Study Data Tabulation Model

Prepared by the

CDISC Submission Data Standards Team

Principal Editor: Wayne Kubick

Principal Contributors: Fred Wood, Diane Wold, Tom Guinter, Julie Evans, CDISC SDS Team

Notes to Readers

This is the released Version 1.0 of the Study Data Tabulation Model Document, previously posted for comment by the CDISC Submissions Data Standards team. This document, which supersedes all prior versions, reflects changes from two comment periods: an initial comment period in March/April 2004 through CDISC and the HL7 Regulated Clinical Research Information Management Technical Committee, and a second review period from May 27 to June 10, 2004 through CDISC.

Revision History

Date	Version	Summary of Changes	Primary Authors
20040625	Version 1.0	Released version reflecting all changes identified during comment periods.	Kubick, Wood, Evans, Wold, Guinter

CONTENTS

1 Introduction

1.1 PURPOSE

This document describes the Study Data Tabulation Model (SDTM), which defines a standard structure for study data tabulations that are to be submitted as part of a product application to a regulatory authority such as the United States Food and Drug Administration (FDA). This document is based on material prepared by the Submissions Data Standards (SDS) Team of the Clinical Data Interchange Standards Consortium (CDISC). This is the second version of an informative document describing the model, and supersedes all prior versions. The initial version, which was originally balloted through HL7 as the Clinical Trial Data Regulatory Submission Model in April 2003, and by CDISC as the CDISC Submission Data Domain Models Version 3.0 (V3), was approved as an HL7 informative document in June 2003.

This document varies in several significant ways from the prior released version (V3):

- The name has been changed to better reflect a broader scope; the model is intended to apply to data tabulations and not to other data presentations (e.g., data listings, analysis datasets, subject profiles). In addition, it is intended to apply to data collected in both human and animal studies.
- Detailed assumptions, business rules, examples, and specifications for representative data domains have been removed from this document and will now be published as separate implementation guides available through CDISC. Implementation details for Clinical Trials data submissions are published as the CDISC Submission Data Standards Implementation Guide. Implementation details for Animal Toxicology data submissions, which have been prepared by the Standard for Exchange of Nonclinical Data (SEND) Consortium and also based on this model, will be published as the SEND Implementation Guide.
- The model has been expanded to include additional variables for each general observation class, as well as additional concepts. Variables that were marked as deprecated in the prior version have been removed.

Other specific changes to the model content are described below in Sections 1.1 and 1.3.

The SDTM has been prepared by the CDISC Submission Data Standards (SDS) Team to guide the organization, structure, and format of tabulation datasets for study data submitted to regulatory authorities. Data tabulation datasets are one of four ways to represent the human subject Case Report Tabulation (CRT) and equivalent animal data submitted to the FDA. CRTs are also submitted in the format of subject profiles, data listings, and analysis datasets. One benefit to industry of submitting data tabulation datasets that conform to the standard structure is that it minimizes the need to submit the same data in multiple formats.

The availability of standard submission data provides many benefits to regulatory reviewers. Reviewers can now 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 that they will support the FDA's efforts to develop a repository for all submitted studies and a suite of standard review tools to access, manipulate, and view the study data.

This document is intended for companies and individuals involved in the collection, preparation, and analysis of study data submitted to regulatory authorities. Audiences are encouraged to read the CDISC Submission Metadata Model for additional historical background on how to provide metadata for submissions. The primary goal of the Metadata Model, which was developed for the CDISC Version 2 (V2) standards, 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 documents before preparing a regulatory submission based on the SDTM.

1.2 RELATIONSHIP TO PRIOR CDISC DOMAIN MODELS

As stated above, this document is the first formal revision to the CDISC V3 Submission Data Domain Models. V3 represented a major change from the CDISC V2 domain models because it incorporated a general model for representing all types of study data. This version was initially released for comment in March 2003. A final version of V3, which addressed most comments received during the review period, was approved by HL7 as an informative document and released for publication on June 9, 2003. Participants from a group of nine sponsor companies then tested this version in summer 2003 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 by both sponsors and the FDA 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 between the first review version and V3 are summarized below:

- Corrections and amendments to what was previously known in V3 as the General Study Information Model to improve consistency, including the incorporation of new variables and new concepts
- Incorporation of a new "Trial Design" component to the SDTM
- A more thorough solution for defining relationships between datasets, between records in different domains, and between supplemental qualifiers and a parent domain
- Representation of all date/time variables in ISO8601 character format, and the elimination of the concept of a separate Date/Time Precision variable for each date/time variable
- New domain variables to represent additional timing descriptions, flags, and descriptive attributes of an observation (e.g., --SCAT, --DOSRGM, --NRIND)
- Removal of some variables within domains (e.g., --INTP, --DESC, --BLRESC, --BLRESN) that were either deprecated in the prior version or were inconsistent with the intent of the model
- Numerous changes to variables, labels, formats, and notes to reduce ambiguity and improve consistency.

While this released version is expected to be an implementation-ready version for clinical studies involving human drug products, future improvements and enhancements may be necessary to support a broader range of regulated products. Efforts will continue to further evaluate the model for human and animal studies involving other regulated products including food additives; therapeutic biologics; blood derivatives; vaccines; cellular, tissue, and gene therapy; and devices. Structured evaluation pilots of the SDTM are planned for these products, and the lessons learned from these pilots would be used in developing future enhancements to the standard. For example, specific incremental requirements for animal toxicity data will be further addressed in future versions after an ongoing FDA pilot project is completed. Implementation guides for applying the model to each type of data and guidance on controlled terminology will be published separately.

1.3 SIGNIFICANT CHANGES FROM SDTM VERSION 0.8

The first review version (Version 0.8) of the SDTM was posted for comment on the CDISC website and used in the HL7 ballot process in March and April of 2004. A second review version (Version 0.92) of the SDTM incorporated changes that were identified as a result of the initial comment period. The 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 between Version 0.8 and 0.92 are summarized below:

- Revised cover page to remove references to HL7 (since this document will not be submitted to HL7 for reballot at this time)
- Made substantial revisions to last paragraph of Section 1.2
- Added new sections 1.3 and 1.4
- Added description of Rule variables in Section 2.1
- Replaced the qualifier variable--SPREF with the identifier variable --SPID in all 3 general classes (tables 2.2.1, 2.2.2, 2.2.3, 2.2.4
- Removed the variable --TRTEM from the Events class (table 2.2.2); this variable, which requires an Evaluator, should be represented in Supplemental Qualifiers

- Removed the variable --SOTHC from the Events class (table 2.2.2); this variable, when used, should be represented in Supplemental Qualifiers
- Renamed the variable --RELOTH from the Events class (table 2.2.2) to --RELNST to retain consistency of meaning of the SDS "OTH" naming fragment
- Renamed the variable --SOTH from the Events class (table 2.2.2) to --SMIE to retain consistency of meaning of the SDS "OTH" naming fragment
- Redefined and renamed the variable --TOXCAT from the Findings class (table 2.2.3) to --TOX to more accurately describe the intended concept
- Removed the variable --LNKSEQ from the Findings class (table 2.2.3); this variable was determined not to be necessary, given subsequent revisions to the general structure for relationships defined in this version
- Redefined the variable --TOXGR from the Findings class (table 2.2.3) to more accurately describe the intended concept.
- Redefined the timing variables EPOCH, --STRF and --ENRF (table 2.2.5) to more accurately describe the intended concept
- Removed the variable RACEOTH from the Demographics special purpose domain (table 2.2.6)
- Added variable TEDUR to the Trial Design model (table 3.2.1)
- Renamed the variable TICRIT from the Trial Design Model (table 3.4.1) to TIRL to more accurately describe the intended concept
- Redefined the general structure for expressing relationships among records in Section 4.0. This resulted in several variable name changes in tables 2.2.7, 4.2.1 and 4.4.1 and the removal of table 4.3.1 (which is now addressed in table 4.2.1 under the new approach).

