

# Study Data Tabulation Model

Prepared by the

**CDISC Submission Data Standards Team**

---

Principal Editor:

Wayne Kubick

Principal Contributors:

Fred Wood, Diane Wold, Tom Guinter, Mary Lenzen, Gary Walker, Gail Stoner, Christine Tolk, CDISC SDS Team

## Notes to Readers

This is Version 1.2 of the Study Data Tabulation Model Document, posted for comment by the CDISC Submissions Data Standards team. This document includes additional variables related to human clinical trials and other corrections and clarifications to the text. A full description of all changes is provided in Section [6.3](#).

## Revision History

Date	Version	Summary of Changes
2004-06-25	Version 1.0	<ul style="list-style-type: none"><li>• First released version reflecting all changes identified during comment periods.</li></ul>
2005-04-28	Version 1.1 Final	<ul style="list-style-type: none"><li>• Final version incorporating minor corrections to address comments submitted during public review period.</li></ul>
2007-07-10	Version 1.2 Draft	<ul style="list-style-type: none"><li>• Draft for comment.</li></ul>

## CONTENTS

<b>1 Introduction</b> .....	<b>3</b>
<b>1.1 Purpose</b> .....	<b>3</b>
<b>1.2 Relationship to Prior CDISC Models</b> .....	<b>3</b>
<b>1.3 Significant Changes from the Prior Version</b> .....	<b>4</b>
<b>1.4 Relationship to HL7 Models</b> .....	<b>4</b>
<b>2 Model Fundamentals</b> .....	<b>5</b>
<b>2.1 Model Concepts and Terms</b> .....	<b>5</b>
<b>2.2 The General Observation Classes</b> .....	<b>6</b>
2.2.1 The Interventions Observation Class .....	7
2.2.2 The Events Observation Class .....	8
2.2.3 The Findings Observation Class .....	9
2.2.4 Identifiers for All Classes .....	10
2.2.5 Timing Variables for All Classes .....	11
2.2.6 The Demographics Clinical Domain .....	12
2.2.7 The Comments Domain .....	13
2.2.8 The Subject Elements Table.....	14
2.2.9 The Subject Visits Table .....	14
<b>3 The Trial Design Model</b> .....	<b>15</b>
<b>3.1 Introduction</b> .....	<b>15</b>
<b>3.2 Planned Elements, Arms, and Visits</b> .....	<b>17</b>
3.2.1 Trial Elements .....	17
3.2.2 Trial Arms .....	18
3.2.3 Trial Visits.....	18
<b>3.3 Trial Inclusion/Exclusion Criteria</b> .....	<b>19</b>
3.3.1 Trial Inclusion/Exclusion Table .....	19
<b>3.4 Trial Summary Information</b> .....	<b>19</b>
3.4.1 Trial Summary Table .....	19
<b>4 Representing Relationships Among Datasets and Records</b> .....	<b>20</b>
<b>4.1 Relating Groups of Records Within a Domain</b> .....	<b>20</b>
<b>4.2 Relating Records in Separate Datasets</b> .....	<b>20</b>
4.2.1 Related Records Dataset .....	21
<b>4.3 Relating Datasets</b> .....	<b>21</b>
<b>4.4 Relating Non-Standard Variable Values to a Parent Domain</b> .....	<b>21</b>
4.4.1 Supplemental Qualifiers Dataset .....	22
<b>4.5 Relating Comments to a Parent Domain</b> .....	<b>22</b>
<b>5 Using the Model for Regulatory Submissions</b> .....	<b>23</b>
<b>6 SDTM Version history</b> .....	<b>24</b>
<b>6.1 Changes from CDISC Submission Data Domain Models V3 to SDTM V1.0</b> .....	<b>24</b>
<b>6.2 Changes from SDTM V1.0 to SDTM V1.1</b> .....	<b>25</b>
<b>6.3 Changes from SDTM V1.1 to SDTM V1.2</b> .....	<b>26</b>

# 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 document, which will supersede all prior versions, includes numerous changes from the prior Version 1.1 which are described in Appendix 6.3.

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 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 originally developed for the CDISC SDS 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 MODELS

As stated previously, this document is a descendant of what was formerly known in prior versions as the CDISC Submission Data Standards or Submission Domain Models. While Version 1.0 SDTM was designated as the first implementation-ready version for clinical studies involving human drug products, future improvements and enhancements have been incorporated in subsequent versions to support a broader range of regulated products, including the needs of non-clinical animal toxicity studies. Future 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. 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 THE PRIOR VERSION

The SDTM has been designed for backward compatibility; datasets prepared with V1.2 should be fully compatible with prior versions. In most cases, this means that later versions may add new variables or correct textual errors, but not eliminate variables or structures incorporated in prior versions. V1.2 has been expanded to include the following new variables: --PRESF, --VAMT, --VAMTU, --OBJ, --STRTP, --ENRTP, --STTP, --ENTP, TIVERS and TSGRPID. V1.2 also includes numerous text corrections, clarifications and some reordering, including the transfer of domains for subject elements and subject visits from Section 3 to Section 2.

