. 2. The LOINC version should be provided in the comments column in the data definition file.	Perm	SDSIG 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	SDSIG 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 both EGSTRESC and EGSTRESN having the derived value.	Perm	SDSIG 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. Controlled terminology will consist of values such as INVESTIGATOR, ADJUDICATION COMMITTEE, VENDOR.	Exp	
VISITNUM	Visit Number	Num		CRF or Derived	Timing	1. Clinical encounter number. 2. Numeric version of VISIT, used for sorting.	Req	SDSIG 7.8; SDSIG 4.1.4.5

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
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	SDSIG 7.8; SDSIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDSIG 7.8; SDSIG 4.1.4.5
EGDTC	Date/Time of ECG	Char	ISO 8601	CRF or Derived	Timing		Exp	SDSIG 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	SDSIG 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.	Perm	SDSIG 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.

6.3.1.0 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.

6.3.2 Inclusion/Exclusion Exceptions — IE

IE.xpt, Inclusion/Exclusion Exceptions — Findings, Version 3.1, June 25, 2004. 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 ID	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	Exception 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	Exception Criterion	Char	*	CRF	Synonym Qualifier	Verbatim description of the inclusion or exclusion criterion that was the exception for the subject.	Req	
IECAT	Exception Criteria Category	Char	**INCLUSION, EXCLUSION	Sponsor Defined	Grouping Qualifier	Used to define a category of related records.	Req	SDSIG 2.1
IESCAT	Exception Criteria Subcategory	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of the exception criterion. Can be used to distinguish criteria for a sub-study.	Perm	SDSIG 2.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
IEORRES	Criterion Met	Char	**Y, N	CRF or Derived	Result Qualifier	Inclusion or Exclusion criterion met?	Req	SDSIG 4.1.5.1
IESTRESC	Character Result/Finding in Std Format	Char	**Y, N	Derived	Result Qualifier	Inclusion or Exclusion criterion met? (Same as IEORRES for the IE Domain).	Req	SDSIG 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	SDSIG 7.8; SDSIG 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	SDSIG 7.8; SDSIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDSIG 7.8; SDSIG 4.1.4.5
IEDTC	Date/Time of Collection	Char	ISO 8601	CRF or Derived	Timing		Exp	SDSIG 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	SDSIG 4.1.4.4 SDTM 2.2.6

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

6.3.2.0   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 will be developed for protocol deviations/violations in a future version of the Implementation Guide.

6.3.3 Laboratory Test Results — LB

LB.xpt, Labs — Findings, Version 3.1, July 14, 2004. 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	SDSIG 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 ID	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	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
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	SDSIG 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	SDSIG 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	SDSIG 4.1.5.1
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	SDSIG 4.1.3.2
LBORNRLO	Normal Range Lower Limit in Orig Units	Char		CRF or Derived	Variable Qualifier	Lower end of normal range for continuous measurements in original units. Should be populated only for continuous results.	Exp	
LBORNRHI	Normal Range Upper Limit in Orig Units	Char		CRF or Derived	Variable Qualifier	Upper end of normal 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	SDSIG 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.).	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	SDSIG 4.1.5.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
LBSTRESU	Standard Units	Char	*	CRF or Derived	Variable Qualifier	Standardized unit used for LBSTRESC or LBSTRESN.	Exp	
LBSTNRLO	Normal Range Lower Limit-Standard Units	Num		CRF or Derived	Variable Qualifier	Lower end of normal range for continuous measurements in standardized units. SI units suggested for lab data. Should be populated only for continuous results.	Exp	
LBSTNRHI	Normal Range Upper Limit-Standard Units	Num		CRF or Derived	Variable Qualifier	Upper end of normal range for continuous measurements in standardized units. SI units suggested for lab data. Should be populated only for continuous results.	Exp	
LBNRIND	Reference Range Indicator	Char	*	CRF or Derived	Variable Qualifier	Indicates where value falls with respect to reference range defined by ---ORNRLO and --ORNRHI or by --STNRC. Examples: NORMAL, ABNORMAL, H, L.	Exp	
LBSTAT	Lab Status	Char	**NOT DONE	CRF or Derived	Result Qualifier	Used to indicate exam not done. Should be Null if a result exists in LBORRES.	Perm	SDSIG 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	
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	SDSIG 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: SWAB, NEEDLE PUNCTURE, BRONCHIAL BRUSHING.	Perm	
LBBLFL	Baseline Flag	Char	**Y or Null	CRF or Derived	Record Qualifier	Indicator used to identify a baseline value.	Exp	SDSIG 4.1.3.7

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
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 both LBSTRESC and LBSTRESN having the derived value.	Perm	SDSIG 4.1.5.1 SDSIG 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).	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	SDSIG 7.8; SDSIG 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	SDSIG 7.8; SDSIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDSIG 7.8; SDSIG 4.1.4.5
LBDTC	Date/Time of Specimen Collection	Char	ISO 8601	CRF or Derived	Timing		Exp	SDSIG 4.1.4.1
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	SDSIG 4.1.4.4
LBTPT	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 LBTPTNUM and LBTPTREF. Examples: Start, 5 min post.	Perm	

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
LBTPTNUM	Planned Time Point Number	Num		CRF or Derived	Timing	Numerical version of LBTPT 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 (LBTPTREF) such as 'PREVIOUS DOSE' or 'PREVIOUS MEAL'. This variable is useful where there are repetitive measures. Not a clock time or a date time variable.	Perm	SDSIG 4.1.4.3
LBTPTREF	Time Point Reference	Char		Sponsor Defined	Timing	Name of the fixed reference point referred to by LBELTM, LBTPTNUM, and LBTPT. Examples: PREVIOUS DOSE, PREVIOUS MEAL.	Perm	

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

6.3.4 Physical Examinations — PE

PE.xpt, Physical Examination — Findings, Version 3.1, June 25. One record per body system 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	SDSIG 2.1; SDTM 2.2.4

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
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 for the body system. 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	SDSIG 4.1.3.6
PECAT	Category for Examination	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of examination. Examples: GENERAL, NEUROLOGICAL.	Perm	SDSIG 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	SDSIG 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	
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	SDSIG 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	SDSIG 4.1.5.1
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	SDSIG 4.1.5.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
PESTAT	Examination Status	Char	**NOT DONE	CRF or Derived	Result Qualifier	Used to indicate exam not done. Should be Null if a result exists in ORRES.	Perm	SDSIG 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	
PESPID	Sponsor ID	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
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	SDSIG 7.8; SDSIG 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	SDSIG 7.8; SDSIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDSIG 7.8; SDSIG 4.1.4.5
PEDTC	Date/Time of Examination	Char	ISO 8601	CRF or Derived	Timing		Exp	SDSIG 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	SDSIG 4.1.4.4

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

6.3.4.0 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, June 25, 2004. 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	SDSIG 2.1; SDTM 2.2.4
QSSPID	Sponsor ID	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. Examples: HAMILTON DEPRESSION SCLAE, SF36, ADAS.	Req	SDSIG 2.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
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	SDSIG 2.1
QSORRES	Finding in Original Units	Char		CRF or Derived	Result Qualifier	Finding as originally received or collected (e.g. RARELY, SOMETIMES).	Exp	SDSIG 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.	Perm	
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	SDSIG 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	SDSIG 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	Result Qualifier	Used to indicate a questionnaire or response to a questionnaire was not done. Should be Null if a result exists in QSORRES.	Perm	SDSIG 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	
QSBLFL	Baseline Flag	Char	**Y or Null	CRF or Derived	Record Qualifier	Indicator used to identify a baseline value.	Exp	SDSIG 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 subscores 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 both QSSTRESC and QSSTRESN having the derived value.	Perm	SDSIG 4.1.5.1 SDSIG 4.1.3.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
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	SDSIG 7.8; SDSIG 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	SDSIG 7.8; SDSIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDSIG 7.8; SDSIG 4.1.5
QSDTC	Date/Time of Finding	Char	ISO 8601	CRF or Derived	Timing		Exp	SDSIG 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	SDSIG 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	
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.	Perm	SDSIG 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	

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

6.3.6 Subject Characteristics — SC

SC.xpt, Subject Characteristics — Findings, Version 3.1, July 14, 2004. 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	SDSIG 2.1; 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. Examples: EYECOLOR, SUBJINIT.	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: Eye Color, Subject Initials.	Req	
SCCAT	Category for Subject Characteristic	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records.	Perm	SDSIG 2.1
SCSCAT	Subcategory for Subject Characteristic	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of the subject characteristic.	Perm	SDSIGA 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	SDSIG 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	SDSIG 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	SDSIG 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	Result Qualifier	Used to indicate that the measurement was not done. Should be Null if a result exists in SCORRES.	Perm	SDSIG 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	
SCSPID	Sponsor ID	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
SCDTC	Date/Time of Collection	Char	ISO 8601	CRF or Derived	Timing		Exp	SDSIG 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	SDSIG 4.1.4.4

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

6.3.6.0 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, eye color, childbearing status, allergy status, etc.

6.3.7 Vital Signs — VS

VS.xpt, Vital Signs — Findings, Version 3.1, June 25, 2004. 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	SDSIG 2.1; SDTM 2.2.4
VSSPID	Sponsor ID	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	SDSIG 2.1
VSCAT	Category for Vital Signs	Char	*	Sponsor Defined	Grouping Qualifier	Used to define a category of related records.	Perm	SDSIG 2.1

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	Reference
VSSCAT	Subcategory for Vital Signs	Char	*	Sponsor Defined	Grouping Qualifier	A further categorization of a measurement or examination.	Perm	SDSIG 2.1
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.	Exp	
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	SDSIG 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	SDSIG 4.1.5.1
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	SDSIG 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	SDSIG 4.1.5.1
VSSTRESU	Standard Units	Char	*	CRF or Derived	Variable Qualifier	Standardized unit used for VSSTRESC or VSSTRESN.	Exp	SDSIG 4.1.5.1
VSSTAT	Vitals Status	Char	**NOT DONE	CRF or Derived	Result Qualifier	Used to indicate that a vital sign measurement was not done. Should be Null if a result exists in VSORRES.	Perm	
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	SDSIG 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 comments column in the data definition file.	Perm	SDSIG 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	SDSIG 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	ExpSdS	SDSIG 4.1.5.1
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 subscores 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 both VSSTRESC and VSSTRESN having the derived value.	Perm	SDSIG 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	SDSIG 7.8; SDSIG 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	SDSIG 7.8; SDSIG 4.1.4.5
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing		Perm	SDSIG 7.8; SDSIG 4.1.4.5
VSDTC	Date/Time of Measurements	Char	ISO 8601	CRF or Derived	Timing		Exp	SDSIG 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	SDSIG 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.	Perm	SDSIG 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, 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.