In addition, there were several other minor text edits and changes to the controlled terminology (*) indicator in several of the tables.

Version 1.0 reflects the following additional corrections that were identified in the second review period:

- Abridgement of certain variable labels (to conform with the SAS Transport 40 character limit)
- Identified additional variables that may be subject to controlled terminology
- Added a new variable --METHOD to the Findings domain
- Corrected minor typographical errors throughout.

1.4 RELATIONSHIP TO HL7 MODELS

Readers from the HL7 community will notice that the SDTM was not developed under the HL7 Development Framework as an HL7 model – it describes the content of submission data, rather than the entities, acts, roles and participations typically modeled by HL7. Instead, CDISC has developed the SDTM to provide a standardized approach for submitting study data to regulatory authorities consistent with current industry practices and regulatory requirements, which specify use of the SAS Version 5 transport format. However, CDISC believes that the essential model concepts can be mapped to the HL7 V3 RIM and adapted for HL7 data types in the future.

2 Model Fundamentals

2.1 MODEL CONCEPTS AND TERMS

The SDTM provides a general framework for describing the organization of information collected during human and animal studies.

The model is built around the concept of observations, which consist of discrete pieces of information collected during a study. Observations normally correspond to rows in a dataset. A collection of observations on a particular topic is considered a domain. For example, "Subject 101 had mild nausea starting on Study Day 6" is an observation belonging to the Adverse Events domain in a clinical trial. Similarly, "Animal 525 weighed 250 grams on Study Day 6" would represent an observation belonging to the Body Weights domain in an animal toxicity study.

Each observation can be described by a series of named variables. Each variable, which normally corresponds to a column in a dataset, 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 five major roles:

- A common set of *Identifier* variables, which identify the study, the subject (individual human or animal) involved in the study, 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), and vary according to the type of observation.
- A common set of *Timing* variables, which describe the timing of an 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). The list of Qualifier variables included with a domain will vary considerably depending on the type of observation and the specific domain
- *Rule* variables, which 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 (e.g., HEMATOLOGY as a category classification for laboratory results)
- *Result Qualifiers* describe the specific results associated with the topic variable for a finding (e.g., the original result and the standardized result)
- *Synonym Qualifiers* specify an alternative name for a particular variable in an observation (e.g., the verbatim term and the preferred term for an adverse event)
- *Record Qualifiers* define additional attributes of the observation record as a whole (e.g., the reason a medication was taken)
- *Variable Qualifiers* are used to further modify or describe a specific variable within an observation (e.g., the units for a laboratory result).

For the example 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.

Observations are reported in a series of domains, usually corresponding to data that were collected together. A domain is defined as a collection of observations with a topic-specific commonality about a subject. Each dataset is distinguished by a unique, two-character identifier that should be used consistently throughout the submission. Standardized dataset codes are available in the CDISC SDS and SEND implementation guides.

The dataset structure for a collection of observations is a flat file representing a table with one or more rows and columns, with each row representing an observation and each column representing a variable. Normally, one dataset is submitted for each domain. 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. The Define document describes each variable in the dataset using seven distinct metadata attributes to be defined for each dataset variable included.

- The unique *Variable Name* based upon those described in the SDTM
- 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); an asterisk (*) after the Char type indicates that a discrete set of values (controlled terminology) is expected to be made available for the variable
- The set of controlled Terminology for the value or the presentation format of the data (e.g., Y, N; or ISO 8601)
- The *Origin* or source of each variable (e.g., CRF, Derived, Sponsor Defined)
- The *Role* of the variable (e.g., Identifier, Topic, Timing, or Qualifier), which describes how the variable is used in the dataset (this attribute need not be submitted except for sponsor extensions to standard roles)
- *Comments* or other relevant information about the variable or its data.

Data stored in these 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). The SDTM describes the name, label, role, and type for the standard variables; the origin and terminology would be sponsor defined for each particular study. Note that current types are restricted to character and number for compatibility with SAS; it is expected that additional, more descriptive datatypes (e.g., integer, float, date, date/time) will be used in the future.

When creating submissions, a sponsor may 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. Sponsors should consult the appropriate implementation guides which specifically describe which variables are required, expected, or permissible to use in specific domains based on the general observation classes.

2.2 THE GENERAL OBSERVATION CLASSES

The majority of observations collected during a study can be divided among three general classes: Interventions, Events, or Findings:

- The *Interventions* class, described in Table 2.2.1, 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”), or preceding or coincident with the study assessment period (e.g., “concomitant medications”)
- The *Events* class, described in Table 2.2.2, captures planned protocol milestones such as randomization and study completion (“disposition”), and occurrences or incidents independent of planned study evaluations occurring during the trial (e.g., “adverse events”) or prior to the trial (e.g., “medical history”).
- The *Findings* class, described in Table 2.2.3, captures the observations resulting from planned evaluations to address specific questions such as observations made during a physical examination, laboratory tests, histopathology, ECG testing, and questions listed on questionnaires.

Datasets based on any of the general observation classes share a set of common Identifier variables and Timing variables. The set of Unique Identifier variables used for all observations is described in Table 2.2.4. The set of Timing variables that should be used for all three general observation classes is included in Table 2.2.5. As a general rule, any valid optional Identifier or Timing variable is permissible for use in any submission dataset. While the choice of Timing variables may vary between studies for a given submission, the meaning of a Timing variable should be consistent within a submission.

In the tables below, the presence of two hyphens before the variable name (e.g., --TRT) is used to indicate the required use of a prefix based on the 2-character domain code. The domain code is used as a prefix to support the submission of datasets in the SAS Version 5 Transport format, currently designated as the standard format required

by the FDA. The use of this prefix may be deprecated in the future once datasets are submitted in an alternative format such as XML, which is not subject to the same limitations as SAS Version 5 Transport format.

In addition to the three general observation classes, a submission may include a set of other special purpose datasets of specific standardized structures to represent additional important information. Examples include:

- A Demographics special-purpose domain is included with human studies, described in Section 2.2.6
- Datasets to describe the design of a trial, described in Section 3
- Datasets to represent the relationships between datasets and records (including a general Comments domain introduced in Section 2.2.7), described in Section 4.