Details of significant changes are described in Section 6.

## 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 tabulations 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 Reference Information Model (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 and submitted to regulatory authorities.

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 describes the type of information conveyed by the variable about each distinct observation and how it can be used. SDTM variables can be classified into five major roles:

- *Identifier* variables, such as those that 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).
- *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).
- *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 or equivalent coded value 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 indication for which a drug was taken or a body position for a vital sign measurement).
- *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. A domain is defined as a collection of logically related observations with a common topic. 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 SDTM Implementation Guide for Human Clinical Trials 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 and values used in the dataset. The metadata are described in a data definition document named "Define" that is submitted along with the data. The Define data definition 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); some variables may be subject to controlled terminologies assigned in implementation guides
- 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., a reference to a CRF or derivation algorithm)
- 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 — such as age). 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 version 5 transport files; 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 Define data definition document, but new variables must not be added, and existing variables must 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, therapeutic and other treatments that are administered to the subject (with some actual or expected physiological effect) either as specified by the study protocol (e.g., “exposure”), coincident with the study assessment period (e.g., “concomitant medications”), or other substances self-administered by the subject (such as alcohol, tobacco, or caffeine).
- The *Events* class, 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 tests or questions such as 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 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 Identifier or Timing variable is permissible for use in any submission dataset based on a general observation class.

In the tables in this section, 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 variable prefix to minimize the risk of difficulty when merging/joining domains for reporting purposes.

In addition to the three general observation classes, a submission will generally 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
<b>Topic Variable</b>			
--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 interval for the observation.
<b>Qualifier Variables</b>			
--MODIFY	Modified Treatment Name	Char	If the value for --TRT is modified for coding purposes, 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.
--CAT	Category	Char	Used to define a category of topic variables.
--SCAT	Subcategory	Char	Used to define a further categorization level for a group of --CAT values.
--PRESP	Pre-specified	Char	Used when a specific intervention is pre-specified on a CRF. Values should be "Y" or null.
--OCCUR	Occurrence	Char	Used only when the occurrence of specific interventions is solicited.
--STAT	Completion Status	Char	Used to indicate that an assessment about an intervention was not performed. Should be null or have a value of NOT DONE.
--REASND	Reason Not Done	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, used with a coding dictionary.
--CLASCD	Class Code	Char	Used to represent dictionary codes for --CLAS.
--DOSE	Dose	Num	Amount of --TRT given. Not used when --DOSTXT is used.
--DOSTXT	Dose Description	Char	Dosing information collected in text form. Examples: <1 per day, 200-400. Not used when --DOSE is used.
--DOSU	Dose Units	Char	Units for --DOSE. Examples: ng, mg, mg/kg. Can be used with --DOSE, --DOSTOT, --DOSTXT.
--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. To be used in addition to and not in place of --DOSE.
--DOSRGM	Intended Dose Regimen	Char	Text description of the (intended) schedule or regimen for the Intervention. Examples: TWO WEEKS ON, TWO WEEKS OFF. Generally at a less granular level than --FRQ.
--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	Anatomical location of an intervention, such as an injection site. Example: RIGHT ARM for an injection.
--TRTV	Treatment Vehicle	Char	Vehicle for administration of treatment, such as a liquid in which the treatment drug is dissolved. Example: SALINE.
--VAMT	Treatment Vehicle Amount	Num	Amount of the treatment vehicle administered or given.
--VAMTU	Treatment Vehicle Amount Units	Char	Units for the treatment vehicle. Examples: mL, puffs.
--ADJ	Reason for Dose Adjustment	Char	Used only when dose is adjusted.