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, June 25, 2004. One record per Element, Tabulation

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, which must be TE.	Req
ETCD	Element Code	Char	*	Sponsor Defined	Record Qualifier	Short 8-characterversion of ELEMENT, used for programming.	Req
ELEMENT	Description of Element	Char	*	Sponsor Defined / Protocol	Topic	Topic variable for an Element -- 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.	Req
TEDUR	Planned Duration of Element	Char		Sponsor Defined/Protocol	Timing	Planned Duration of Element in ISO 8601 format. For use when the rule for ending the Element is to end 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, June 25, 2004. One record per Element per Arm, Tabulation

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, which must be TA.	Req
ARMCD	Arm Code (Treatment Code)	Char	*	Derived	Topic	Short 8-character version of ARM, used for programming. If the timing of visits for a trial does not depend on which ARM a subject is in, then ARMCD should be Null.	Req
ARM	Description of Arm (Treatment Group)	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 8-character version of ELEMENT, used for programming.	Req
ELEMENT	Description of Element	Char	*	Sponsor Defined / Protocol	Synonym Qualifier	Topic variable for an Element -- 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	*		Rule	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, June 25, 2004. One Record per Visit per Arm, Tabulation

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, which must be TV.	Req
VISIT	Visit Name	Char		CRF or Derived	Topic	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.	Req
VISITNUM	Visit Sequence Number	Num		CRF or Derived	Identifier	1. Clinical encounter number. 2. Numeric version of VISIT, used for sorting.	Req
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 (Treatment Code)	Char	*	Derived	Record Qualifier	Short 8-character version of ARM, used for programming.	*Exp
ARM	Description of Arm	Char	*		Synonym Qualifier	1. Description of the Arm. 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			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, June 25, 2004. One record per actual Element per Subject, Tabulation

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	**SA	Derived	Identifier	Two-character abbreviation for the domain most relevant to the observation, which must be SA.	Req
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req
ETCD	Subject Element Code	Char	*	Sponsor Defined	Topic	Short 8-character version of ELEMENT, used for programming. 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, June 25, 2004. One record per Subject per actual Visit , Tabulation

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, which must be SV.	Req
USUBJID	Unique Subject Identifier	Char		Sponsor Defined	Identifier	Unique subject identifier within the submission.	Req
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 as a text description of the clinical encounter.	Perm
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
VISITDY	Planned Study Day of Visit	Num		CRF or Derived	Timing	Planned study day of VISIT, used for sorting. (Decimal numbering may be useful for inserting unplanned Visits.)	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 treatment.	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
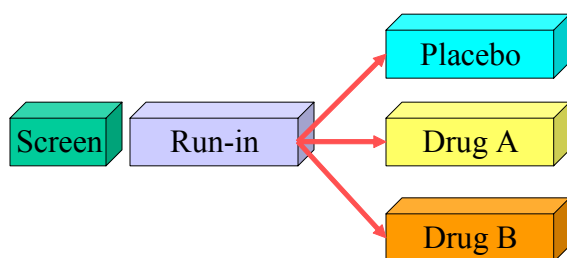
Figure 1

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
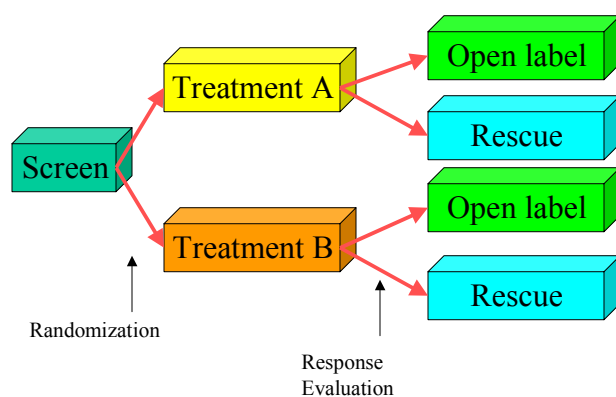
Figure 2

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
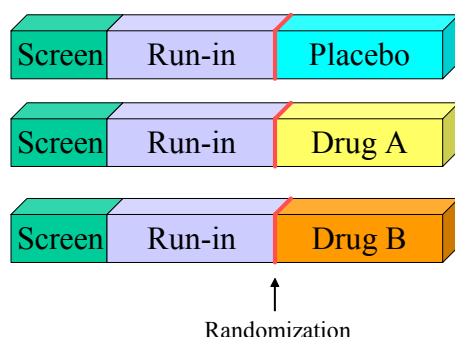
Figure 3

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.

ARM	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
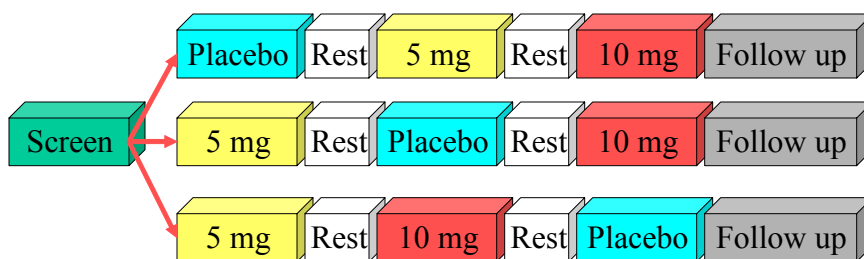
Figure 4

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
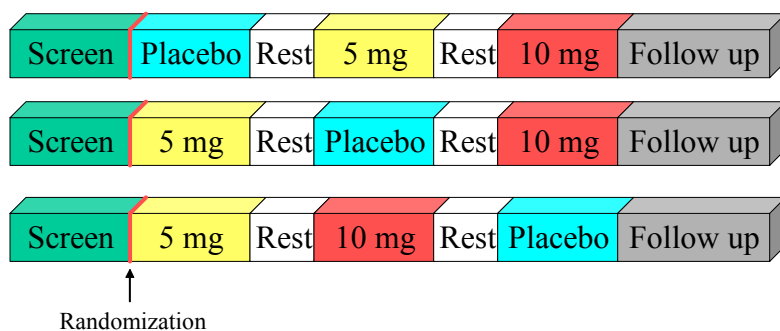
Figure 5

The trial design is represented with three parallel Arms. This transformation is similar to that shown in the Parallel Trial example (previously shown Figures1-3).

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

Repeat until disease progression
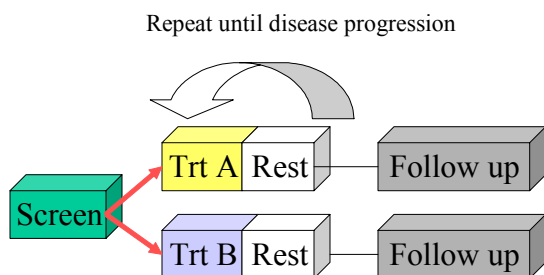
Figure 6

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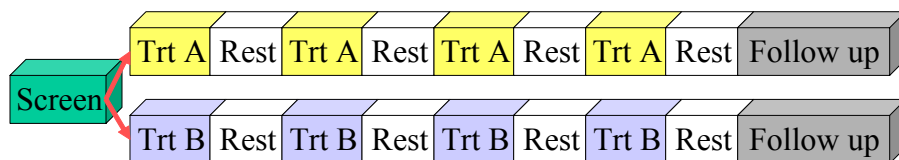
Note: If protocol does not specify maximum number of cycles,
use actual maximum number of cycles that occurred in trial.

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
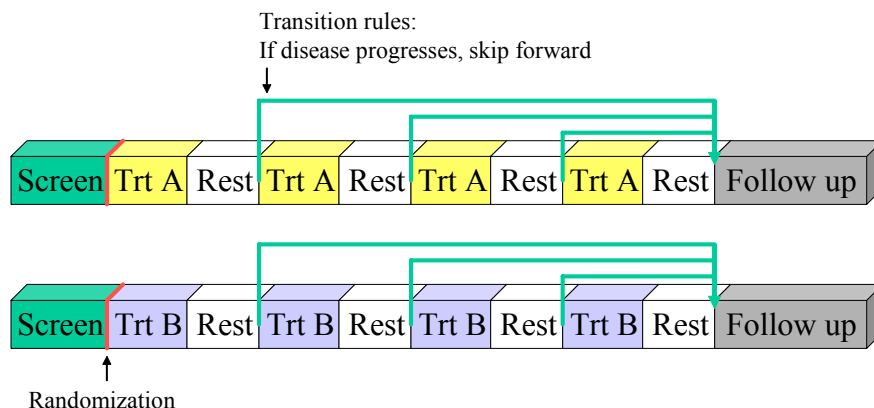
Figure 8

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	Treatment A		
A	3	Rest		If disease progression, go to 10th Element.
A	4	Treatment A		
A	5	Rest		If disease progression, go to 10th Element.
A	6	Treatment A		
A	7	Rest		If disease progression, go to 10th Element.
A	8	Treatment 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 this example, 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
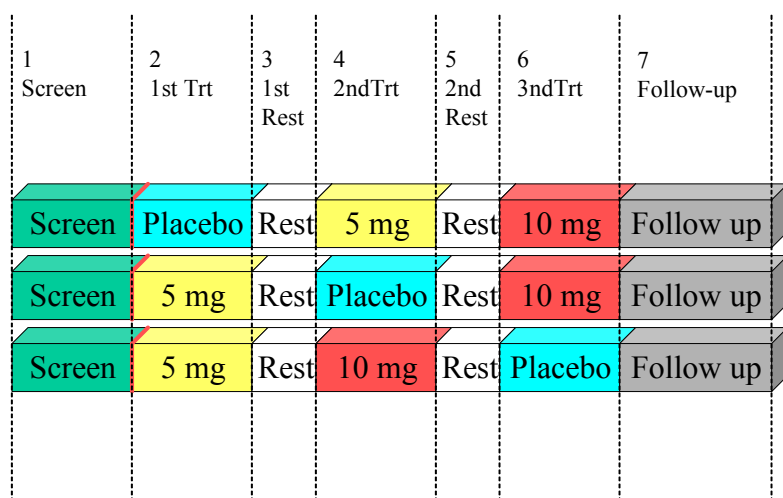
Figure 9

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 and Duration from the TE dataset have been added).

ARM	TAETORD	ELEMENT	TESTRL	TEDUR	TABRANCH
P-5-10	1	Screen	Informed consent	14D	Randomized to Placebo/5 mg/10 mg
P-5-10	2	Placebo	First dose of placebo	14D	
P-5-10	3	Rest	48 hrs after last dose drug	7D	
P-5-10	4	5 mg	First dose of 5 mg drug	14D	
P-5-10	5	Rest	48 hrs after last dose drug	7D	
P-5-10	6	10 mg	First dose of 10 mg drug	14D	
P-5-10	7	Follow-up	48 hrs after last dose drug	21D	
5-P-10	1	Screen	Informed consent	14D	Randomized to 5 mg/Placebo/10 mg
5-P-10	2	5 mg	First dose of 5 mg drug	14D	
5-P-10	3	Rest	48 hrs after last dose drug	7D	
5-P-10	4	Placebo	First dose of placebo	14D	
5-P-10	5	Rest	48 hrs after last dose drug	7D	
5-P-10	6	10 mg	First dose of 10 mg drug	14D	
5-P-10	7	Follow-up	48 hrs after last dose drug	21D	
5-10-P	1	Screen	Informed consent	14D	Randomized to 5 mg/10 mg/Placebo
5-10-P	2	5 mg	First dose of 5 mg drug	14D	
5-10-P	3	Rest	48 hrs after last dose drug	7D	
5-10-P	4	10 mg	First dose of 10 mg drug	14D	
5-10-P	5	Rest	48 hrs after last dose drug	7D	
5-10-P	6	Placebo	First dose of placebo	14D	
5-10-P	7	Follow-up	48 hrs after last dose drug	21D	

Sorting this dataset TAETORD (the value in the Order column), displays the similarity between Elements with the same value of TAETORD. For each value of TAETORD, the End Rules are the same and the Start Rules are either the same or differ only by drug dose. Each value of TAETORD corresponds to an Epoch, and the congruent 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	ELEMENT	TESTRL	TEDUR	TABRANCH
5-P-10	1	Screen	Informed consent	14D	Randomized to 5 mg/Placebo/10 mg
P-5-10	1	Screen	Informed consent	14D	Randomized to Placebo/5 mg/10 mg
5-10-P	1	Screen	Informed consent	14D	Randomized to 5 mg/10 mg/Placebo
5-P-10	2	5 mg	First dose of 5 mg drug	14D	
P-5-10	2	Placebo	First dose of placebo	14D	
5-10-P	2	5 mg	First dose of 5 mg drug	14D	
5-P-10	3	Rest	48 hrs after last dose drug	7D	
P-5-10	3	Rest	48 hrs after last dose drug	7D	
5-10-P	3	Rest	48 hrs after last dose drug	7D	
5-P-10	4	Placebo	First dose of placebo	14D	
P-5-10	4	5 mg	First dose of 5 mg drug	14D	
5-10-P	4	10 mg	First dose of 10 mg drug	14D	
5-P-10	5	Rest	48 hrs after last dose drug	7D	
P-5-10	5	Rest	48 hrs after last dose drug	7D	
5-10-P	5	Rest	48 hrs after last dose drug	7D	
5-P-10	6	10 mg	First dose of 10 mg drug	14D	
P-5-10	6	10 mg	First dose of 10 mg drug	14D	
5-10-P	6	Placebo	First dose of placebo	14D	
5-P-10	7	Follow-up	48 hrs after last dose drug	21D	
P-5-10	7	Follow-up	48 hrs after last dose drug	21D	
5-10-P	7	Follow-up	48 hrs after last dose drug	21D	