2.2.1 The Interventions Observation Class

Table 2.2.1: Interventions — Topic and Qualifier Variables, One Record per Intervention (--TRT)

Variable Name	Variable Label	Type	Description
	Topic Variable		
--TRT	Name of Treatment	Char	The topic for the intervention observation, usually the verbatim name of the treatment, drug, medicine, or therapy given during the dosing period for the observation. Use --MODIFY and --DECOD for modifying and coding the verbatim name, respectively.
	Qualifier Variables		
--MODIFY	Modified Treatment Name	Char	If the value for --TRT is modified as part of a defined procedure, then the modified text is placed here.
--DECOD	Standardized Treatment Name	Char*	Standardized or dictionary-derived name of the topic variable, --TRT, or the modified topic variable (--MODIFY), if applicable. Equivalent to the generic drug name in WHO Drug, or a term in SNOMED, ICD9, or other published or sponsor-defined dictionaries. The dictionary name and version should be provided in the comments column in the data definition document.
--CAT	Category	Char*	Used to define a category of related records.
--SCAT	Subcategory	Char*	Used to define a further categorization level for a group of related records.
--OCCUR	Occurrence	Char*	Used only when the occurrence of specific interventions is solicited. Valid values include Y and N.
--STAT	Status	Char*	Used to indicate that a planned intervention was not performed. Should be null or have a value of NOT DONE.
--REASND	Reason	Char	Reason not done. Used in conjunction with --STAT when value is NOT DONE.
--INDC	Indication	Char	Denotes the indication for the intervention (e.g., why the therapy was taken or administered).
--CLAS	Class	Char*	Class for a medication or treatment when the dictionary used codes to a single class.
--CLASCD	Class Code	Char*	Used to store dictionary codes for --CLAS when the dictionary used codes to a single class.
--DOSE	Dose	Num	Amount of --TRT given.
--DOSTXT	Dose Description	Char	Dosing amounts or a range of dosing information collected in text form. Example: 200-400.
--DOSU	Dose Units	Char*	Units for --DOSE such as ng, mg, mg/kg.
--DOSFRM	Dose Form	Char*	Dose form for the treatment. Examples: TABLET, CAPSULE.
--DOSFRQ	Dosing Frequency per Interval	Char*	Usually expressed as the number of doses given per a specific interval. Examples: BID, TID, QID.
--DOSTOT	Total Daily Dose Using DOSU	Num	Total daily dose of --TRT using the units in --DOSU.
--DOSRGM	Intended Dose Regimen	Char	Text description of the (intended) schedule or regimen for the Intervention. Examples: two weeks on, two weeks off.
--ROUTE	Route of Administration	Char*	Route of administration for the intervention. Examples: ORAL, INTRAVENOUS.
--LOT	Lot Number	Char	Lot number for the intervention.
--LOC	Location of Dose Administration	Char*	Specifies anatomical location of an intervention, such as an injection site.

* *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 or from an external published source such as MedDRA. See the Implementation Guide for details on which terminology is sponsor-controlled or standardized.*

2.2.2 The Events Observation Class

Table 2.2.2: Events — Topic and Qualifier Variables, One Record per Event (--TERM)

Variable Name	Variable Label	Type	Description
	Topic Variable		
--TERM	Reported Term	Char	Topic variable for an event observation, which is the verbatim name of the event.
	Qualifier Variables		
--MODIFY	Modified Reported Term	Char	If the value for --TERM is modified as part of a defined procedure, then the modified text is placed here.
--DECOD	Dictionary-Derived Term	Char*	Dictionary or sponsor-defined derived text description of the topic variable, --TERM, or the modified topic variable (--MODIFY), if applicable. Equivalent to the Preferred Term (PT in MedDRA). The dictionary name and version should be provided in the comments column in the Define document.
--CAT	Category	Char*	Used to define a category of related records.
--SCAT	Subcategory	Char*	Used to define a further categorization level for a group of related records.
--OCCUR	Occurrence	Char*	Used when the occurrence of specific events is solicited. Valid values include Y and N. Values are null for spontaneously reported events.
--STAT	Status	Char*	Used to indicate when a question about the occurrence of an event was not answered. Should be null or have a value of NOT DONE.
--REASND	Reason	Char	Reason not done. Used in conjunction with --STAT when its value is NOT DONE.
--BODSYS	Body System or Organ Class	Char*	Body system or organ class that is involved in an event or measurement from the standard hierarchy (e.g., Primary SOC in MedDRA).
--LOC	Location of the Reaction	Char*	Describes anatomical location relevant for the event. Example: LEFT ARM for skin rash.
--SEV	Severity/Intensity	Char*	The severity or intensity of the event. Examples: MILD, MODERATE, SEVERE.
--SER	Serious Event	Char*	Is this is a serious event?
--ACN	Action Taken with Study Treatment	Char*	Describes changes to the study treatment as a result of the event.
--ACNOTH	Other Action Taken	Char	Describes other actions taken as a result of the event.
--REL	Causality	Char*	Records the investigator's opinion regarding the causality of the event to the treatment. Examples: DEFINITELY NOT RELATED, PROBABLY NOT RELATED, POSSIBLY RELATED, PROBABLY RELATED, or DEFINITELY RELATED.
--RELNST	Relationship to Non-Study Treatment	Char	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."
--PATT	Pattern of Event	Char*	Used to indicate the pattern of the event over time as collected. Examples: INTERMITTENT, CONTINUOUS, SINGLE EVENT.
--OUT	Outcome of Event	Char*	Description of the outcome of an event. Examples: RECOVERED/RESOLVED, RECOVERING/RESOLVING, NOT RECOVERED/NOT RESOLVED, RECOVERED/RESOLVED WITH SEQUELAE, FATAL, or UNKNOWN.
--SCAN	Involves Cancer	Char*	Was the event associated with the development of cancer?
--SCONG	Congenital Anomaly or Birth Defect	Char*	Was the event associated with congenital anomaly or birth defect?
--SDISAB	Persist or Signif Disability/Incapacity	Char*	Did the event result in persistent or significant disability/incapacity?
--SDTH	Results in Death	Char*	Did the event result in death?
--SHOSP	Requires or Prolongs Hospitalization	Char*	Did the event require or prolong hospitalization?
--SLIFE	Is Life Threatening	Char*	Was the event life threatening?
--SOD	Occurred with Overdose	Char*	Did the event occur with an overdose?
--SMIE	Other Medically Important Serious Event	Char*	Do additional categories for seriousness apply?

Variable Name	Variable Label	Type	Description
--CONTRT	Concomitant or Additional Trtmnt Given	Char*	Was another treatment given because of the occurrence of the event?
--TOXGR	Toxicity Grade	Char*	Records toxicity grade using a standard toxicity scale (such as the NCI CTCAE). Sponsor should specify which scale and version is used in sponsor comments.