## 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
	<b>Topic Variable</b>		
--TERM	Reported Term	Char	Topic variable for an event observation, which is the verbatim or pre-specified name of the event.
	<b>Qualifier Variables</b>		
--MODIFY	Modified Reported Term	Char	If the value for --TERM is modified for coding purposes, 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).
--CAT	Category	Char	Used to define a category of topic variables.
--SCAT	Subcategory	Char	Used to define a further categorization level for a group of --CAT values.
--PRESP	Pre-specified	Char	Used to indicate whether the event described by --TERM was pre-specified on a CRF. Value is Y for pre-specified events, null for spontaneously reported events.
--OCCUR	Occurrence	Char	Used to record whether an event occurred when information about the occurrence of a specific event is solicited.
--STAT	Completion 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 Not Done	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 system organ class that a standard hierarchy such as MedDRA associated with an event.
--LOC	Location	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? Valid values are 'Y' and 'N'.
--ACN	Action Taken with Study Treatment	Char	Describes changes made to the study treatment as a result of the event. Examples: DOSE INCREASED, DOSE NOT CHANGED.
--ACNOTH	Other Action Taken	Char	Describes other actions taken as a result of the event that are unrelated to dose adjustments of study treatment.
--REL	Causality	Char	An opinion as to whether the event may have been due to the study treatment. Examples: NOT RELATED or POSSIBLY.
--RELNST	Relationship to Non-Study Treatment	Char	An opinion as to whether the event may have been due to a treatment other than study drug. Example: "More likely related to aspirin use."
--PATT	Pattern of Event	Char	Used to indicate the pattern of the event over time. Examples: INTERMITTENT, CONTINUOUS.
--OUT	Outcome of Event	Char	Description of the outcome of an event. Examples: RECOVERED/RESOLVED, FATAL.
--SCAN	Involves Cancer	Char	Was the event associated with the development of cancer? Valid values are 'Y' and 'N'.
--SCONG	Congenital Anomaly or Birth Defect	Char	Was the event associated with congenital anomaly or birth defect? Valid values are 'Y' and 'N'.
--SDISAB	Persist or Signif Disability/Incapacity	Char	Did the event result in persistent or significant disability/incapacity? Valid values are 'Y' and 'N'.
--SDTH	Results in Death	Char	Did the event result in death? Valid values are 'Y' and 'N'.
--SHOSP	Requires or Prolongs Hospitalization	Char	Did the event require or prolong hospitalization? Valid values are 'Y' and 'N'.
--SLIFE	Is Life Threatening	Char	Was the event life threatening? Valid values are 'Y' and 'N'.
--SOD	Occurred with Overdose	Char	Did the event occur with an overdose? Valid values are 'Y' and 'N'.
--SMIE	Other Medically Important Serious Event	Char	Do additional categories for seriousness apply? Valid values are 'Y' and 'N'.
--CONTRT	Concomitant or Additional Trtmt Given	Char	Was another treatment given because of the occurrence of the event? Valid values are 'Y' and 'N'.
--TOXGR	Toxicity Grade	Num	Records toxicity grade using a standard toxicity scale (such as the NCI CTCAE). Sponsor should specify which scale and version is used in the Sponsor Comments column of the Define data definition document.

## 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
	<b>Topic Variable</b>		
--TESTCD	Measurement, Test or Exam Short Name	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.
	<b>Qualifier Variables</b>		
--TEST	Name of Measurement, Test or Examination	Char	Long name For --TESTCD.. Examples: Platelet Count, Systolic Blood Pressure, PR Interval, Eye Examination.
--OBJ	Object of Measurement	Char	Used in domains modeled as Findings about Events or Findings about Interventions. Describes the event or intervention whose property is being measured in --TEST. Example: an event of vomiting which has findings, where OBJ = 'VOMIT' and the volume of VOMIT is being measured with a --TESTCD value of VOLUME.
--CAT	Category	Char	Used to define a category of topic variables. Examples: HEMATOLOGY, URINALYSIS, CHEMISTRY, HAMILTON DEPRESSION SCALE, SF36.
--SCAT	Subcategory	Char	Used to define a further categorization level for a group of --CAT values. Example: DIFFERENTIAL.
--POS	Position of Subject During Observation	Char	Position of the subject during a measurement or examination. Examples: SUPINE, STANDING, SITTING.
--ORRES	Result or Finding in Original Units	Char	Result of the measurement or finding as originally received or collected. Examples: 120, <1.
--ORRESU	Original Units	Char	Unit for --ORRES. Examples: in, ft, lb, g, L, g/L.
--ORNRL0	Normal Range Lower Limit-Original Units	Char	Lower end of normal range or reference range for results stored in --ORRES.
--ORNRHI	Normal Range Upper Limit-Original Units	Char	Upper end of normal range or reference range for results stored in --ORRES.
--MODIFY	Modified Term	Char	If the value of --ORRES is modified for coding purposes, then the modified text is placed here.
--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".
--STRESN	Numeric Result or Finding in Std 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 and --STRESN. Example: mmol/L.
--BODSYS	Body System or Organ Class	Char	Body System or Organ Class that is involved for a finding from the standard hierarchy for dictionary-coded results. Example: the Primary SOC in MedDRA.
--STNRLO	Normal Range Lower Limit-Standard Units	Num	Lower end of normal range or reference range for results stored in --STRESN.
--STNRHI	Normal Range Upper Limit-Standard Units	Num	Upper end of normal range or reference range for results stored in --STRESN.
--STNRC	Reference Range for Char Rslt-Std Units	Char	For normal range or reference range values for results stored in --STRESC that are character in ordinal or categorical scale. Example: Negative to Trace.
--NRIND	Normal/Reference Range Indicator	Char	Used to indicate the value is outside the normal range or reference range. May be defined by --ORNRL0 and --ORNRHI or other objective criteria. Examples: Y, N or HIGH, LOW.
--RESCAT	Result Category	Char	Used to categorize the result of a finding. Example: MALIGNANT or BENIGN for tumor findings.
--STAT	Completion 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 Not Done	Char	Reason not done. Used in conjunction with --STAT when value is NOT DONE.