Inconsistent numbers and durations of Elements Example: Figure 10

Trial with Inconsistent Numbers and Durations of Elements
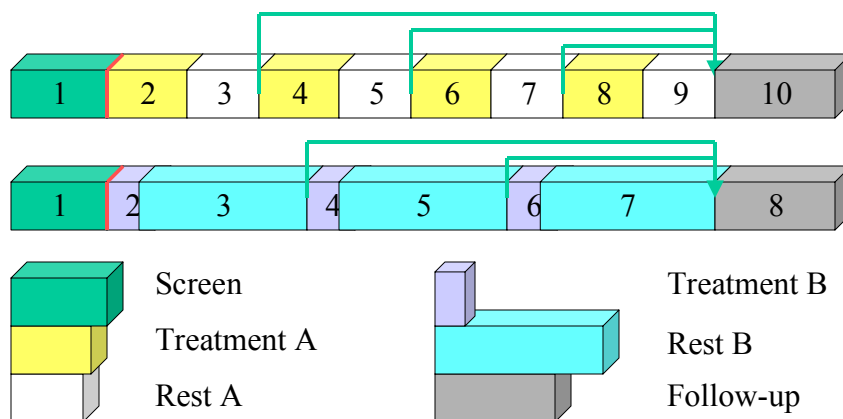
Figure 10

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	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	Drug A	First dose A	9D		
B	2	Drug B	First dose B	7D		
A	3	Rest A	48 hrs after last dose A	12D		If disease progresses, go to 10th Element
B	3	Rest B	48 hrs after last dose B	21D		If disease progresses, go to 8th Element
A	4	Drug A	First dose A	9D		
B	4	Drug B	First dose B	7D		
A	5	Rest A	48 hrs after last dose A	12D		If disease progresses, go to 10th Element
B	5	Rest B	48 hrs after last dose B	21D		If disease progresses, go to 8th Element

ARM	TAETORD	ELEMENT	TESTRL	TEDUR	TABRANCH	TATRANS
A	6	Drug A	First dose A	9D		
B	6	Drug B	First dose B	7D		
A	7	Rest A	48 hrs after last dose A	12D		If disease progresses, go to 10th Element
B	7	Rest B	48 hrs after last dose B	21D		
A	8	Drug 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
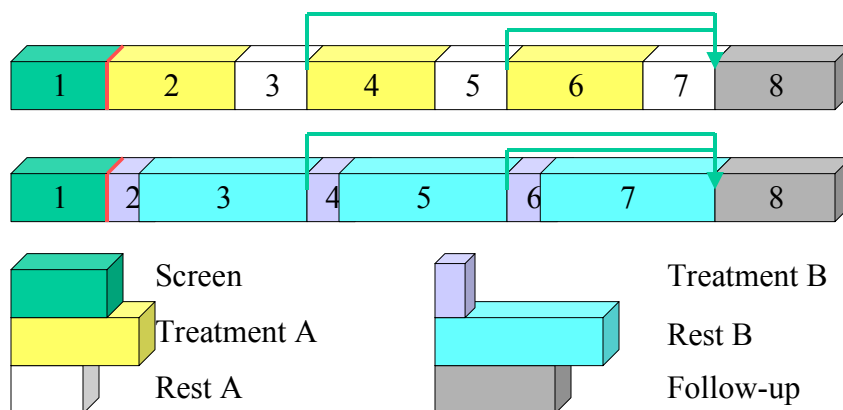
Figure 11

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

Trial with Dissimilar Treatments
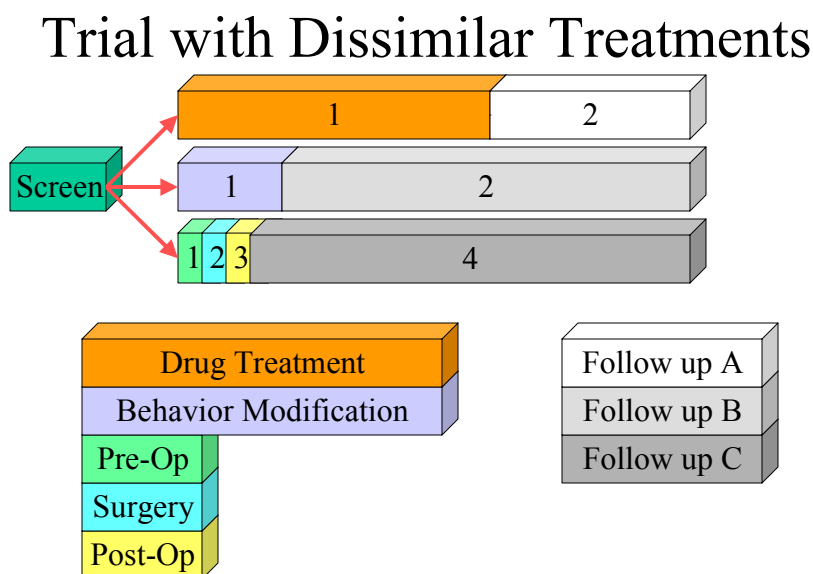
Figure 12

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, 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 be 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	Optional. 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.

Arms and Elements, Parallel Trial

9 records in
Trial Arms table

5 records in Trial Elements table
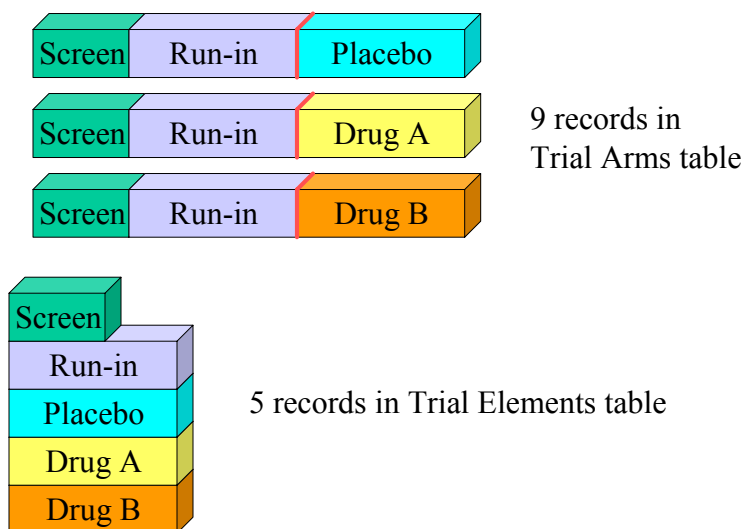
Figure 13

Arms and Elements, Crossover Trial

Screen	Placebo	Rest	5 mg	Rest	10 mg	Follow up
Screen	5 mg	Rest	Placebo	Rest	10 mg	Follow up
Screen	5 mg	Rest	10 mg	Rest	Placebo	Follow up

21 records in Trial Arms table

Screen
Placebo
5 mg
10 mg
Rest
Follow up

6 records in Trial Elements table
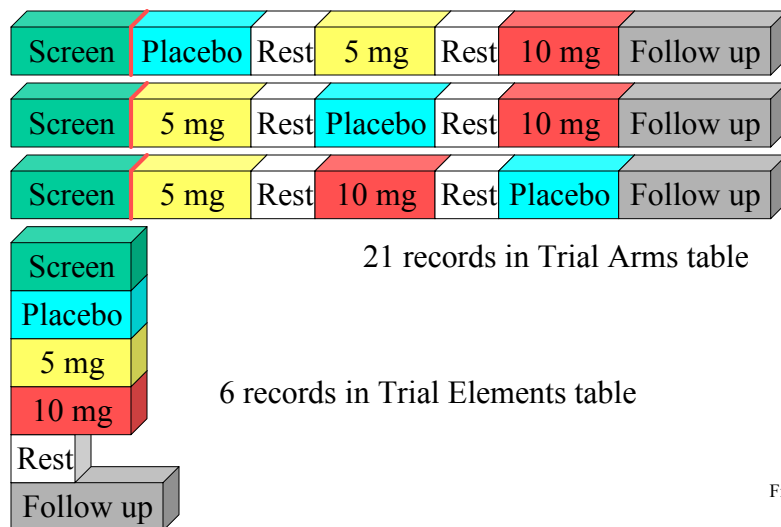
Figure 14

Arms and Elements, Oncology Trial

| Screen | Trt A | Rest | Trt A | Rest | Trt A | Rest | Trt A | Rest | Follow up |
| Screen | Trt B | Rest | Trt B | Rest | Trt B | Rest | Trt B | Rest | Follow up |

21 records in Trial Arms table

Screen
Trt A
Trt B
Rest
Follow up

5 records in Trial Elements table
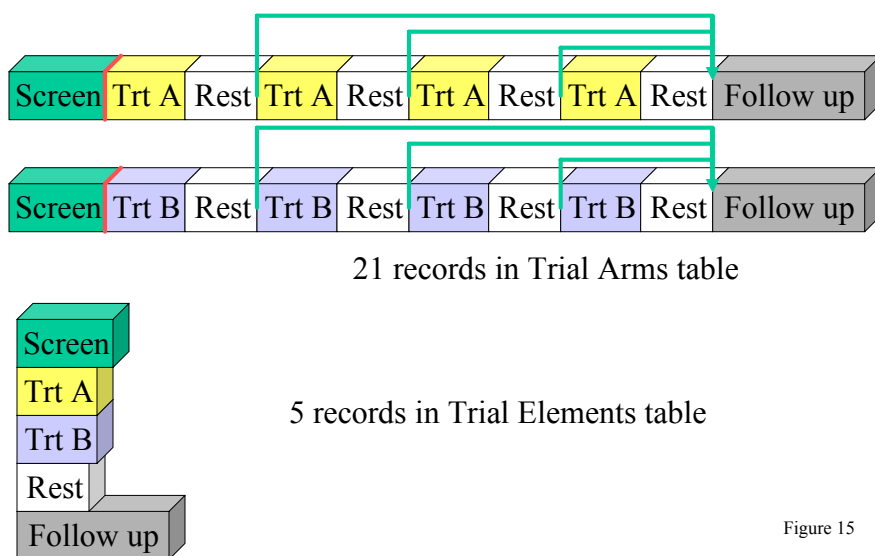
Figure 15

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

Transition rules:
If disease progresses, skip forward

Screen | Trt A | Trt A | Trt A | Trt A | Follow up

Screen | Trt B | Trt B | Trt B | Trt B | Follow up

Randomization
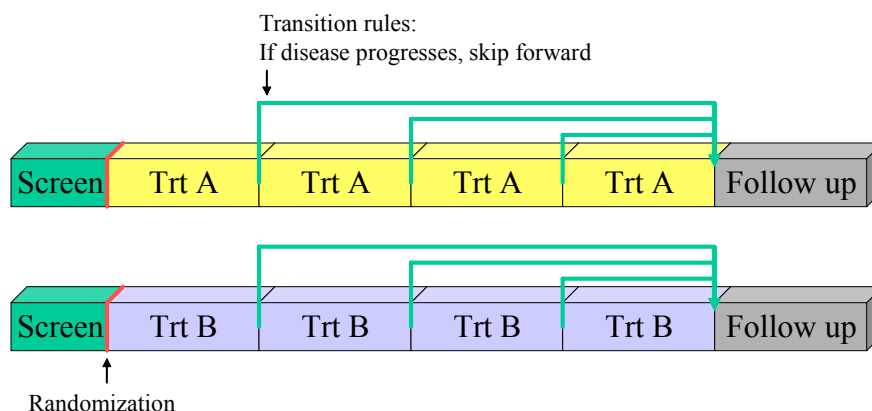
Figure 16

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 <u>does</u> end, but the conditions under which it <u>should</u> 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	7D
Run-In	Eligibility confirmed	2 weeks after start of element	14D
Placebo	First dose of placebo	2 weeks after start of element	14D
Drug A	First dose of Drug A	2 weeks after start of element	14D
Drug B	First dose of Drug B	2 weeks after start of element	14D