* *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 or from an external published source such as MedDRA. See the Implementation Guide for details on which terminology is sponsor-controlled or standardized.*

2.2.3 The Findings Observation Class

Table 2.2.3: Findings — Topic and Qualifier Variables, One Record per Finding (--TESTCD)

Variable Name	Variable Label	Type	Description
	Topic Variable		
--TESTCD	Short Name of Measurement, Test or Examination	Char*	Short character value for --TEST used as a column name when converting a dataset from a vertical format to a horizontal format. The short value can be up to 8 characters. Examples: PLATELET, SYSBP, PR, EYEEXAM.
	Qualifier Variables		
--TEST	Name of Measurement, Test or Examination	Char*	Verbatim name, corresponding to the topic variable, of the test or examination used to obtain the measurement or finding. Examples: Platelet Count, Systolic Blood Pressure, PR Interval, Eye Examination.
--MODIFY	Modified Term	Char	If the value of --TEST is modified as part of a defined procedure, then the modified text is placed here.
--CAT	Category	Char*	Used to define a category of related records. Examples: HEMATOLOGY, URINALYSIS, CHEMISTRY, HAMILTON DEPRESSION SCALE, SF36.
--SCAT	Subcategory	Char*	Used to define a further categorization level for a group of related records. Example: DIFFERENTIAL.
--POS	Position of Subject During Observation	Char*	Position of the subject during a measurement or examination. Examples: SUPINE, STANDING, SITTING.
--BODSYS	Body System or Organ Class	Char*	Body System or Organ Class that is involved in an event or measurement from the standard hierarchy. Example: the Primary SOC in MedDRA.
--ORRES	Result or Finding in Original Units	Char	Result of the measurement or finding as originally received or collected.
--ORRESU	Original Units	Char*	Unit for --ORRES. Examples: IN, FT, LB, g, L, g/L.
--ORNRLO	Normal Range Lower Limit-Original Units	Char	Lower end of normal range or reference range for continuous measurements or findings in original units.
--ORNRHI	Normal Range Upper Limit-Original Units	Char	Upper end of normal range or reference range for continuous measurements or findings in original units.
--STRESC	Result or Finding in Standard Format	Char	Contains the result value for all findings, copied or derived from --ORRES in a standard format or in standard units. --STRESC should store all results or findings in character format; if results are numeric, they should also be stored in numeric format in --STRESN. For example, if various tests have results 'NONE', 'NEG', and 'NEGATIVE' in --ORRES and these results effectively have the same meaning, they could be represented in standard format in --STRESC as "NEGATIVE". For other examples, see the implementation guides.
--STRESN	Numeric Result or Finding in Standard Units	Num	Used for continuous or numeric results or findings in standard format; copied in numeric format from --STRESC. --STRESN should store all numeric test results or findings.
--STRESU	Standard Units	Char*	Standardized units used for --STRESC or --STRESN. Example: mmol/L.
--STNRLO	Normal Range Lower Limit-Standard Units	Num	Lower end of normal range or reference range for continuous results or findings in standard units.

Variable Name	Variable Label	Type	Description
--STNRHI	Normal Range Upper Limit-Standard Units	Num	Upper end of normal range or reference range for continuous results or findings in standard units.
--STNRC	Normal Range for Character Results-Standard Units	Char	For normal range or reference range values that are character in ordinal scale. Example: Negative to Trace.
--NRIND	Normal/Reference Range Indicator	Char*	Used to indicate the value is outside the normal range or reference range defined by --ORNRLO and --ORNRHI. Examples include Y, N, H, L..
--STAT	Status	Char*	Used to indicate that a question was not asked or a test was not done. Should be null or have a value of NOT DONE.
--REASND	Reason	Char	Reason not done. Used in conjunction with --STAT when value is NOT DONE.
--XFN	External Filename	Char	Filename for an external file, such as one for an ECG waveform or a medical image.
--NAM	Laboratory/Vendor Name	Char	Name or identifier of the vendor (e.g., laboratory) that provided the test results.
--LOINC	LOINC Code	Char*	LOINC Code for the topic variable such as a lab test.
--SPEC	Specimen Material Type	Char*	Defines the type of specimen used for a measurement. Examples: SERUM, PLASMA, URINE.
--SPCCND	Specimen Condition	Char*	Defines the condition of the specimen. Example: cloudy.
--LOC	Location Used for the Measurement	Char*	Location relevant to the collection of the measurement. Examples: "ORAL" for temperature or "V1" for an ECG lead.
--METHOD	Method of Test or Examination	Char*	Method of the test or examination. Examples: SWAB, NEEDLE PUNCTURE, BRONCHIAL BRUSHING.
--BLFL	Baseline Flag	Char*	Indicator used to identify a baseline value. Should be Y or null.
--FAST	Fasting Status	Char*	Indicator used to identify fasting status such as Y, N, U, (for "Yes", "No", or "Unknown") or null if not relevant.
--DRVFL	Derived Flag	Char*	Used to indicate a derived record. Should be Y or null. Example: a record that represents the average of other records (such as a computed baseline).
--EVAL	Evaluator	Char*	Role of the person who provided the evaluation. Used only for results that are subjective (e.g., assigned by a person or a group). Should be Null for records that contain collected or derived data. Examples: INVESTIGATOR, ADJUDICATION COMMITTEE, VENDOR.
--TOX	Toxicity	Char*	Description of toxicity quantified by --TOXGR. Sponsor should specify which scale and version is used in sponsor comments. Example: NCI CTCAE Short Name.
--TOXGR	Toxicity Grade	Char*	Records toxicity grade value using a standard toxicity scale. Example: NCI CTCAE Grade.

* *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 or from an external published source such as MedDRA. See the Implementation Guide for details on which terminology is sponsor-controlled or standardized.*

2.2.4 Unique Identifiers for All Classes

Table 2.2.4: All Observations — Unique Identifiers

Variable Name	Variable Label	Type	Description
	Identifier Variables		
STUDYID	Study Identifier	Char	Unique identifier for a study within the submission.
DOMAIN	Domain Abbreviation	Char*	Two-character abbreviation for the domain most relevant to the observation. The Domain abbreviation is also used as a prefix for variables in SAS Transport datasets necessary to ensure that columns of data based upon the same variable (e.g., TEST) are not corrupted during merges.
USUBJID	Unique Subject Identifier	Char	Unique subject identifier within the submission.
--SEQ	Sequence Number	Num	Sequence number given to ensure uniqueness of records within a dataset for a subject.
--GRPID	Group ID	Char	Optional group identifier, used to link together a block of related records for a subject in a single domain.
--REFID	Reference ID	Char	Optional internal or external identifier. Examples: lab specimen ID, or UUID for an ECG waveform or a medical image.
--SPID	Sponsor ID	Char	Optional Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Line number on a Concomitant Medications page.

* *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 or from an external published source such as MedDRA. See the Implementation Guide for details on which terminology is sponsor-controlled or standardized.*

2.2.5 Timing Variables for All Classes

All of the following timing variables are permissible for use in any domain based on one of the three general classes.

Table 2.2.5: All Observations — Timing Variables

Variable Name	Variable Label	Type	Description
	Timing Variables		
VISITNUM	Visit Number	Num	Clinical encounter number. Numeric version of VISIT, used for sorting.
VISIT	Visit Name	Char	Protocol-defined description of a clinical encounter.
VISITDY	Planned Study Day of Visit	Num	Planned study day of VISIT.
TAETORD	Order of Element within Arm	Num	Number that gives the order of the element within the arm (see Section 3.2.2).
EPOCH	Epoch	Char*	Epoch associated with the start date/time of the observation, or the date of collection if start date/time is not collected. (see Section 3.2.2).
--DTC	Date/Time of Collection	Char	Collection date and time of an observation represented in IS0 8601 character format.
--STDTC	Start Date/Time of Observation	Char	Start date/time of an observation represented in IS0 8601 character format.
--ENDTC	End Date/Time of Observation	Char	End date/time of the observation represented in IS0 8601 character format.
--DY	Study Day of Visit/Collection/Exam	Num	Actual study day of Visit/Collection/Exam measured in integer days. Algorithm for calculations must be relative to the sponsor-defined RFSTDTC in Demographics. The formula should be consistent across the submission.
--STDY	Study Day of Start of Observation	Num	Start of observation expressed as study day relative to the sponsor-defined RFSTDTC.