Variable Name	Variable Label	Type	Description
--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. Example: RECTAL for temperature, LEFT ARM for blood pressure or V1 for an ECG lead.
--METHOD	Method of Test or Examination	Char	Method of the test or examination. Examples: EIA (Enzyme ImmunoAssay), ELECTROPHORESIS, DIPSTICK.
--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. Valid values include Y, N, U or null if not relevant.
--DRVFL	Derived Flag	Char	Used to indicate a derived record (e.g., a record that represents the average of other records such as a computed baseline). Should be Y or null.
--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). Examples: INVESTIGATOR, ADJUDICATION COMMITTEE, VENDOR.
--TOX	Toxicity	Char	Description of toxicity quantified by --TOXGR such as NCI CTCAE Short Name. Examples: HYPERCALCEMIA, HYPOCALCEMIA.
--TOXGR	Toxicity Grade	Num	Records toxicity grade using a standard toxicity scale (such as the NCI CTCAE). Sponsor should specify which scale and version is used in the Sponsor Comments column of the Define data definition document.
--SEV	Severity	Char	Describes the severity or intensity of a particular finding. Examples: MILD, MODERATE, SEVERE.
--DTHREL	Relationship to Death	Char	Describes the relationship of a particular finding to the death of a subject.
--LLOQ	Lower Limit of Quantitation	Num	Indicates the lower limit of quantitation for an assay. Units will be those used for --STRESU.

## 2.2.4 Identifiers for All Classes

All of the following Identifier variables are available for use in any domain based on one of the three general classes, STUDYID, DOMAIN, USUBJID, and --SEQ are required in all such domains.

**Table 2.2.4: All Observation Classes —Identifiers**

Variable Name	Variable Label	Type	Description
<b>Identifier Variables</b>			
STUDYID	Study Identifier	Char	Unique identifier for a study.
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 to ensure uniqueness when datasets are merged.
USUBJID	Unique Subject Identifier	Char	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
--SEQ	Sequence Number	Num	Sequence number given to ensure uniqueness of records within a dataset for a subject (or within a study, in the case of the Trial Summary domain).
--GRPID	Group ID	Char	Optional group identifier, used to link together a block of related records within a subject in a domain. --GRPID is also used to link together a block of related records in the Trial Summary dataset.
--REFID	Reference ID	Char	Optional internal or external identifier such as lab specimen ID, or UUID for an ECG waveform or a medical image.
--SPID	Sponsor ID	Char	Sponsor-defined reference number. Example: pre-printed line identifier on a Concomitant Medications page.



## 2.2.6 The Demographics Clinical Domain

Each human clinical study must include one standardized set of observations in a specific structure, which is the Demographics domain described in [Table 2.2.6](#). 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. 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
<b>Identifier Variables</b>			
STUDYID	Study Identifier	Char	Unique identifier for a study.
DOMAIN	Domain Abbreviation	Char	Two-character abbreviation for the domain, which must be DM.
USUBJID	Unique Subject Identifier	Char	Unique subject identifier.
<b>Topic Variable</b>			
SUBJID	Subject Identifier for the Study	Char	Subject identifier used within the study. Often the ID of the subject as collected on a CRF.
<b>Qualifier Variables</b>			
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.
SITEID	Study Site Identifier	Char	Unique identifier for a study site.
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 in ISO 8601 character format.
AGE	Age	Num	Age expressed in AGEU. 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	F, M, U for Female, Male, Unknown.
RACE	Race	Char	The race of the subject.
ETHNIC	Ethnicity	Char	Ethnicity of the subject. Examples: HISPANIC OR LATINO, NOT HISPANIC OR LATINO.
ARMCD	Planned Arm Code	Char	Eight-character name of ARM.
ARM	Description of Planned Arm	Char	Name of the Arm to which the subject was assigned.
COUNTRY	Country	Char	Country of the investigational site at which the subject participated in the trial in ISO 3166 3-character format.
<b>Timing Variables</b>			
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 Collection measured as integer days.