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	14D
Placebo	First dose of placebo	2 weeks after start of element	14D
5 mg	First dose of 5 mg drug	2 weeks after start of element	14D
10 mg	First dose of 10 mg drug	2 weeks after start of element	14D
Rest	48 hrs after last dose drug	1 week after start of element	7D
Follow-up	48 hrs after last dose drug	3 weeks after start of element	21D

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	14D
Trt A	First dose drug	5 days after start of element	5D
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	28D

A subject in the 3rd Element 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**
	1st	2nd	3rd	4th	5th	6th	7th
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 (1st, 2nd, 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 affects 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 in are discussed further in

Trial Visits Example: Figure 17

Trial Visits
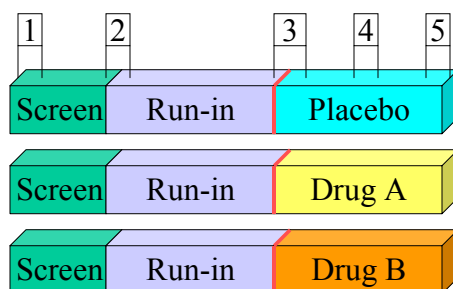
Figure 17

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

VISITNUM	TVSTRL	TVENRL
1	Start of 1st Element	1 hour after start of visit
2	30 minutes before end of 1st Element	30 minutes after start of 2nd Element
3	30 minute before end of 2nd Element	1 hour after start of 3rd Element
4	1 week after start of 3rd Element	1 hour after start of visit
5	2 weeks after start of 4th 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. If visit schedule does depend on Arm, names the 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, then 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 as the end date/time for the last subject 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	Optional 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.

Tools to assign dates to a subject Element will have to assign date/times equal to an actual end/start date/time to either the ending Element or the starting Element. It will be particularly important to specify this convention if date/times are collected only to the nearest day.

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	Optional variable. Blank except for unplanned Visits. Describes the unplanned Visit.

VISIT, VISITNUM, VISITDY – variables that name and order the visits

SVSTDTC and SVENDTC – date/times of start and end of this visit for this subject.

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, June 4, 2004. One record per I/E criterion, Tabulation

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. Must be TI.	Req
IETESTCD	Exception Criteria Test or Exam Short Name	Char	*	Sponsor Defined	Topic		Req
IETEST	Exception Criteria Test or Exam	Char	*	Sponsor Defined / Protocol	Qualifier	Full text of the inclusion or exclusion criterion	Req
IECAT	Category for Test or Examination Criteria	Char	*INCLUSION, EXCLUSION	Sponsor Defined	Topic		Req
TIRL	Inclusion/Exclusion Criterion Rule	Char		Sponsor Defined	Rule	Optional rule that expresses the criterion in computer-executable form.	Perm

7.10 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 Relationships among Datasets or Records

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 identifies five distinct types of relationships, which are described in more detail in this Section:

- 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 dependent 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) in a different general class, such as lesion measurements associated with the identification (and classification) of a lesion.

- Section 8.4 describes a dependent relationship between data that cannot be represented by a standard variable 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.

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 identifier, such as an identifier collected on the CRF as an explicit identifier, identified in the protocol, or provided in an electronic data transfer. Some examples are lab-specimen identifiers, ECG identifiers (sometimes used to reference waveform ECGs), and lesion names/numbers (identified for relating the periodic measurements to the specific lesion identification and classification). --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 the next four 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 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 does not include all variables that would be used in CM; only a subset of the variables is shown to illustrate the use of --GRPID. 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 DOMAINS

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. Relationships can be defined for single records (by using --SEQ in IDVAR and the appropriate --SEQ value in IDVARVAL) or groups of records (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).

The relationship is defined by including a RELREC record that identifies the key(s) for each of the records to be related, and by assigning the same unique character value to the RELID variable for each of the related records. The value of RELID can be any constant value chosen by the sponsor.

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 dataset must be used to describe this 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 general classes, 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 (SUPPQUAL) dataset, which is described in Section 8.4.

RELREC records identify relationships between records in SDS domains using the keys STUDYID, DOMAIN, and USUBJID variables, which are populated with the values for STUDYID, DOMAIN, and USUBJID from the SDS domain record(s), along with the IDVAR and IDVARVAL variables.

8.2.1 RELREC Dataset

Variable	Variable Label	Type	Description
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	Identifier Variable	Char**	Value of the identifier variable in the general class dataset that identifies the related record(s). Examples include --SEQ and --GRPID.
IDVARVAL	Identifier Variable Value	Char	Value of identifier 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 or ** external controled 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 DEPENDENT RECORDS IN DIFFERENT DATASETS

The Related Records (RELREC) dataset can also be used to identify relationships between datasets that have a common topicality (and thus may be considered to belong to the same logical domain), such as microbiology specimen information with susceptibility testing of the identified organisms. 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 share a common topicality. 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 parent record(s) key variable identifiers that are necessary to make the association.

8.3.1 RELREC Split Datasets Example

This example shows how to use the RELREC dataset to represent microbiology information that is submitted as two datasets:

1. Specimen information organism identification and results from tests done on the entire specimen are represented in the M1 dataset, considered to be Findings
2. Susceptibility testing of the identified organisms is represented in the M2 dataset, also considered to be Findings.

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
EFC1234	M1		M1SPID		ONE	A
EFC1234	M2		M2SPID		MANY	A

In both of the datasets, the --SPID variable is used to identify the specimen, and could be used to merge/join the two datasets. The values for the --SPID variable in the two datasets are sponsor defined.

Since IDVAR in this case (when using the RELREC dataset to identify dataset-to-dataset relationships) identifies the merge/join keys that can be used to merge/join all records between the datasets, the root values for IDVAR must be the same for both records with the same RELID and will likely be either --REFID or --SPID. --SEQ and --GRPID cannot be used in this case 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) dataset 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 the SUPPQUAL dataset. The SUPPQUAL dataset 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 SUPPQUAL 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 SUPPQUAL 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.

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 SUPPQUAL 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

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	Identifier Variable	Char*	Identifier variable in the dataset that identifies the related record(s). Examples: --SEQ, --GRPID.	

Variable	Variable Label	Type	CDISC Notes	References
IDVARVAL	Identifier Variable Value	Char	Value of identifier variable of the parent record(s).	
QNAM	Qualifier Variable Name	Char*	The short name of the Qualifier variable. This variable should contain a text value less than or equal to 8 characters in length, so the value could be used as a column name in a domain view with data from the parent domain(s). This will often be the column name in the sponsor's original dataset. QNAM values can only include alphanumeric characters and the underscore (_) and cannot start with a number.	SDSIG 4.1.5.2
QLABEL	Qualifier Variable Label	Char	This is the long name or label associated with QNAM. This should be less than or equal to 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. Controlled terminology: 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. Controlled terminology will consist of values such as ADJUDICATION COMMITTEE, STATISTICIAN, DATABASE ADMINISTRATOR, CLINICAL COORDINATOR, PRIMARY INVESTIGATOR, etc.	

*Indicates variable may be subject to sponsor-controlled or ** external controlled terminoloty.*

A record(s) in the SUPPQUAL dataset merge/join back to their 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 when the Intent To Treat population flag (ITT) and Safety population flag (SAFETY) are applied at the subject level only, will have both IDVAR and IDVARVAL set to Null, as the only key needed to identify the relationship/association to all subject records (for that subject) in all domains is USUBJID.

The 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, the value in EVAL should reflect the role of the person or institution assigning the value (e.g., PRIMARY INVESTIGATOR or ADJUDICATION COMMITTEE).

8.4.2 SUPPQUAL Examples

Examples of how to use the SUPPQUAL dataset are listed in the table below:

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVAR VAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
1996001	DM	99-401			ITT	Intent to Treat	Y	DERIVED	SPONSOR
1996001	DM	99-401			PPE	Per Protocol Evaluable	N	DERIVED	SPONSOR
1996001	AE	99-401	AESEQ	1	AESOTHC	Desc of Other Med Impt Serious Event	Spontaneous Abortion	CRF	
1996001	AE	99-401	AESEQ	2	AETRTEM	Treatment Emergent	N	DERIVED	SPONSOR

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

- Subject-level objective data that fits in Subject Characteristics (SC)
- ECG Interpretation -- this should be added as an additional test called EGINTP in EGTESTCD, and given the same EGGRPID or EGREFID 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 text comments collected on separate CRFs or as part of CRFs with their own topicality, such as adverse events, concomitant medications, or comments supplied by central reviewers. Comments may be related to a Subject, a domain for a Subject, or to specific parent records in any SDS domain. The Comments domain is structured similarly to the SUPPQUAL 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 SDS domains. STUDYID, USUBJID, and DOMAIN 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 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 SDS domain records for a subject would populate the RDOMAIN, IDVAR, and IDVARVAL variables with values from the parent record(s).

Again, the use of --GRPID can be used to identify comments with relationships to multiple SDS 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 each of the adverse events (AE), exposure (EX), and vital signs (VS) pages, one on the PE page, and another on a separate comments 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	PAGES 320-355		Some delays in taking medications because of vacation	PRINCIPAL INVESTIGATOR
1234	CO	VS	1234-2004	3	VSSEQ	5	PAGE 120		Value confirmed in range, but questionable based on other related measurements.	ADJUDICATIO N COMMITTEE
1234	CO	PE	1234-2001	4			Page 40	2003-11-08	Subject was very uncooperative during physical exam.	PRINCIPAL INVESTIGATOR
1234	CO		1234-2004	5			PAGE 800	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 general 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 dataset.

When determining how to model a new type of data, it is best to identify the topicality first and classify its primary observation class as Events, Interventions, or Findings (see section 2.4). To do this, begin by evaluating the 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.

Once the 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. For information that is logically part of one domain (having a single topicality), but that cannot be entirely represented in the same general observation class, it will be necessary to do one of two things:

1. Submit that data in the SUPPQUAL (Supplemental Qualifiers) special-purpose dataset (see Section 8.4.1) 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. Section 8.4 describes how information should be stored in

the Supplemental Qualifiers (SUPPQUAL) dataset when it does not have either Identifier, Topic, or Timing variables that are different from the data which do fit into an observation class. In other words, the information consists only of what would be classified as additional Qualifiers, and would not be considered an entire 'record' of information 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 or in SDS V2 submission datasets.

2. Submit that data as a separate dataset if the information that does not fit consists of 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, its own Timing variables and additional Qualifier(s) that need to stand separately.

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.

    Additional findings for data already determined to fit into the Findings observation class would normally be submitted as additional rows within the same Domain dataset, and not as a second dataset.

2. The topicality of the information determines the Parent (or master) dataset general observation class (Events or Interventions). A Parent (master) dataset in the Events or Interventions general observation class is required, even if the records have limited information. The variable --OCCUR can be used to indicate Y/N if the only information known about an Event or Intervention is whether or not it occurred (e.g., a responses of Yes or No to the question, "did the subject experience nausea [an event]?" or "did the subject take aspirin [an intervention]?").