Variable Name	Variable Label	Type	Description
--ENDY	Study Day of End of Observation	Num	End of observation expressed as study day relative to the sponsor-defined RFSTDTC.
--DUR	Duration	Char	Collected duration of an event, intervention, or finding represented in ISO 8601 character format. Used only if collected on the CRF and not derived from --STDTC and --ENDTC.
--TPT	Planned Time Point Name	Char	Text description of time when a measurement or observation should be taken within a visit, as defined in the protocol. This may be represented as an elapsed time relative to a fixed reference point, such as time of last dose. See --TPTNUM and --TPTREF.
--TPTNUM	Planned Time Point Number	Num	Numeric version of planned time point used in sorting.
--ELTM	Elapsed Time from Reference Point	Char	Elapsed time in ISO 8601 character format relative to a planned fixed reference (--TPTREF) 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, but an interval.
--TPTREF	Time Point Reference	Char	Description of the fixed reference point referred to by --ELTM, --TPTNUM, and --TPT. Examples: Previous Dose, Previous Meal.
--STRF	Start Relative to Reference Period	Char*	Identifies the start of the observation 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. --STRF 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 --STRF.
--ENRF	End Relative to Reference Period	Char*	Identifies the end of the observation 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. Observations that are ongoing at the end of the reference period should have a value of AFTER for this variable. -- ENRF 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 --ENRF.

* *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 or from an external published source such as MedDRA. See the Implementation Guide for details on which terminology is sponsor-controlled or standardized.*

2.2.6 The Demographics Clinical Domain

Each human clinical study should include one standardized set of observations in a specific structure, which is the Demographics domain described in Table 2.2.6 below. The Demographics domain describes the essential characteristics of the study subjects, and is used by reviewers for selecting populations for analysis. The Demographics domain, as with other datasets, includes Identifiers, a Topic variable, Timing variables, and Qualifiers, but unlike domains built from the general observation classes, it should not include any Identifiers or Timing variables other than those listed below. Demographics is the parent domain for all other observations for human clinical subjects, and should be identified with the domain code of "DM".

Table 2.2.6: Subject Demographics Domain Variables

Variable Name	Variable Label	Type	Description
	Identifier Variables		
STUDYID	Study Identifier	Char	Unique identifier for a study within the submission.
DOMAIN	Domain Abbreviation	Char	Two-character abbreviation for the domain which must be "DM".
USUBJID	Unique Subject Identifier	Char	Unique subject identifier within the submission.
	Topic Variable		
SUBJID	Subject Identifier for the Study	Char	Subject identifier used within the study as collected on CRF.
	Timing Variables		
RFSTDTC	Subject Reference Start Date/Time	Char	Reference Start Date/time for the subject in ISO 8601 character format. Usually equivalent to date/time when subject was first exposed to study treatment.
RFENDTC	Subject Reference End Date/Time	Char	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 exposure to study treatment.
	Qualifier Variables		
SITEID	Study Site Identifier	Char	Unique identifier for a study site within a submission.
INVID	Investigator Identifier	Char*	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.
INVNAM	Investigator Name	Char	Name of the investigator for a site.
BRTHDTC	Date/Time of Birth	Char	Date/time of birth of the subject.
AGE	Age in AGEU at Reference Start Date/Time	Num	Usually derived as RFSTDTC-BRTHDTC, but BRTHDTC may not be available in all cases (due to subject privacy concerns).
AGEU	Age Units	Char*	Age units in YEARS, MONTHS, or DAYS.
SEX	Sex	Char*	M, F, or U for Male, Female, Unknown.
RACE	Race	Char*	The race of the subject. In cases of mixed race, primary race should be listed. 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.
ETHNIC	Ethnicity	Char*	Ethnicity of the subject (HISPANIC or NON-HISPANIC).
ARMCD	Planned Arm Code	Char*	Short 8-character version of ARM, used for programming. (Formerly TRTCD)
ARM	Description of Planned Arm	Char	Name given to the arm a subject was assigned to (formerly TRTGRP).
COUNTRY	Country	Char*	Country of the investigational site at which the subject participated in the trial.
DMDTC	Date/time of Collection	Char	Date/time of collection of the demographic information in ISO 8601 character format.
DMDY	Study Day of Collection	Num	Study day of Visit/Collection measured in integer days. Algorithm for calculations must be relative to the sponsor-defined RFSTDTC. The formula should be consistent across the submission.

* *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 or from an external published source such as MedDRA. See the Implementation Guide for details on which terminology is sponsor-controlled or standardized.*

2.2.7 The Comments Domain

Comments are collected during the conduct of many studies. These are normally supplied by a principal investigator, but might also be collected from others such as central reviewers. When collected, comments should be submitted in a single Comments domain, which is defined in Table 2.2.7 below. Like Demographics, the Comments domain may not include any Identifiers or Timing variables other than those listed below.

Table 2.2.7: Comments Domain Variables

Variable	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Unique identifier for a study within the submission.
DOMAIN	Domain Abbreviation	Char*	Two-character abbreviation for the domain which must be "CO".
RDOMAIN	Related Domain Abbreviation	Char*	Domain Abbreviation of the parent record(s). Null for records collected on CRFs used to collect general comments.
USUBJID	Unique Subject Identifier	Char*	Unique Subject Identifier within the submission.
COSEQ	Sequence Number	Num	Sequence Number given to ensure uniqueness within a domain.
IDVAR	Identifier Variable Name	Char*	Identifier variable in the dataset that identifies the related record(s). Examples --SEQ, --GRPID. Null for comments collected on separate CRFs.
IDVARVAL	Identifier Variable	Char	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.
COREF	Comment Reference	Char	Indicates a reference such as a CRF or equivalent animal study page of the parent record(s) to which the comment refers, or the page on which the comment was collected. Page may be the page number (e.g., 650), the page name (e.g., VITALS), or a combination of both (e.g., 650-VITALS).
CODTC	Date/Time of Comment	Char	Date of comment on dedicated comment form, if collected. Represented in ISO 8601 character format. Should be null if this is a child record of another domain or if comment date was not collected.
COVAL	Comment	Char	The text of the comment. COVALUE cannot be null – a value is required for the record to be valid.
COEVAL	Evaluator	Char*	Used to describe the originator of the comment. Controlled terminology will consist of values such as CENTRAL REVIEWER, PRINCIPAL INVESTIGATOR.

* *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 or from an external published source such as MedDRA. See the Implementation Guide for details on which terminology is sponsor-controlled or standardized.*

3 The Trial Design Model

3.1 INTRODUCTION

The Trial Design Model defines a standard structure for representing the planned sequence of events and the treatment plan for the trial. The model provides a standard way to define the treatment groups and planned visits and assessments that will be experienced by trial subjects. The model also defines a way to capture 'actual' subject progress that can be compared against the plan.