## 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](#). The Comments domain should not include any additional Identifiers other than those listed, but may include Timing variables when appropriate.

**Table 2.2.7: Comments Domain Variables**

Variable Name	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Unique identifier for a study.
DOMAIN	Domain Abbreviation	Char	Two-character abbreviation for the domain, which must be CO.
USUBJID	Unique Subject Identifier	Char	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
COSEQ	Sequence Number	Num	Sequence Number given to ensure uniqueness within a domain.
COVAL	Comment	Topic	The text of the comment. COVAL cannot be null. A value is required for the record to be valid.
RDOMAIN	Related Domain Abbreviation	Char	Domain Abbreviation of the parent record(s). Null for records collected on general comments or additional information CRF page.
IDVAR	Identifying Variable	Char	Identifying variable in the parent dataset that identifies the record(s) to which the comment applies. Examples AESEQ, CMGRPID. Null for comments collected on separate CRFs. Used only when individual comments are related to domain records.
IDVARVAL	Identifying Variable Value	Char	Value of identifying variable of the parent record(s). Null for comments collected on separate CRFs.
COREF	Comment Reference	Char	Reference to the parent record(s) to which the comment refers. May be the page number (e.g., 650) or the module name (e.g. DEMOG), or a combination of identification that identifies the reference (e.g., 650-VITALS).
COEVAL	Evaluator	Char	Used to describe the originator of the comment. Examples: CENTRAL REVIEWER, PRINCIPAL INVESTIGATOR.
CODTC	Date/Time of Comment	Char	Date/time of comment on dedicated comment form, if collected. Represented in ISO 8601 character format.

## 2.2.8 The Subject Elements Table

The Subject Elements table describes the actual order of Elements that were traversed by the subject, together with the start date/time and end date/time for each Element. These correspond to the planned Elements described in the Trial Elements (Section 3.2.1) of the Trial Design Model. Because actual data does not always follow the plan, the model allows for descriptions of an unplanned Element for subjects.

**Table 2.2.8: Subject Elements - All Observations, One Record per Actual Element per Subject**

Variable Name	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Unique identifier for a study.
DOMAIN	Domain Abbreviation	Char	Two-character abbreviation for the domain, which must be SE.
USUBJID	Unique Subject Identifier	Char	Unique subject identifier within the submission.
SESEQ	Sequence Number	Num	Sequence number given to ensure uniqueness within dataset. Should be assigned to be in a consistent chronological order.
ETCD	Element Code	Char	Short 8-character name for 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 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 null.
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.

## 2.2.9 The Subject Visits Table

The Subject Visits table describes the actual start and end date/time for each visit of each individual subject. These correspond to the planned visits described in the Trial Design Model Trial Visits table (see Section 3.2.3). Because actual data does not always follow the plan, the model allows for descriptions of unplanned visits for subjects.

**Table 2.2.9: 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.
DOMAIN	Domain Abbreviation	Char	Two-character abbreviation for the domain, which must be SV.
USUBJID	Unique Subject Identifier	Char	Unique subject identifier within the submission.
VISIT	Visit Name	Char	1. Protocol-defined description of clinical encounter or description of unplanned visit. 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 subject's visit, represented in ISO 8601 character format.
SVENDTC	End Date/Time of Visit	Char	End date/time of a subject's visit, represented in ISO 8601 character format.
SVUPDES	Description of Unplanned Visit	Char	Description of what happened to the subject during an unplanned visit. Null for protocol-defined visits.

# 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. The variables corresponding to these concepts are used in many domains based on the SDTM.

### Elements

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”.

### Arms

An Arm is a planned sequence of Elements, and is typically equivalent to a planned 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 previous two examples, 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 Arms. 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 3<sup>rd</sup> Element, by 4<sup>th</sup> 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.

Figure 3-1 diagrams the crossover example described above, with its Arms, Elements, and Epochs.