3. Both datasets should include the identifier variable (--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 the Domain variable, and the variable prefix. The dataset names for the two datasets should be the two-character Domain codes chosen for the datasets (MH and MX for example). Since it is not self-evident that a Domain is captured in multiple datasets using this combination of dataset names (nobody would automatically know that MH and MX 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 V3.1 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, the 4th subject (USUBJID=4), who is recorded as 'ASIAN' for RACE could have the value 'INDIAN' in SC where SCTESTCD equals 'RACEOTH.'

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

June 25, 2004 Final

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 (USUBID=1), each instance is recorded separately, and frequency (CMDOSFRQ) is ONCE. For the second subject (USUBID=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 (USUBID=3) records are collapsed 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.

STUDYID	DOMAIN	USUBJID	CMSEQ	CMTRT	CMDOSFRQ	CMSTDTC	CMENDTC
ABC123	CM	1	1	Aspirin	ONCE	2004-01-01	2004-01-01
ABC123	CM	1	2	Aspirin	ONCE	2004-01-02	2004-01-02
ABC123	CM	1	3	Aspirin	ONCE	2004-01-03	2004-01-03
ABC123	CM	1	4	Aspirin	ONCE	2004-01-07	2004-01-07
ABC123	CM	1	5	Aspirin	ONCE	2004-01-07	2004-01-07
ABC123	CM	1	6	Aspirin	ONCE	2004-01-09	2004-01-09
ABC123	CM	2	1	Aspirin	QD	2004-01-01	2004-01-03
ABC123	CM	2	2	Aspirin	BID	2004-01-07	2004-01-07
ABC123	CM	2	3	Aspirin	QD	2004-01-09	2004-01-09
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	EXDOSTOT	EXROUTE	EXSTDTC	EXENDTC	EXSTDY	EXENDY
Row 1	QD	40	ORAL	2002-01-10	2002-03-08	1	58
Row 2	BID	0	ORAL	2002-01-10	2002-03-08	1	58
Row 3	QD	20	ORAL	2002-01-10	2002-03-07	1	57
Row 4	BID	0	ORAL	2002-01-10	2002-03-07	1	57
Row 5	QD	0	ORAL	2002-01-11	2002-02-01	1	22
Row 6	BID	300	ORAL	2002-01-11	2002-02-01	1	22
Row 7	QD	0	ORAL	2002-02-04	2002-03-06	25	55
Row 8	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
Row 1	56789	EX	56789001	1	1	DRUG A	20	MG	CAPSULE
Row 2	56789	EX	56789001	2	1	DRUG A	20	MG	CAPSULE
Row 3	56789	EX	56789001	3	1	DRUG A	20	MG	CAPSULE
Row 4	56789	EX	56789001	4	2	DRUG B	30	MG	TABLET, COATED
Row 5	56789	EX	56789001	5	2	DRUG B	30	MG	TABLET, COATED
Row 6	56789	EX	56789001	6	2	DRUG B	30	MG	TABLET, COATED
Row 7	56789	EX	56789003	1	1	DRUG B	30	MG	TABLET, COATED
Row 8	56789	EX	56789003	2	1	DRUG B	30	MG	TABLET, COATED
Row 9	56789	EX	56789003	3	1	DRUG B	30	MG	TABLET, COATED
Row 10	56789	EX	56789003	4	2	DRUG A	20	MG	CAPSULE
Row 11	56789	EX	56789003	5	2	DRUG A	20	MG	CAPSULE
Row 12	56789	EX	56789003	6	2	DRUG A	20	MG	CAPSULE

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

9.2.3 SU Examples:

Examples and Usage Notes for Substance Use (SU) Domain Model

In SDS V3.0 the Substance Use domain was modeled using the Findings general observation class. As a result of the FDA/SDS V3.0 pilot, it was determined that the Interventions general observation class was a more appropriate structure for substance use data. The examples presented in these Usage Notes reflect the change to an Intervention-based structure for SU.

The use of the Interventions general observation class to model substance use data meets the needs of the simple medical-history-based information as well as the complex, detailed dosing information requirements for analysis.

In many clinical trials, detailed substance use information as represented by this domain may not be required. In these instances this information is usually captured for the purpose of evaluating a subject's medical history, or to assess ongoing substance use without requiring additional detailed dosing information for a specific substance. The data collected in this case could be used in analyses to support secondary hypotheses defined in the protocol, or to provide supportive evidence for sub-group analyses of the study therapy. An example of a Case Report Form (CRF) that captures substance use data in this limited scope is presented in Figure 9.2.3.1.

Table 1 presents data that could be captured using the simple smoking history CRF presented in Figure 9.2.3.1. Table 2 presents the data as structured using the SDS V3.1 SU model.

Other clinical trials may be conducted where detailed dosing information regarding substance use is required (e.g., smoking cessation trials). In such cases, detailed information (e.g., visit, dose, total daily dose, start and stop dates) about the substance consumed is collected. The data collected in this case could be used in primary efficacy analyses to support therapy effectiveness claims. Figure 9.2.3.2 presents a CRF that captures data of this nature.

Table 3 presents data that could be captured using the detailed smoking history CRF presented in Figure 9.2.3.2, and Table 4 presents the data as structured using the SDS V3.1 SU model.

Not all variables specified in the CDISC SU domain are present on either of the example CRFs. Since they are not used, the variables are not present on the SU dataset. However, all required variables are present and populated.

Figure 9.2.3.1: Sample Case Report Form for Smoking History - Simple

| Visit x | SMOKING | Smoking History |

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

SMOKCAT

☐ 1   Non-smoker (never smoked)

☐ 2   Passive smoker (a non-smoker who is frequently exposed to second-hand smoke)

☐ 3   Ex smoker

☐ 4   Current smoker

Table 1: Data Captured Using the Simple Smoking History CRF

	STUDYID	DOMAIN	USUBJID	VISIT	SEQ	SMOKCAT
Row 1	1234	SU	1234001	1	1	NON
Row 2	1234	SU	1234002	1	1	PASSIVE
Row 3	1234	SU	1234003	1	1	EX
Row 4	1234	SU	1234004	1	1	PRESENT

Table 2: Data from Table 1 Modeled in SDS V3.1 SU Domain

	STUDYID	DOMAIN	USUBJID	SUSEQ	SUTRT	SUDECOD	SUCAT	SUOCCUR	SUDOSTXT	VISITNUM
Row 1	1234	SU	1234001	1	TOBACCO	NICOTINE	TOBACCO	N		1
Row 2	1234	SU	1234002	1	TOBACCO	NICOTINE	TOBACCO	Y	PASSIVE	1
Row 3	1234	SU	1234003	1	TOBACCO	NICOTINE	TOBACCO	Y	EX	1
Row 4	1234	SU	1234004	1	TOBACCO	NICOTINE	TOBACCO	Y	PRESENT	1

Figure 9.2.3.2 - Sample Case Report Form for Smoking History – Detailed

Visit x	SMOKING	Smoking History

Please identify **one** of the following smoking categories for the subject: SMOKCAT   ☐₁ Not assessed   NOTDONE

☐₁ Non-smoker (never smoked)

Start date   SMOKSDT   Stop date   SMOKEDT

☐₂ Ex smoker

d d m m y y y y   d d m m y y y y

☐₃ Current smoker

d d m m y y y y

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

SMOKTYPE

☐₁ Cigars   per day. Round to next highest whole number. Specify average daily amount smoked during the time the subject smoked.

☐₂ Pipe   bowls per day. Round to next highest whole number. Specify average daily amount smoked during the time the subject smoked.

☐₃ Cigarettes   per day. Round to next highest whole number. Specify average daily amount smoked during the time the subject smoked.

TOBACONO

Table 3: Data Captured Using the Detailed Smoking History CRF

	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	01-01-2000	04-02-2003	
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	1	2	CURRENT	CIGARETTES	20	1965		
Row 8	1234	1234010	1	1						1

Table 4: Data from Example 3 Modeled in SDS V3.1 SU Domain

	STUDYID	DOMAIN	USUBJID	SUSEQ	SUTRT	SUDECOD	SUCAT	SUSTAT	SUDOSTXT	SUDOSFRQ	SUDOSTOT
Row 1	1234	SU	1234005	1	TOBACCO	NICOTINE	TOBACCO		NON		
Row 2	1234	SU	1234006	1	CIGARS	NICOTINE	TOBACCO		EX	/DAY	1
Row 3	1234	SU	1234007	1	PIPE	NICOTINE	TOBACCO		EX	/DAY	1
Row 4	1234	SU	1234007	2	CIGARETTES	NICOTINE	TOBACCO		EX	/DAY	10
Row 5	1234	SU	1234008	1	CIGARETTES	NICOTINE	TOBACCO		CURRENT	/DAY	20
Row 6	1234	SU	1234009	1	CIGARS	NICOTINE	TOBACCO		EX	/DAY	1
Row 7	1234	SU	1234009	2	CIGARETTES	NICOTINE	TOBACCO		CURRENT	/DAY	20
Row 8	1234	SU	1234010	1	ALL			NOT DONE			

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

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	AEBODSYS
Row 1	ABC123	AE	123101	1	POUNDING HEADACHE	2003-10-12	2003-10-12	HEADACHE	HEADACHE	NERVOUS SYSTEM DISORDERS
Row 2	ABC123	AE	123101	2	BACK PAIN FOR 6 HOURS	2003-10-13T13:05	2003-10-13T19:00	BACK PAIN	BACK PAIN	MUSCULOSKELETAL AND CONNECTIVE TISSUE DISORDERS
Row 3	ABC123	AE	123101	3	PULMONARY EMBOLISM	2003-10-21			PULMONARY EMBOLISM	VASCULAR DISORDERS

	AESEV	AESER	AEACN	AEREL	AEOUT
Row 1	SEVERE	N	NOT APPLICABLE	DEFINITELY NOT RELATED	RECOVERED/RESOLVED
Row 2	MODERATE	N	DOSE REDUCED	PROBABLY RELATED	RECOVERED/RESOLVED
Row 3	MODERATE	Y	DOSE REDUCED	PROBABLY NOT RELATED	RECOVERING/RESOLVING

	AESCONG	AESDISAB	AESDTH	AESHOSP	AESLIFE	AESMIE	AESTDY	AEENDY	AEENRF
Row 1							-1	-1	
Row 2							1	1	
Row 3				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.

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

	STUDYID	DOMAIN	USUBJID	DSSEQ	DSTERM	DSDECOD	DSCAT	VISITNUM	EPOCH	DSDTC	DSSTDTC
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	45	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

9.3.3 MH Example:

In this example, a Medical History CRF collects descriptions of conditions and events by body system (e.g., Endocrine, Metabolic). These events include medical history, a primary diagnosis, and risk factors. The events for the Medical History and Risk Factors (MHCAT = 'MEDICAL HISTORY' or MHCAT = 'RISK FACTORS') are pre-specified on the CRF, and there is a prompt for 'OTHER MEDICAL HISTORY' which does not have pre-specified terms. The sponsor codes findings using a standard dictionary.

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 would be placed in the Comments domain (CO).

   c. If a subject had no conditions or events reported, no records are included for that subject in the MH dataset.

   d. If a body system is not done at all then this is indicated by a value of MHSTAT = 'NOT DONE'

   e. If terms are pre-specified on the CRF then MHOCCUR is populated. In the case where the CRF has allowed for free text, then MHOCCUR is NULL and MHTERM is populated. This is represented in the example for the record MHSEQ = 5 and MHCAT = 'PRIMARY DIAGNOSIS'.