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

- A description of what happens to the subject during the Element. This may be a description of study treatment a subject receives (e.g., Drug A, Drug B, Drug A + Drug B; Oral, IV; 5 mg, 10 mg, 20 mg) or a description of a time without treatment (e.g., Screening, Run-in, Wash-out, Rest, Follow-up).
- A definition of the start of the Element. This is the time point that marks the beginning of the Element. For treatment Elements, the start of the Element is usually defined as the start of treatment. For non-treatment Elements, the definition of the start of the Element may be in terms of the end of treatment in an earlier Element (e.g., "24 hours after last dose of study drug"). Other definitions of Element starts are also possible.
- A rule for ending the Element. The most common type of rule involves a planned duration for the Element (e.g., "continue until 2 weeks have passed since the start of the Element"). Examples of other rules:
  - "continue until 16 days have passed since the start of the Element and white blood count has recovered"
  - "continue until hospital discharge"
  - "continue until surgery is complete"
  - "continue until 24 weeks have passed since the start of treatment with study drug"

An Arm is a planned sequence of Elements, and is typically equivalent to a treatment group. Generally, each subject is assigned to an Arm, and the design of the study is reflected in the number and composition of the individual arms.

Some examples of Arm descriptions (showing Elements within the Arm are as follows):

- Simple Parallel-Group Design. In this example there are four arms (A, B, C, D) and six Elements (Screen, Drug A, Drug B, Drug A + Drug B, Placebo, Follow-up):

  Arm A: Screen, Drug A, Follow-up

  Arm B: Screen, Drug B, Follow-up

  Arm C: Screen, Drug A + Drug B, Follow-up

  Arm D: Screen, Placebo, Follow-up

- Crossover Design. In this example there are two arms (A, B) and five Elements (Baseline, IV drug, Wash-out, Oral drug, Follow-up):

  Arm A: Baseline, IV drug, Wash-out, Oral drug, Follow-up

  Arm B: Baseline, Oral drug, Wash-out, IV drug, Follow-up

From these examples, it is evident that an ordered list of Elements is needed to describe an Arm. Two other characteristics are included in the model:

- The points where arms of the Trial Design diverge or branch and how this occurs (e.g., via a randomization) are noted.
- There is a place to note transition rules more complex than the default rule, "go to the next Element in sequence."

Branches are assumed to take place between one Element and the next. In the two examples above, the different arms "branch" after the Screen (simple parallel design) or Baseline (crossover design) Element, before the first treatment Element. The example above does not specify that subjects are assigned to an Arm via a randomization, though that is the most common mechanism. Other examples of branching: further treatment depends on whether a subject is a responder or non-responder; in an escalating-dose cohort study, a subject is assigned a dose-level (Arm) depending on the responses of other subjects.

Some designs allow for some flexibility of Elements within an Arm. For example, cancer chemotherapy trials often allow subjects to skip some of the planned cycles of treatment if the disease progresses. Thus, an Arm might call for six cycles of treatment, as follows:

Arm A: Screen	Drug A, Rest Cycle 1	Drug A, Rest Cycle 2	Drug A, Rest Cycle 3	Drug A, Rest Cycle 4	Drug A, Rest Cycle 5	Drug A, Rest Cycle 6	Follow-up

However, the protocol would allow treatment to be cut short after 1, 2, 3, 4, or 5 cycles if the disease progresses. Thus, at the end of each Rest Element in the Arm described above, the transition rule is not "go to the next Element" but "if disease has progressed, go to Follow-up, otherwise go to the next Element"

Epochs

The term Epoch describes a phase or segment of a trial and is a useful concept to apply during study conduct -- especially while the trial is blinded. In parallel-design trials, the different trial arms are similar in that they have the same numbers of Elements and the same pattern of treatment and non-treatment Elements. As a result, one can divide the entire trial as a series of Epochs, which are often synonymous with periods, phases, or time segments of a trial. An Epoch is useful when it is meaningful to group data, across trial arms, by ordered Elements within Arm (e.g., by 3rd Element, by 4th Element).

The concept of trial Epochs is optional, since the Elements of the different arms of a trial do not necessarily line up. The Elements of blinded trials usually do fall into Epochs, since the patterns of time on and off treatment must be the same for all arms in order to maintain blinding.

The following diagram shows the crossover example described above, with its Arms, Elements, and Epochs.

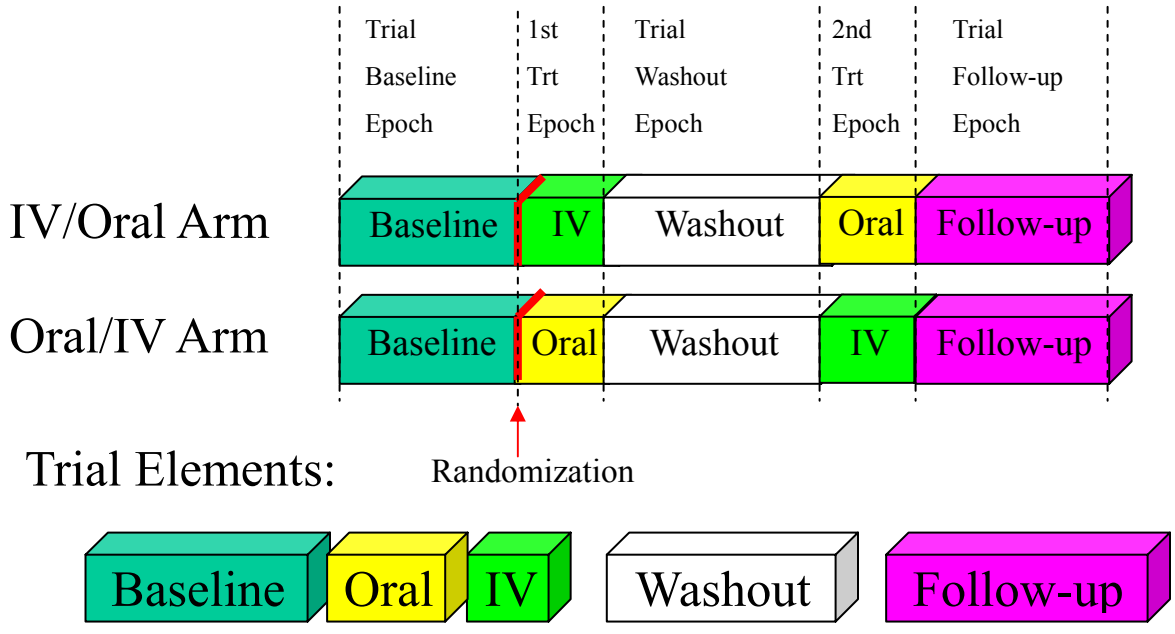

Figure: Arms, Elements and Epochs for a Crossover Trial

Visits

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. Visit information is generally recorded in a Clinical Data Management System but is not always evident from submission data. A visit has a start and an end, each described

with a rule. A visit need not be nested within a single Element. In other words, it may start in one Element and end in another.