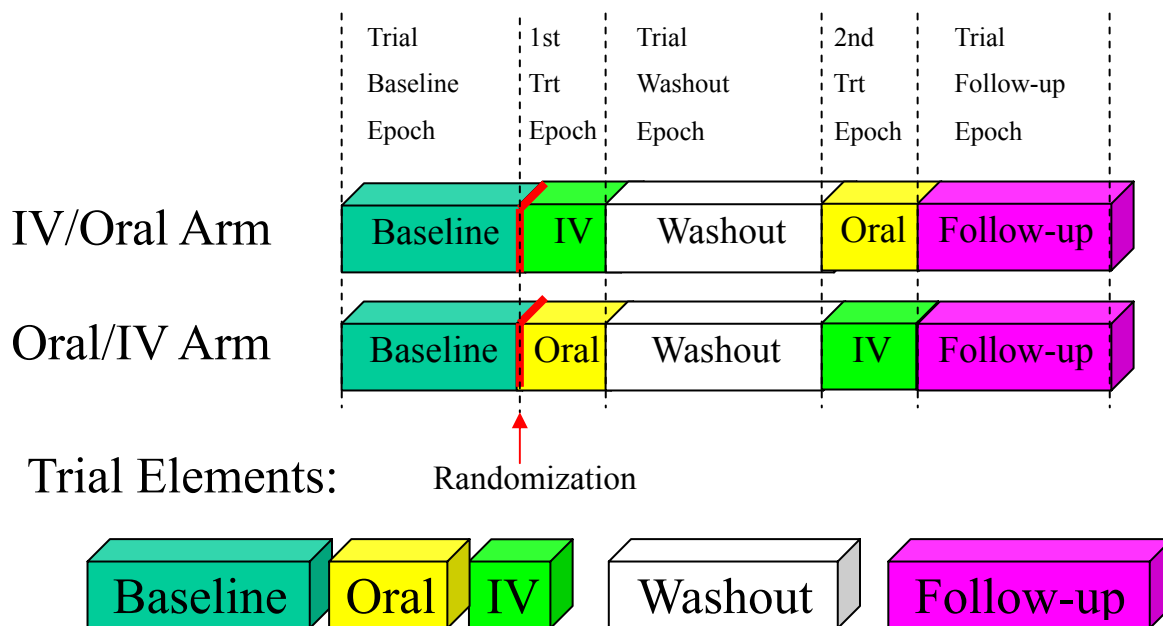


Figure 3-1: Arms, Elements and Epochs for a Crossover Trial

## 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 Epochs 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 parallel-design, blinded trials, the timing of visits is the same for all subjects in all Arms. In these cases, the ARM variable in the Trial Visits dataset (see [Table 3.2.3](#)) 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 Elements 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. In the case when visits vary for each Arm, there would be a separate record per visit. 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 Name	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Unique identifier for a study.
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. Used when the rule for ending the Element is to end after a fixed duration.

### 3.2.2 Trial Arms

**Table 3.2.2: Trial Arms — All Observations, One Record per Planned Element per Arm**

Variable Name	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Unique identifier for a study.
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.
ARM	Description of Planned Arm	Char	Name given to Arm or treatment group.
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 and sorting.
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	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 Planned Trial Visit**

Variable Name	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Unique identifier for a study.
DOMAIN	Domain Abbreviation	Char	Two-character abbreviation for the domain, which must be TV.
VISITNUM	Visit Sequence Number	Num	1. Clinical encounter number. 2. Numeric version of VISIT, can be used for sorting.
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.
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	1 Short name of ARM, used for programming and sorting. 2. If the timing of visits for a trial does not depend on which ARM a subject is in, then ARMCD should be null.
ARM	Description of Planned Arm	Char	1. Name given to an Arm or Treatment Group. 2. If the timing of Visits for a trial does not depend on which Arm a subject is in, then Arm should be left blank.
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.
TVENRL	Visit End Rule	Char	Rule describing when the visit ends, in relation to the sequence of Elements.

### 3.3 TRIAL INCLUSION/EXCLUSION CRITERIA

The Trial Inclusion Exclusion Domain (TI) contains one record for each of the inclusion and exclusion criteria for the trial.

#### 3.3.1 Trial Inclusion/Exclusion Table

Table 3.3.1: 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.
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	Used for categorization of the Inclusion/Exclusion Criterion: INCLUSION, EXCLUSION.
TIRL	Trial Inclusion/Exclusion Rule	Char	Rule that expresses the inclusion/exclusion criterion in computer-executable form.
TIVERS	Protocol Criteria Versions	Char	The number of this version of the Inclusion/Exclusion criteria. May be omitted if there is only one version.

### 3.4 TRIAL SUMMARY INFORMATION

The Trial Summary Information Domain (TS) contains one record for each trial summary characteristic. Trial Summary is used to record basic information about the trial, such as trial phase, protocol title and design objectives.