Sample data that illustrates the above:

	STUDYID	DOMAIN	USUBJID	MHSEQ	MHTERM	MHDECOD	MHCAT	MHOCCUR	MHSTAT	MHREASND	MHBODSYS
Row 1	ABC123	MH	123101	1	ALL		MEDICAL HISTORY		NOT DONE	Subject left office before examination was completed	GASTOINTESTINAL DISORDERS
Row 2	ABC123	MH	123101	2	ASTHMA	ASTHMA	MEDICAL HISTORY	Y			RESPIRATORY SYSTEM DISORDERS
Row 3	ABC123	MH	123101	3	FREQUENT HEADACHES	HEADACHE	MEDICAL HISTORY	N			CENTRAL AND PERIPHERAL NERVOUS SYSTEM DISORDERS

CDISC. © 2004. All rights reserved
Final

	STUDYID	DOMAIN	USUBJID	MHSEQ	MHTERM	MHDECOD	MHCAT	MHOCCUR	MHSTAT	MHREASND	MHBODSYS
Row 4	ABC123	MH	123101	4	BROKEN LEG	BONE FRACTURE	OTHER MEDICAL HISTORY				MUSCULOSKELTAL SYSTEM DISORDERS
Row 5	ABC123	MH	123101	5	STROKE		PRIMARY DIAGNOSIS				CARDIOVASCULAR
Row 6	ABC123	MH	123101	6	HYPERTENSION		RISK FACTORS	Y			CARDIOVASCULAR
Row 7	ABC123	MH	123101	7	TIA		RISK FACTOR	N			CARDIOVASCULAR

9.4 FINDINGS EXAMPLES

9.4.1 EG Examples

ECG Example 1:

This is an example of an ECG dataset from Company A 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.

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

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

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

4. **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.

5. **Row 12** - Shows how interpretations are represented.

	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

	STUDYID	DOMAIN	USUBJID	EGSEQ	EGGRPID	EGREFID	EGTESTCD	EGTEST	EGXFN	EGNAM	EGPOS
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	ECG Interpretation			SUPINE

	EGORRES	EGORRESU	EGSTRESC	EGSTRESN	EGSTRESU	EGEVAL	VISIT	VISITNUM	EGDTC	EGDY
Row 1	62	BPM	62	62	BPM	ECGCRO	SCREENING 1	1	2003-04-15T11:58	-36
Row 2	0.15	SEC	150	150	MS	ECGCRO	SCREENING 1	1	2003-04-15T11:58	-36
Row 3	0.103	SEC	103	103	MS	ECGCRO	SCREENING 1	1	2003-04-15T11:58	-36
Row 4	0.406	SEC	406	406	MS	ECGCRO	SCREENING 1	1	2003-04-15T11:58	-36
Row 5	0.406	SEC	406	406	MS	ECGCRO	SCREENING 1	1	2003-04-15T11:58	-36
Row 6	0.406	SEC	406	406	MS	ECGCRO	SCREENING 1	1	2003-04-15T11:58	-36
Row 7	ATRIAL FIBRILLATION		ATRIAL FIBRILLATION			ECGCRO	SCREENING 1	1	2003-04-15T11:58	-36
Row 8	ATRIAL FLUTTER		ATRIAL FLUTTER			ECGCRO	SCREENING 1	1	2003-04-15T11:58	-36
Row 9	QT PROLONGATION		QT PROLONGATION			ECGCRO	SCREENING 1	1	2003-04-15T11:58	-36
Row 10	LEFT VENTRICULAR HYPERTROPHY		LEFT VENTRICULAR HYPERTROPHY			ECGCRO	SCREENING 1	1	2003-04-15T11:58	-36

	EGORRES	EGORRESU	EGSTRESC	EGSTRESN	EGSTRESU	EGEVAL	VISIT	VISITNUM	EGDTC	EGDY
Row 11	INCORRECT ELECTRODE PLACEMENT		INCORRECT ELECTRODE PLACEMENT			ECGCRO	SCREENING 1	1	2003-04-15T11:58	-36
Row 12	ABNORMAL, CLINICALLY INSIGNIFICANT		ABNORMAL, CLINICALLY INSIGNIFICANT				SCREENING 1	1	2003-04-15T11:58	-36

ECG Example 2:

Example 2 shows results from Company B 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 principle investigator and a cardiologist. In this example the EGGRPID is the same number and the EGSEQ increments by one.

1. **Row 2** - The only one selected as Baseline.

2. **Row 3** - Shows a date/time in ISO 8601 representation where 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	INTP	ECG Interpretation	SUPINE	NORMAL	NORMAL
Row 2	ABC	EG	ABC-99-CA-456	2	2	INTP	ECG Interpretation	SUPINE	ABNORMAL	ABNORMAL
Row 3	ABC	EG	ABC-99-CA-456	3	3	INTP	ECG Interpretation	SUPINE	ABNORMAL	ABNORMAL
Row 4	ABC	EG	ABC-99-CA-456	4	4	INTP	ECG Interpretation	SUPINE	ABNORMAL	ABNORMAL
Row 5	ABC	EG	ABC-99-CA-456	5	4	INTP	ECG Interpretation	SUPINE	ABNORMAL	ABNORMAL

	EGSTRESN	EGBLFL	EGEVAL	VISIT	VISITNUM	VISITDY	EGDTC	EGDY
Row 1			PI	SCREEN I	1	-2	2003-11-26	-2
Row 2		Y	PI	SCREEN II	2	-1	2003-11-27	-1

	EGSTRESN	EGBLFL	EGEVAL	VISIT	VISITNUM	VISITDY	EGDTC	EGDY
Row 3			PI	DAY 10	3	10	2003-12-07T09:02	10
Row 4			PI	DAY 18	4	15	2003-12-12	15
Row 5			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	Evidence of clinically significant organic disease that would prevent the subject from completing?	EXCLUSION	17	Y	Y
Row 2	XYZ	IE	XYZ-0023	1	INCL03	Clinically acceptable screening bilateral mammogram as determined by the local radiologist?	INCLUSION	3	N	N
Row 3	XYZ	IE	XYZ-0047	1	INCL03	Clinically acceptable screening bilateral mammogram as determined by the local radiologist?	INCLUSION	3	N	N
Row 4	XYZ	EG	XYZ-0096	1	INCL03	Clinically acceptable screening bilateral mammogram as determined by the local radiologist?	INCLUSION	3	N	N

	VISIT	VISITNUM	VISITDY	IEDTC	IEDY
Row 1	WEEK-8	1	-63	1999-01-10	-58
Row 2	WEEK-8	1	-63	1999-01-06	-62
Row 3	WEEK-8	1	-63	1999-01-12	-56

	VISIT	VISITNUM	VISITDY	IEDTC	IEDY
Row 4	WEEK-8	1	-63	1999-01-13	-55

9.4.3 LB Examples

The example 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 LBORNRHI or LBSTNRHI; however, a sponsor may decide, for example, to enter '0' into LBORNRLO and '199' in LBORNRHI. The sponsor must define the appropriate value for each of the normal range variables.

1. **Row 1**   - Shows a value collected in one unit, but converted to a standard unit.

2. **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.

3. **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, such as 'COAGULATION'.

4. **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.

5.

	STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBSCAT	LBORRES	LBORRESU	LBORNRHI	LBORNRLO
Row 1	ABC	LB	ABC-001-001	1	ALB	Albumin	CHEMISTRY		34	g/L	50	35
Row 2	ABC	LB	ABC-001-001	2	ALKP	Alkaline Phosphatase	CHEMISTRY		398	IU/L	160	40
Row 3	ABC	LB	ABC-001-001	3	ALKP	Alkaline Phosphatase	CHEMISTRY		350	IU/L	160	40
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	11	4
Row 6	ABC	LB	ABC-001-001	6	LYMPH	Lymphocytes	HEMATOLOGY	DIFFERENTIAL	6.7	%	40	25
Row 7	ABC	LB	ABC-001-001	7	NEUT	Neutrophils	HEMATOLOGY	DIFFERENTIAL	5.1	10*9/L	8	2
Row 8	ABC	LB	ABC-001-001	8	PH	pH	URINALYSIS		7.5		9.0	5.0
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	199*	0*
Row 11	ABC	LB	ABC-001-001	11	WBC	WBC Count	HEMATOLOGY		5.9	10*9/L	11	4

	LBSTRESC	LBSTRESN	LBSTRESU	LBSTNRHI	LBSTNRLO	LBSTAT	LBREAS ND	LBBLFL	LBFAST	LBDRVFL	VISIT	VISITNUM	LBDTC
Row 1	3.4	3.4	g/dL	5	3.5			Y	Y		BASELINE	1	1999-06-19
Row 2	398	398	U/L	160	40				Y		BASELINE	1	1999-06-19
Row 3	350	350	U/L	160	40				Y		BASELINE	1	1999-06-20
Row 4	374	374	U/L	160	40			Y	Y	Y	BASELINE	1	1999-06-19
Row 5	5.9	5.9	10*9/uL	11	4			Y	Y		BASELINE	1	1999-06-19
Row 6	0.4	0.4	10*9/L	3	1.2			Y	Y		BASELINE	1	1999-06-19
Row 7	5.1	5.1	10*9/L	8	2			Y	Y		BASELINE	1	1999-06-19
Row 8	7.5			9.00	5.00			Y	Y		BASELINE	1	1999-06-19
Row 9						NOT DONE	INSUFFI CIENT SAMPLE		Y		VISIT 1	2	1999-07-21
Row10	229	5.92	mmol/L	5.85*	0*				Y		VISIT 1	2	1999-07-21
Row 11	5.9	5.9	10*9/uL	11	4				Y		VISIT 1	2	1999-07-21

* The actual value used by the sponsor may depend upon the nature (e.g., precision) of the data.

9.4.4 PE Examples

The example shows data for one subject collected at two different visits. In all of the records except Rows 8 and 13, the data comes from the general physical examination.

Row 1	Shows how PESTRESC is populated if result is 'NORMAL'.
Row 2	Shows the proper use of PESTAT to indicate 'NOT DONE', and when PEREASND is used to indicate why a body system (PETEST) was not examined.
Rows 4-6	Show how PESPID is used to show the sponsor's sequence number used for identifying abnormalities within a body system.
Rows 4-7	Show how PESTRESC is populated if abnormality is dictionary coded.
Rows 8, 13	Show how the use of 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	PEMODIFY	PECAT	PELOC	PEBODSYS
Row 1	ABC	PE	ABC-001-001	1	HEAD	Head		GENERAL		
Row 2	ABC	PE	ABC-001-001	2	RESP	Respiratory		GENERAL		
Row 3	ABC	PE	ABC-001-001	3	ENT	Ear/nose/throat		GENERAL		
Row 4	ABC	PE	ABC-001-001	4	SKIN	Skin		GENERAL	FACE	SKIN AND
Row 5	ABC	PE	ABC-001-002	5	SKIN	Skin		GENERAL	HAND	SKIN AND
Row 6	ABC	PE	ABC-001-003	6	SKIN	Skin		GENERAL	LEFT ARM	SKIN AND
Row 7	ABC	PE	ABC-001-001	5	CARDIO	Cardiovascular		GENERAL		INVESTIGATIONS
Row 8	ABC	PE	ABC-001-001	6	FUNDOSCP	Fundoscopic		OPHTHALMOLOGIC		
Row 9	ABC	PE	ABC-001-001	7	RESP	Respiratory		GENERAL		
Row 10	ABC	PE	ABC-001-001	8	OTHO	Ear/nose/throat		GENERAL		
Row 11	ABC	PE	ABC-001-001	9	NECK	Neck		GENERAL		
Row 12	ABC	PE	ABC-001-001	10	CARDIO	Cardiovascular		GENERAL		
Row 13	ABC	PE	ABC-001-001	11	FUNDOSCP	Fundoscopic		OPHTHALMOLOGIC		

	PEORRES	PESTRESC	PESTAT	PEREASND	PESPID	VISIT	VISITNUM	VISITDY	PEDTC	PEDY
Row 1	NORMAL	NORMAL			1	BASELINE	1	1	1999-06-06	-3
Row 2			NOT DONE	INVESTIGATOR ERROR	1	BASELINE	1	1	1999-06-06	-3
Row 3	NORMAL	NORMAL			1	BASELINE	1	1	1999-06-06	-3
Row 4	Acne	ACNE NOS			1	BASELINE	1	1	1999-06-06	-3
Row 5	Allergic	DERMATITI			2	BASELINE	1	1	1999-06-07	-2
Row 6	Skin Rash	RASH			3	BASELINE	1	1	1999-06-08	-1
Row 7	SYSTOLIC	CARDIAC			1	BASELINE	1	1	1999-06-06	-3
Row 8	NORMAL	NORMAL			1	BASELINE	1	1	1999-06-06	-3
Row 9	NORMAL	NORMAL			1	VISIT 1	2	45	1999-07-21	45
Row 10	NORMAL	NORMAL			1	VISIT 1	2	45	1999-07-21	45
Row 11	NORMAL	NORMAL			1	VISIT 1	2	45	1999-07-21	45
Row 12	NORMAL	NORMAL			1	VISIT 1	2	45	1999-07-21	45
Row 13	NORMAL	NORMAL			1	VISIT 1	2	45	1999-07-21	45

9.4.5 QS Examples

Questionnaire 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.