In blinded trials, one will not know, during the blinded portion of the trial, which Element within an Arm a subject is in. For instance, in a simple parallel group design with one treatment and two arms (A and B), one does not know, during that treatment, whether a subject is in the Drug A Element, or the Drug B Element. Therefore, Visits are usually tied to Element Sequences (or Epoch in a blinded study), rather than Elements. Even though it is not known whether a subject is the Drug A or the Drug B Element, it *would* be known they are in the second Element of each Arm.

In most blinded trials, the timing of visits is the same for all subjects in all arms, regardless of the Arm to which they have been assigned. In these cases, the ARM variable in the Trial Visits dataset (described below) is not needed to describe the timing of visits, and is left blank. If the timing of visits depends on Arm, then the complete set of visits for each Arm should be represented in the Trial Visits dataset (Table 3.2.3).

3.2 PLANNED ELEMENTS, ARMS, AND VISITS

Under the model, planned information is presented in a series of three tables:

- The Trial Element table (Table 3.2.1) describes the Element Code (unique for each Element), the Element description, and the rules for starting and ending an Element. A rule could be expressed as pseudo code or as executable code for determining transitions from one Element to another.
- The Trial Arms table (Table 3.2.2) describes each planned Arm in the trial. An Arm is described as an ordered sequence of Elements, and the same Element may occur more than once in a given Arm. In order to accommodate complex Trial Designs, this table allows for rules for branching from one Element to another when a choice is available, and a rule for transitions to allow a subject to either skip ahead to another Element rather than proceed linearly.
- The Trial Visits table (Table 3.2.3) describes the planned order and number of visits in the study within each Arm. It describes the allowable or planned values for VISIT, VISITNUM and VISITDY in the trial (which are subsequently used as Timing Variables for the collected study data), and rules for starting and ending each visit. In most blinded trials, the timing of visits is the same for all subjects in all Arms.

These datasets are essential to determine whether data comparisons are feasible across different studies.

3.2.1 Trial Elements

Table 3.2.1: Trial Elements – All Observations, One Record per Trial Element

Variable	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Unique identifier for a study within the submission.
DOMAIN	Domain Abbreviation	Char*	Two-character abbreviation for the domain which must be TE.
ETCD	Element Code	Char*	Short 8-character name for ELEMENT, used for programming.
ELEMENT	Description of Element	Char*	The name of the Element.
TESTRL	Rule for Start of Element	Char	Expresses rule for beginning the Element.
TEENRL	Rule for End of Element	Char	Expresses rule for ending the Element.
TEDUR	Planned Duration of Element	Char	Planned Duration of Element in ISO 8601 format. For use when the rule for ending the Element is to end after a fixed duration.

* *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 or from an external published source such as MedDRA. See the Implementation Guide for details on which terminology is sponsor-controlled or standardized.*

3.2.2 Trial Arms

Table 3.2.2: Trial Arms – All Observations, One Record per Element per Arm

Variable Name	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Unique identifier for a study within the submission.
DOMAIN	Domain Abbreviation	Char*	Two-character abbreviation for the domain which must be TA.
ARMCD	Planned Arm Code	Char*	Short 8-character version of ARM, used for programming (formerly TRTCD).
ARM	Description of Planned	Char*	Name given to the Arm to which a subject was assigned (formerly TRTGRP).
TAETORD	Order of Element within Arm	Num	Number that gives the order of the Element within the Arm.
ETCD	Element Code	Char*	Short 8-character version of ELEMENT, used for programming.
ELEMENT	Description of Element	Char*	The name of the Element.
TABRANCH	Branch	Char	Condition subjects meet, at a "branch" in the Trial Design at the end of this Element, to be included in this Arm. Example: Randomization to Drug A.
TATRANS	Transition Rule	Char	If the Trial Design allows subjects 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. Example: Responders go to washout.
EPOCH	Trial Epoch	Char*	Name of the Trial Epoch with which this Element of the Arm is associated.

Note: The same Element may occur more than once within an Arm, but each occurrence would have a different value for TAETORD and EPOCH, and may have different values for TABRANCH and TATRANS.

3.2.3 Trial Visits

Table 3.2.3: Trial Visits– All Observations, One Record per Trial Visit

Variable Name	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Unique identifier for a study within the submission.
DOMAIN	Domain Abbreviation	Char*	Two-character abbreviation for the domain which must be TV.
VISIT	Visit Name	Char	1. Protocol-defined description of the clinical encounter. 2. May be used in addition to VISITNUM and/or VISITDY as a text description of the clinical encounter.
VISITNUM	Visit Sequence Number	Num	1. Clinical encounter number. 2. Numeric version of VISIT, can be used for sorting.
VISITDY	Planned Study Day of Visit	Num	1. Planned study day of VISIT. 2. Due to its sequential nature, can be used for sorting.
ARMCD	Planned Arm Code	Char*	Short 8-character version of ARM, used for programming (formerly TRTCD).
ARM	Description of Planned	Char*	Name given to the Arm to which a subject was assigned (formerly TRTGRP). If the timing of visits for a trial does not depend on which Arm a subject is in, then ARM should be left blank
TVSTRL	Visit Start Rule	Char	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. Example: When subject experiences symptoms.

Variable Name	Variable Label	Type	Description
TVENRL	Visit End Rule	Char	Rule describing when the visit ends, in relation to the sequence of Elements.

3.3 SUBJECT ELEMENT SEQUENCES AND VISITS

While the Trial Elements, Trial Arms, and Trial Visits tables describe the planned design of the study, it is also necessary to capture corresponding information to depict the actual treatments undergone and visits (or the values observed) for each subject. This information is described in two additional tables:

- The Subject Elements table (Table 3.3.1 below) describes the actual order of Elements followed by the subject, together with the start date/time and end date/time for each Element. Because actual data does not always follow the plan, the model allows for a description of an unplanned Element(s) for a subject.
- The Subject Visits table (Table 3.3.2 below) describes the actual start and end date/time for each visit of each individual subject. Because actual data does not always follow the plan, the model allows for a description of an unplanned Visit(s) for a subject.

Because the subject Arm is included in the Demographics dataset, it does not require a separate Trial Design table.

By including these tables, it is possible to precisely track the progression of a subject through a given study.

3.3.1 Subject Elements Table

Table 3.3.1: Subject Elements - All Observations, One Record Subject Element per Subject

Variable Name	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Unique identifier for a study within the submission.
DOMAIN	Domain Abbreviation	Char*	Two-character abbreviation for the domain which must be SE.
USUBJID	Unique Subject Identifier	Char	Unique identifier within the submission.
ETCD	Subject Element Code	Char*	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.
ELEMENT	Description of Subject Element	Char*	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 blank.
SESTDTC	Start Date/Time of Element	Char	Start date/time for an Element for each subject, represented in ISO 8601 character format.
SEENDTC	End Date/Time of Element	Char	End date/time of an Element for each subject, represented in ISO 8601 character format.
SEUPDES	Description of Unplanned Element	Char	Description of what happened to the subject during an unplanned Element. Used only if ETCD has the value of UNPLAN.

3.3.2 Subject Visits Table

Table 3.3.2: Subject Visits– All Observations, One Record per Subject Visit, per Subject

Variable Name	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Unique identifier for a study within the submission.
DOMAIN	Domain Abbreviation	Char*	Two-character abbreviation for the domain which must be SV.
USUBJID	Unique Subject Identifier	Char	Unique identifier within the submission.