#### 3.4.1 Trial Summary Table

Table 3.4.1: Trial Summary — All Observations, One Record per Trial Summary Parameter

Variable Name	Variable Label	Type	Description
STUDYID	Study Identifier	Char	Unique identifier for a study.
DOMAIN	Domain Abbreviation	Char	Two-character abbreviation for the domain, which must be TS.
TSSEQ	Sequence Number	Num	Sequence number given to ensure uniqueness within the dataset.
TSGRPID	Group ID	Char	Used to tie together a group of related records.
TSPARMCD	Trial Summary Parameter Short Name	Char	Short character value for the trial summary characteristic described in TSPARM. The short value can be up to 8 characters. Examples: DESIGN, MASK, COMPRT.
TSPARM	Trial Summary Parameter	Char	Term for the Trial Summary Parameter. The value in TSPARM cannot be longer than 40 characters. Examples: DESCRIPTION OF TRIAL DESIGN, TRIAL BLINDING SCHEMA, COMPARATIVE NAME OF TREATMENT.
TSVAL	Parameter Value	Char	Value of the TS Parameter. Example: "ASTHMA" when TSPARM is 'Indication'.

# 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).
- 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 observation 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 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 a set of subject records within the same domain. 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 RECORDS IN SEPARATE DATASETS

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	Unique identifier for a study.
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	Identifying Variable	Char	Identifying variable in the dataset that identifies the related record(s). Examples: <a href="#">--SEQ</a> and <a href="#">--GRPID</a> .
IDVARVAL	Identifying Variable Value	Char	Value of identifying variable of the related parent record(s). Used only when individual records are being related.
RELTYPE	Relationship Type	Char	Identifying the hierarchical level of the records in the relationship. Values must be either ONE or MANY. Used only when datasets are being related.
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 DATASETS

The Related Records (RELREC) dataset can also be used to identify relationships between separate datasets that may have a logical relationship for the purpose of data review or analysis. 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 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 INVESTIGATOR 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 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 observation classes. Details on file naming, expected SUPPQUAL values and appropriate controlled terminology, such as the population flags for 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 Name	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	Identifying Variable	Char	Identifying variable in the dataset that identifies the related record(s). Examples: --SEQ, --GRPID.
IDVARVAL	Identifying Variable Value	Char	Value of identifying variable of the parent record(s).
QNAM	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. Examples include CRF, ASSIGNED, or DERIVED.
QEVAL	Evaluator	Char	Used only for results that are subjective (e.g., assigned by a person or a group). Should be Null for records that contain objectively collected or derived data. Some examples include ADJUDICATION COMMITTEE, STATISTICIAN, DATABASE ADMINISTRATOR, CLINICAL COORDINATOR, PRIMARY INVESTIGATOR.

## 4.5 RELATING COMMENTS TO A PARENT DOMAIN

The Comments (CO) 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 describes the basic concepts and general structures of the model. The implementation guides, such as those for the CDISC Study Data Tabulation Model Implementation Guide for Human Clinical Trials 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.

# 6 SDTM Version history

Version 1.2 represents the third formal release of the Study Data Tabulation Model. The original version was released as the Study Data Tabulation Model Version 1.0 in June, 2004. The last prior version was issued in June 2005. Much of SDTM V1.0 was extracted from the CDISC Submission Data Domain Models Version 3.0 (V3), which was approved in June 2003. V3 was also balloted through HL7 as the Clinical Trial Data Regulatory Submission Model in April 2003.

V3 represented a major change from the earlier CDISC data domain models because it incorporated a general model for representing all types of study data. It 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 the SDTM Version 1.0 and its implementation guides. 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 version 1.0.

Version 1.0 of the SDTM varied in several significant ways from the prior released version (V3):

- The name was changed to better reflect a broader scope; the model now applies to data tabulations and not to other CRT data presentations (e.g., data listings, analysis datasets, subject profiles). In addition, it applies to data collected in both human and animal studies.
- Detailed assumptions, business rules, examples, and specifications for representative data domains were removed from this document and published as separate implementation guides available through CDISC. Implementation details for Clinical Trials data submissions are published by [CDISC as the Study Data Tabulation Model Implementation Guide](#) for Human Clinical Trials. Implementation details for Animal Toxicology data submissions, which have been prepared by the Standard for Exchange of Non-clinical Data (SEND) group, also based on this model, have been published as the [SEND Implementation Guide](#).
- The model was expanded to include additional variables for each general observation class, as well as additional concepts.