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

Questionnaire Example 2:

This example shows 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 starts again at 1.

Row 11 - Shows derived records. Notice how QSORRES is blank for derived records.

	STUDYID	DOMAIN	USUBJID	QSSEQ	QSTESTCD	QSTEST	QSCAT	QSSCAT	QSORRES	QSSTRESC
Row 1	STUDYX	QS	P0001	1	COG01T02	ARM	ADAS	WORD RECALL	NO	N
Row 2	STUDYX	QS	P0001	2	COG01T02	ARM	ADAS	WORD RECALL	NO	N
Row 3	STUDYX	QS	P0001	3	COG01T02	ARM	ADAS	WORD RECALL	NO	N
Row 4	STUDYX	QS	P0001	4	COG01T03	BUTTER	ADAS	WORD RECALL	NO	N
Row 5	STUDYX	QS	P0001	5	COG01T03	BUTTER	ADAS	WORD RECALL	NO	N
Row 6	STUDYX	QS	P0001	6	COG01T03	BUTTER	ADAS	WORD RECALL	NO	N
Row 7	STUDYX	QS	P0001	7	COG01T04	CABIN	ADAS	WORD RECALL	NO	N

	STUDYID	DOMAIN	USUBJID	QSSEQ	QSTESTCD	QSTEST	QSCAT	QSSCAT	QSORRES	QSSTRESC
Row 8	STUDYX	QS	P0001	8	COG01T04	CABIN	ADAS	WORD RECALL	NO	N
Row 9	STUDYX	QS	P0001	9	COG01T04	CABIN	ADAS	WORD RECALL	NO	N
Row 10	STUDYX	QS	P0001	10	COG01T09	GRASS	ADAS	WORD RECALL	NO	N
Row 11	STUDYX	QS	P0001	11	COG01X	WORD RECALL	ADAS	WORD RECALL		9

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

9.4.6 SC Examples

Subject Characteristics 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 child-bearing status, eye color, and initials were collected. Row 13 illustrates use of an additional Race response. In this case, the value of RACE in DM would be 'ASIAN'.

	STUDYID	DOMAIN	USUBJID	SCSEQ	SCTESTCD	SCTEST	SCORRES	SCSTRESC	VISIT	VISITNUM	SCDTC	SCDY
Row 1	ABC	SC	ABC-001-001	1	CHILD	Child Bearing Status	NO	N	BASELINE	1	1999-06-19	1
Row 2	ABC	SC	ABC-001-001	2	EYE	Eye Color	BROWN	BROWN	BASELINE	1	1999-06-19	1
Row 3	ABC	SC	ABC-001-001	3	INITIALS	Subject Initials	HLT	HLT	BASELINE	1	1999-06-19	1
Row 4	ABC	SC	ABC-001-002	1	CHILD	Child Bearing Status	YES	Y	BASELINE	1	1999-03-19	1
Row 5	ABC	SC	ABC-001-002	2	EYE	Eye Color	BLUE	BLUE	BASELINE	1	1999-03-19	1
Row 6	ABC	SC	ABC-001-002	3	INITIALS	Subject Initials	BAM	BAM	BASELINE	1	1999-03-19	1
Row 7	ABC	SC	ABC-001-003	1	CHILD	Child Bearing Status	YES	Y	BASELINE	1	1999-05-03	1
Row 8	ABC	SC	ABC-001-003	2	EYE	Eye Color	GREEN	GREEN	BASELINE	1	1999-05-03	1
Row 9	ABC	SC	ABC-001-003	3	INITIALS	Subject Initials	ALM	ALM	BASELINE	1	1999-05-03	1
Row 10	ABC	SC	ABC-002-001	1	CHILD	Child Bearing Status	YES	Y	BASELINE	1	1999-06-14	1
Row 11	ABC	SC	ABC-002-001	2	EYE	Eye Color	HAZEL	HAZEL	BASELINE	1	1999-06-14	1
Row 12	ABC	SC	ABC-002-001	3	INITIALS	Subject Initials	CQH	CQH	BASELINE	1	1999-06-14	1
Row 13	ABC	SC	ABC-002-001	4	RACEOTH	Additional Race	JAPANESE	JAPANESE	BASELINE	1	1999-06-14	1

9.4.7 VS Examples

The example below shows one subject with two visits, Baseline and Visit 1. In this example the subject had two baseline temperatures taken and the second one, Row 5, was used as the baseline. That record has a 'Y' in the VSBLFL variable and the other temperature record has Null for VSBLFL.

1. **Rows 4, 5** - Show two records for the same visit, but only one is flagged as the one to use as baseline.

2. **Row 7**   - Shows a value that was collected once (not necessarily as part of a Vitals Signs CRF module), but given the nature of the data it was mapped to the VS domain.

3. **Row 10**   - Shows a value collected in one unit, but converted to selected standard unit.

4. **Row 11**   - Shows the proper use of the STAT variable to indicate 'NOT DONE' where a reason was collected when a test was not done.

	STUDYID	DOMAIN	USUBJID	VSSEQ	VSTESTCD	VSTEST	VSPOS	VSLOC	VSORRES	VSORRESU
Row 1	ABC	VS	ABC-001-001	1	SYSBP	Systolic Blood Pressure	SITTING	LEFT ARM	154	mmHg
Row 2	ABC	VS	ABC-001-001	2	DIABP	Diastolic Blood Pressure	SITTING	LEFT ARM	44	mmHg
Row 3	ABC	VS	ABC-001-001	3	PULSE	Pulse	SITTING	LEFT ARM	72	bpm
Row 4	ABC	VS	ABC-001-001	4	TEMP	Temperature		MOUTH	34.7	C
Row 5	ABC	VS	ABC-001-001	5	TEMP	Temperature		MOUTH	36.2	C
Row 6	ABC	VS	ABC-001-001	6	WEIGHT	Weight	STANDING		90.5	kg
Row 7	ABC	VS	ABC-001-001	7	HEIGHT	Height	STANDING		157	cm
Row 8	ABC	VS	ABC-001-001	8	SYSBP	Systolic Blood Pressure	SITTING	LEFT ARM	95	mmHg
Row 9	ABC	VS	ABC-001-001	9	DIABP	Diastolic Blood Pressure	SITTING	LEFT ARM	44	mmHg
Row 10	ABC	VS	ABC-001-001	10	TEMP	Temperature		MOUTH	97.2	F
Row 11	ABC	VS	ABC-001-001	11	WEIGHT	Weight				

	VSSTRESC	VSSTRESN	VSSTRESU	VSSTAT	VSREASND	VSBLFL	VISIT	VISITNUM	VISITDY	VSDTC
Row 1	154	154	mmHg			Y	BASELINE	1	1	1999-06-19
Row 2	44	44	mmHg			Y	BASELINE	1	1	1999-06-19
Row 3	72	72	bpm			Y	BASELINE	1	1	1999-06-19
Row 4	34.7	34.7	C				BASELINE	1	1	1999-06-19T08:45
Row 5	36.2	36.2	C			Y	BASELINE	1	1	1999-06-19T09:00
Row 6	90.5	90.5	kg			Y	BASELINE	1	1	1999-06-19
Row 7	157	157	cm			Y	BASELINE	1	1	1999-06-19
Row 8	95	95	mmHg				VISIT 1	2	1	1999-07-21
Row 9	44	44	mmHg				VISIT 1	2	1	1999-07-21
Row 10	36.2	36.2	C				VISIT 1	2	1	1999-07-21
Row 11				NOT DONE	SUBJECT REFUSED		VISIT 1	2	1	1999-07-21

9.5   TRIAL DESIGN EXAMPLES

Defining Epochs – a how-to example using the example crossover study.

Example of a full TA 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	

Example of the TE dataset for the same crossover study:

ELEMENT	TESTRL	TEENRL	TEDUR
Screen	Informed consent	2 weeks after start of element	14D
Placebo	First dose of placebo	2 weeks after start of element	14D
5 mg	First dose of 5 mg drug	2 weeks after start of element	14D
10 mg	First dose of 10 mg drug	2 weeks after start of element	14D
Rest	48 hrs after last dose drug	1 week after start of element	7D
Follow-up	48 hrs after last dose drug	3 weeks after start of element	21D

Example of a join of the TA & TE datasets (merge on Element):

ARM	TAETORD	ELEMENT	TESTRL	TEENRL	TADUR	TABRANCH
P-5-10	1	Screen	Informed consent	2 weeks after start of element	14D	Randomized to Placebo/5 mg/10 mg
P-5-10	2	Placebo	First dose of placebo	2 weeks after start of element	14D	
P-5-10	3	Rest	48 hrs after last dose drug	1 week after start of element	7D	
P-5-10	4	5 mg	First dose of 5 mg drug	2 weeks after start of element	14D	
P-5-10	5	Rest	48 hrs after last dose drug	1 week after start of element	7D	
P-5-10	6	10 mg	First dose of 10 mg drug	2 weeks after start of element	14D	
P-5-10	7	Follow-up	48 hrs after last dose drug	3 weeks after start of element	21D	
5-P-10	1	Screen	Informed consent	2 weeks after start of element	14D	Randomized to 5 mg/Placebo/10 mg
5-P-10	2	5 mg	First dose of 5 mg drug	2 weeks after start of element	14D	
5-P-10	3	Rest	48 hrs after last dose drug	1 week after start of element	7D	
5-P-10	4	Placebo	First dose of placebo	2 weeks after start of element	14D	
5-P-10	5	Rest	48 hrs after last dose drug	1 week after start of element	7D	
5-P-10	6	10 mg	First dose of 10 mg drug	2 weeks after start of element	14D	
5-P-10	7	Follow-up	48 hrs after last dose drug	3 weeks after start of element	21D	
5-10-P	1	Screen	Informed consent	2 weeks after start of element	14D	Randomized to 5 mg/10 mg/Placebo
5-10-P	2	5 mg	First dose of 5 mg drug	2 weeks after start of element	14D	
5-10-P	3	Rest	48 hrs after last dose drug	1 week after start of element	7D	
5-10-P	4	10 mg	First dose of 10 mg drug	2 weeks after start of element	14D	
5-10-P	5	Rest	48 hrs after last dose drug	1 week after start of element	7D	
5-10-P	6	Placebo	First dose of placebo	2 weeks after start of element	14D	
5-10-P	7	Follow-up	48 hrs after last dose drug	3 weeks after start of element	21D	