Variable Name	Variable Label	Type	Description
VISIT	Visit Name	Char	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.
VISITNUM	Visit Number	Num	1. Clinical encounter number. (Decimal numbering may be useful for inserting unplanned visits.) 2. Numeric version of VISIT, used for sorting.
VISITDY	Planned Study Day of Visit	Num	Planned study day of VISIT as a sequential number, used for sorting.
SVSTDTC	Start Date/Time of Visit	Char	Start date/time for a visit, represented in ISO 8601 character format.
SVENDTC	End Date/Time of Visit	Char	End date/time of a visit, represented in ISO 8601 character format.
SVUPDES	Description of Unplanned Visit	Char	Description of what happened to the subject during an unplanned visit.

3.4 TRIAL INCLUSION/EXCLUSION CRITERIA

The Trial Inclusion Exclusion Domain (TI) is not subject-oriented. It contains one record for each of 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 subject-level IE Domain contains records only for inclusion and exclusion criteria that an individual subject did not meet.

3.4.1 Trial Inclusion/Exclusion Table

Table 3.4.1: Trial Inclusion/Exclusion– All Observations, One Record per Trial Inclusion or Exclusion Criterion

Variable Name	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Unique identifier for a study within the submission.
DOMAIN	Domain Abbreviation	Char*	Two-character abbreviation for the domain which must be TI.
IETESTCD	Inclusion/Exclusion Short Name	Char*	Short name for the Inclusion/Exclusion Criterion
IETEST	Inclusion/Exclusion Criterion	Char*	Full text of the Inclusion/Exclusion Criterion
IECAT	Inclusion//Exclusion Category	Char*	Category of the Inclusion/Exclusion Criterion: INCLUSION, EXCLUSION.
TIRL	Trial Inclusion/Exclusion Rule	Char	Rule that expresses the inclusion/exclusion criterion in computer-executable form.

* *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 or from an external published source such as MedDRA. See the Implementation Guide for details on which terminology is sponsor-controlled or standardized.*

4 Representing Relationships among Datasets and Records

There are many occasions when it is necessary or desirable to represent relationships among datasets or records. The SDTM identifies five distinct types of relationships:

- Section 4.1 describes a relationship between a group of records in the same domain.
- Section 4.2 describes a relationship between independent records (whether in the same or separate domains), such as a concomitant medication taken to treat an adverse event.
- Section 4.3 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 4.4 describes a dependent relationship between a non-standard variable and a parent record (or records) in a domain, which provides a way of including additional data captured in variables that are not presently represented in the general class models.
- Section 4.5 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.

The implementation guides define specific details and examples for each of these relationships.

4.1 RELATING GROUPS OF RECORDS IN 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 the same 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 not only meaningful within the domain, but also in RELREC, SUPPQUAL, and COMMENTS 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.

4.2 RELATING PEER RECORDS IN SEPARATE DOMAINS

The Related Records (RELREC) dataset is used to identify 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 4.1) 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.

4.2.1 Related Records Dataset

Table 4.2.1: RELREC Dataset

Variable	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Study Identifier of the domain record(s).
RDOMAIN	Related Domain Abbreviation	Char*	Domain Abbreviation of the domain record(s).
USUBJID	Unique Subject Identifier	Char	Unique Subject Identifier of the domain record(s).
IDVAR	Identifier Variable	Char*	Identifier variable in the dataset that identifies the related record(s). Examples: --SEQ, --GRPID.
IDVARVAL	Identifier Variable Value	Char	Value of identifier variable of the parent record(s). Used only when individual records are being related.
RELTYPE	Relationship Type	Char*	Identifies the hierarchical level of the records in the relationship. Values must be either ONE or MANY.
RELID	Relationship Identifier	Char	Unique value within a 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.

4.3 RELATING DEPENDENT RECORDS IN DIFFERENT DATASETS

The Related Records (RELREC) dataset can also be used to identify relationships between datasets that have the same topicality (and thus may be considered to belong to the same logical domain), such as lesion measurements associated with the identification and classification of a lesion. 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.

4.4 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, which 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 variable being added (QNAM), the label for the value (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 INVESTIGATOR or SPONSOR). Details on expected SUPPQUAL values and appropriate controlled terminology, such as the population flags, Intent To Treat, Per Protocol, and Safety are included in the implementation guides.

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 subjects or subject data according to their evaluability for efficacy analysis. Since it is possible that these attributions may vary by study, SUPPQUAL provides a mechanism for incorporating as many attributions as are necessary. However, it is recognized that sponsors may also need to use SUPPQUAL to capture additional non-standard variables that the sponsor needs to submit, but which cannot be represented in the general classes. Details on expected SUPPQUAL values and appropriate controlled terminology, such as the population flags, Intent To Treat, Per Protocol, and Safety are included in the implementation guides.

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 values (SUPPQUAL records) related to multiple Domain records that could be grouped, such as relating an attribution to a group of ECG measurements.

4.4.1 Supplemental Qualifiers Dataset

Table 4.4.1: SUPPQUAL Dataset - All Variables

Variable	Variable Label	Type	CDISC Notes
STUDYID	Study Identifier	Char	Study Identifier of the Parent record(s).
RDOMAIN	Related Domain Abbreviation	Char*	Domain Abbreviation of the Parent record(s).
USUBJID	Unique Subject Identifier	Char	Unique Subject Identifier of the Parent record(s).
IDVAR	Identifier Variable	Char*	Identifier variable in the dataset that identifies the related record(s). Examples: --SEQ, --GRPID.
IDVARVAL	Identifier Variable Value	Char	Value of identifier variable of the parent record(s).
QNAM	Variable Name	Char*	The short name of the variable test, examination, or judgment. This variable contains text values that are 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.
QLABEL	Variable Label	Char	This is the long name or label associated with QNAM. 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 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, PRINCIPAL INVESTIGATOR.

4.5 RELATING COMMENTS TO A PARENT DOMAIN

The COMMENTS dataset (described in Section 2.2.7) is used to capture unstructured text comments. Comments can be collected on separate CRFs or as part of CRFs with their own topicality, such as adverse events or concomitant medications. Comments may be related to a Subject, a Domain for a Subject, or to specific Parent records in a domain for a subject. The COMMENTS dataset is structured similarly to the SUPPQUAL dataset, in that it uses the same set of keys to identify related records.

Again, use of the optional grouping identifier variable --GRPID in the domains can be a more efficient method of representing relationships in COMMENTS when comments have relationships to multiple domain records that could be grouped, such a comment that applies to a "group" of concomitant medications.

Details on structure and use of the COMMENTS dataset are included in the implementation guides.

5 Using the Model for Regulatory Submissions

The SDTM has been designed to accommodate the broadest range of human and animal study data in a standardized manner. This document has described the basic concepts and general structures of the model. The implementation guides for the CDISC Submission Data Standards and SEND include specific recommendations for numerous domains of data commonly collected in human and animal studies, respectively, identifying which variables from a general observation class may apply in each. These implementation guides also describe basic assumptions and business rules, and provide numerous examples for mapping data to the standard format. Any sponsor wishing to submit data in the standard formats should first consult the implementation guides before preparing a regulatory submission based on the SDTM.