ARM	TAETORD	ELEMENT	TASTRL	TAENRL	TADUR	TABRANCH
5-P-10	1	Screen	Informed consent	2 weeks after start of element	14D	Randomized to 5 mg/Placebo/10 mg
P-5-10	1	Screen	Informed consent	2 weeks after start of element	14D	Randomized to Placebo/5 mg/10 mg
5-10-P	1	Screen	Informed consent	2 weeks after start of element	14D	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	14D	
P-5-10	2	Placebo	First dose of placebo	2 weeks after start of element	14D	
5-10-P	2	5 mg	First dose of 5 mg drug	2 weeks after start of element	14D	
5-P-10	3	Rest	48 hrs after last dose drug	1 week after start of element	7D	
P-5-10	3	Rest	48 hrs after last dose drug	1 week after start of element	7D	
5-10-P	3	Rest	48 hrs after last dose drug	1 week after start of element	7D	
5-P-10	4	Placebo	First dose of placebo	2 weeks after start of element	14D	
P-5-10	4	5 mg	First dose of 5 mg drug	2 weeks after start of element	14D	
5-10-P	4	10 mg	First dose of 10 mg drug	2 weeks after start of element	14D	
5-P-10	5	Rest	48 hrs after last dose drug	1 week after start of element	7D	
P-5-10	5	Rest	48 hrs after last dose drug	1 week after start of element	7D	
5-10-P	5	Rest	48 hrs after last dose drug	1 week after start of element	7D	
5-P-10	6	10 mg	First dose of 10 mg drug	2 weeks after start of element	14D	
P-5-10	6	10 mg	First dose of 10 mg drug	2 weeks after start of element	14D	
5-10-P	6	Placebo	First dose of placebo	2 weeks after start of element	14D	
5-P-10	7	Follow-up	48 hrs after last dose drug	3 weeks after start of element	21D	
P-5-10	7	Follow-up	48 hrs after last dose drug	3 weeks after start of element	21D	
5-10-P	7	Follow-up	48 hrs after last dose drug	3 weeks after start of element	21D	

Example of the same join of TA & TE, sorted by TAETORD, then ARM:

Note that, for each group of records with the same value of TAETORD (the column labeled Order), the Start Rules are the same except for the specific dose being started (Placebo, 5 mg or 10 mg), and the End Rules are the same. This means that these Elements constitute a EPOCH. Each EPOCH has a common Start Rule and a common End Rule, though the 'EPOCH Start Rule' has to be restated as 'First dose of blinded study drug for this epoch.' This 'circular' definition of the start of an epoch in terms of the first dose of the epoch is the reason that epoch (in the form of TAETORD) needs to be collected in EXPOSURE.

Note that EPOCH can be considered a 'label' for TAETORD.

Also note the use of TAETORD as a 'primary' concept that will always exist, even for funny studies with arms that have different patterns.

This is the same table, with the EPOCH variable added.

ARM	TAETORD	ELEMENT	TASTRL	TAENRL	TADUR	EPOCH	TABRANCH
5-P-10	1	Screen	Informed consent	2 weeks after start of element	14D	Trial Screen	Randomized to 5 mg/Placebo/10 mg
P-5-10	1	Screen	Informed consent	2 weeks after start of element	14D	Trial Screen	Randomized to Placebo/5 mg/10 mg
5-10-P	1	Screen	Informed consent	2 weeks after start of element	14D	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	14D	1st Trt	
P-5-10	2	Placebo	First dose of placebo	2 weeks after start of element	14D	1st Trt	
5-10-P	2	5 mg	First dose of 5 mg drug	2 weeks after start of element	14D	1st Trt	
5-P-10	3	Rest	48 hrs after last dose drug	1 week after start of element	7D	1st Rest	
P-5-10	3	Rest	48 hrs after last dose drug	1 week after start of element	7D	1st Rest	
5-10-P	3	Rest	48 hrs after last dose drug	1 week after start of element	7D	1st Rest	
5-P-10	4	Placebo	First dose of placebo	2 weeks after start of element	14D	2nd Trt	
P-5-10	4	5 mg	First dose of 5 mg drug	2 weeks after start of element	14D	2nd Trt	
5-10-P	4	10 mg	First dose of 10 mg drug	2 weeks after start of element	14D	2nd Trt	
5-P-10	5	Rest	48 hrs after last dose drug	1 week after start of element	7D	2nd Rest	
P-5-10	5	Rest	48 hrs after last dose drug	1 week after start of element	7D	2nd Rest	
5-10-P	5	Rest	48 hrs after last dose drug	1 week after start of element	7D	2nd Rest	

ARM	TAETORD	ELEMENT	TASTRL	TAENRL	TADUR	EPOCH	TABRANCH
5-P-10	6	10 mg	First dose of 10 mg drug	2 weeks after start of element	14D	3rd Trt	
P-5-10	6	10 mg	First dose of 10 mg drug	2 weeks after start of element	14D	3rd Trt	
5-10-P	6	Placebo	First dose of placebo	2 weeks after start of element	14D	3rd Trt	
5-P-10	7	Follow-up	48 hrs after last dose drug	3 weeks after start of element	21D	Trial FU	
P-5-10	7	Follow-up	48 hrs after last dose drug	3 weeks after start of element	21D	Trial FU	
5-10-P	7	Follow-up	48 hrs after last dose drug	3 weeks after start of element	21D	Trial FU	

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	Ocagon_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
	Mi Chard Cunion	Aventis
	Gary Cunningham	AstraZeneca
	Kathleen Giordano	Merck
	Dan Godoy	AstraZeneca
	Andreas Gromen	Schering AG
	Joyce Hernandez	IBM
	Sandy Lei	Johnson and Johnson PRD
	Barrie Nelson	Amgen
	Musa Nsereko	AstraZeneca
	Bill Qubeck	Pfizer
	Robert Sarate	Merck
	Sarah Shiue	Aventis
	Gail Stoner	Centocor
	Mike Walega	Covance
	Gary Walker	Quintiles
	Yuguang Zhao	Sanofi-Synthelabo
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
SEND	Standard for Exchange of Nonclinical Data
STDM	Study Data Tabulation Model

10.3 STANDARDIZED AND RESERVED CODES

This appendix includes a 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.

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	Domain Model Type
AE	Adverse Events	Events
AU	Autopsy	Findings
BM	Bone Mineral Density (BMD) Data	Findings
BR	Biopsy	Findings
CM	Concomitant Meds	Interventions
CO	Comments	Special Purpose
CP	Compliance	Findings
DC	Disease Characteristics	Findings
DM	Demographics	Special Purpose
DR	Disease Response	Findings
DS	Disposition	Events
DV	Protocol Deviations	Findings
DY	Diary Data	Findings
EE	EEG	Findings
EG	ECG	Findings
EX	Exposure	Interventions
FH	Family History	Events
HU	Healthcare Resource Utilization	Findings
IE	Inclusion / Exclusion	Findings
IM	Imaging	Findings
LB	Laboratory Data	Findings
MB	Microbiology	Findings
MH	Medical History	Events
ML	Meal Data	Interventions
NE	Neurological Exam	Findings

Code	Domain	Domain Model Type
OM	Organ Measurements	Findings
PC	PK Concentration	Findings
PE	Physical Exam	Findings
PP	PK Parameters	Findings
QS	Questionnaires	Findings
SC	Subject Characteristics	Findings
SE	Subject Elements	Study Design
SG	Surgery	Interventions
SK	Skin Test	Findings
SL	Sleep (Polysomnography) Data	Findings
SS	Signs and Symptoms	Findings or Events
ST	Stress (Exercise) Test Data	Findings
SU	Substance Use	Interventions
SV	Subject Visits	Trial Design
TA	Trial Arms	Trial Design
TE	Trial Elements	Trial Design
TI	Trial Inclusion/Exclusion Criteria	Trial Design
TV	Trial Visits	Trial Design
VS	Vital Signs	Findings

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
FINDING	General ECG Finding
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
BSA	Body Surface Area
DIABP	Diastolic Blood Pressure
FRMSIZE	Frame Size
HEIGHT	Height
HR	Heart Rate
MAP	Mean Arterial Pressure
PULSE	Pulse Rate
RESP	Respiration Rate
SYSBP	Systolic Blood Pressure
TEMP	Body 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 (SUPPQUAL) special-purpose dataset.

QNAM	QLABEL
AESOSP	Other Medically Important SAE
AETRTEM	Treatment Emergent
COMPLT	Completers Population
FULLSET	Full Analysis Set
ITT	Intent to Treat Population Flag
PERPROT	Per Protocol Set
SAFETY	Safety population Flag

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 V3.1 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 naming limit of SAS transport files. When using fragments, the general rule is to use the fragment 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	Keyword(s)	Fragment
ACTION	ACN	ETHNICITY	ETHNIC
ADVERSE EVENT	AE	EXPOSURE	EX
BASELINE	BL	EXTERNAL	X
BIOPSY RESULTS	BR	EVALUATOR	EVAL
BIRTH	BRTH	FAMILY MEDICAL HISTORY	FH
BODY	BOD	FASTING	FAST
CANCER	CAN	FILENAME	FN
CATEGORY	CAT	FLAG	FL
CHARACTER	C	FORMULATION, FORM	FRM
CONDITION	CND	FREQUENCY	FREQ
CLASS	CLAS	GRADE	GR
CODE	CD	GROUP	GRP
COMMENT	COM	HIGHER LIMIT	HI
COMMENTS	CO	HOLTER MONITORING	HM
COMPLIANCE	CP	HOSPITALIZATION	HOSP
CONCOMITANT	CON	IDENTIFIER	ID
CONCOMITANT MEDICATION	CM	IMAGING	IM
CONGENITAL	CONG	INCLUSION / EXCLUSION	IE
DATE TIME - CHARACTER	DTC	INDICATION	INDC
DAY	DY	INDICATOR	IND
DEATH	DTH	INVESTIGATOR	INV
DECODE	DECOD	LABORATORY	LB
DEMOGRAPHY	DM	LESIONS	LS
DERIVED	DRV	LIFE-THREATENING	LIFE
DESCRIPTION	DESC	LOCATION	LOC
DISABILITY	DISAB	LOINC CODE	LOINC
DISEASE CLASSIFICATION	DC	LOWER LIMIT	LO
DISPOSITION	DS	MEDICAL HISTORY	MH
DOSE, DOSAGE	DOS, DOSE	MEDICALLY-IMPORTANT EVENT	MIE
DURATION	DUR	MICROBIOLOGY	MB
ELAPSED	EL	NAME	NAM
ELECTROCARDIOGRAM	EG	NON-STUDY THERAPY	NST
ELEMENT	ET	NORMAL RANGE	NR
EMERGENT	EM	NOT DONE	ND
END	END, EN	NUMBER	NUM

Keyword(s)	Fragment
NUMERIC	N
ONGOING	ONGO
ORIGIN	ORIG
ORIGINAL	OR
OTHER	OTH, O
OUTCOME	OUT
OVERDOSE	OD
PATTERN	PATT
PHARMACOKINETICS	PK
PHYSICAL EXAMINATION	PE
POSITION	POS
PROTOCOL DEVIATIONS	DV
PULMONARY FUNCTION	PF
QUESTIONNAIRES	QS
REASON	REAS
REFERENCE	REF, RF
REGIMEN	RGM
RELATED	REL, R
RELATIONSHIP	REL
RESULT	RES
RULE	RL
SEQUENCE	SEQ
SERIOUS	S, SER
SEVERITY	SEV
SKIN TEST	SK
SPECIMEN	SPEC, SPC
SPONSOR	SP
STANDARD	ST, STD
START	ST
STATUS	STAT
SUBCATEGORY	SCAT
SUBJECT	SUBJ
SUBSTANCE USE	SU
SYSTEM	SYS
TEXT	TXT

Keyword(s)	Fragment
TIME	TM
TIMEPOINT	TPT
TOTAL	TOT
TOXICITY	TOX
TRANSITION	TRANS
TREATMENT	TRT
UNIT	U
UNIQUE	U
UNPLANNED	UP
VARIABLE	VAR
VALUE	VAL
VITAL SIGNS	VS

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 over 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 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, SEQSP, 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, 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 they did not easily fit into a specific observation class; this was in part due to the fact that our 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 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 datasets highlights the importance of and the need for specific controlled terminology and 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 reliationships (one to many to many) exis 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 colleciton, the SDTM modelsare optimnized for electronic data exchange with the FDA. But 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.