## 6.1 CHANGES FROM CDISC SUBMISSION DATA DOMAIN MODELS V3 TO SDTM V1.0

- 1) Extraction of the chapter describing the V3 General Study Information Model as a separate document now known as the SDTM.
- 2) 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
- 3) Incorporation of a new “Trial Design” component to the SDTM
- 4) Creation of a more thorough solution for defining relationships between datasets, between records in different domains, and between supplemental qualifiers and a parent domain
- 5) Representation of all date/time variables and durations in ISO 8601 character format (rather than as seconds since January 1, 1960); elimination of the concept of a separate date/time Precision variable for each date/time variable, and elimination of the --DURU variable
- 6) Addition of new domain variables to represent additional timing descriptions, flags, and descriptive attributes of an observation (e.g., --SCAT, --DOSRGM, --NRIND)
- 7) 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

- 8) Numerous changes to variables, labels, formats, and notes to reduce ambiguity and improve consistency.
- 9) Removal of the variable --TRTEM from the Events observation class (table 2.2.2); this variable, which requires an Evaluator, should be represented in Supplemental Qualifiers
- 10) Removal of the variable --SOTHC from the Events observation class (table 2.2.2); this variable, when used, should be represented in Supplemental Qualifiers
- 11) Renaming of the variable --RELOTH from the Events observation class (table 2.2.2) to --RELNST to retain consistency of meaning of the SDS “OTH” naming fragment
- 12) Renaming of the variable --SOTH from the Events observation class (table 2.2.2) to --SMIE to retain consistency of meaning of the SDS “OTH” naming fragment
- 13) Redefinition of and renaming of the variable --TOXCAT from the Findings class (table 2.2.3) to --TOX to more accurately describe the intended concept
- 14) Removal of the variable --LNKSEQ from the Findings observation class (table 2.2.3); this variable was determined not to be necessary, given subsequent revisions to the general structure for relationships defined in Version 1.1.
- 15) Redefinition of the variable --TOXGR from the Findings observation class (table 2.2.3) to more accurately describe the intended concept.
- 16) Redefinition of the timing variables EPOCH, --STRF and --ENRF (table 2.2.5) to more accurately describe the intended concept
- 17) Removal of the variable RACEOTH from the Demographics special purpose domain (table 2.2.6)
- 18) Addition of variable TEDUR to the Trial Design model (table 3.2.1)
- 19) Renaming of the variable TICRIT from the Trial Design Model (table 3.4.1) to TIRL to more accurately describe the intended concept
- 20) Abridgement of certain variable labels (to conform with the SAS Transport 40 character limit)
- 21) Addition a new variable --METHOD to the Findings general class.

## 6.2 CHANGES FROM SDTM V1.0 TO SDTM V1.1

- 22) Addition of the new variables --TRTV and --ADJ to the Interventions general class
- 23) Addition of the new variables --RESCAT, --SEV, --DTHREL to the Findings general class
- 24) Addition of the new variables --RFTDTC and --EVLINT to the Timing Variables
- 25) Addition of examples and valid values to the Description for several variables in Table 2.2.2
- 26) Correction of the descriptions for several variables in Table 2.2.3, 2.2.4, 2.2.5, 2.2.6 and 2.2.7
- 27) Correction of the label for AGE in Table 2.2.6
- 28) Correction of the label for ARM in tables 3.2.2 and 3.2.3
- 29) Changed label for IDVAR and IDVARVAL from “Identifier” to “Identifying” in tables 2.2.7, 4.2.1, and 4.4.1 to avoid confusion with Identifier variables in table 2.2.4.
- 30) Addition of a new Trial Summary (TS) domain to the Trial Design Model in Section 3.5
- 31) Removal of reference to lesions example in sections 4 and 4.3.
- 32) Application of a number of minor text corrections throughout as appropriate
- 33) Relocation of the list of changes to Appendix 6.1.
- 34) Deleted last sentence in paragraph in Section 2.2, paragraph 2: “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”.

- 35) Used term “Define data definition document” consistently
- 36) Removed “and vary according to the type of observation” from definition of Topic Variable
- 37) Corrected examples for --METHOD in Table 2.2.3
- 38) Corrected examples for --ENRF in Table 2.2.5
- 39) Corrected notes to SUBJID and SITEID in Table 2.2.6, and removed stray text below table
- 40) Corrected Labels for IDVAR and IDVARVAL and note to COREF in Table 2.2.7
- 41) Added sentence to paragraph 2 in Section 3.1 to note global use of trial design variables
- 42) Corrected heading to Table 3.2.2 to “One Record per Planned Element per Arm”
- 43) Corrected heading to Table 3.2.3 to “One Record per Planned Trial Visit”
- 44) Corrected heading to Table 3.3.1 to “One Record per Actual Element per Subject”
- 45) Corrected Examples in TSPARM Note to Table 3.5.1
- 46) Corrected Section Name titles in sections 4.1, 4.2 and 4.3.

### 6.3 CHANGES FROM SDTM V1.1 TO SDTM V1.2

- 47) Deleted detailed history of prior versions in Section 1.1
- 48) Removed “\*” references to controlled terminology; controlled terminology now specified in implementation guides.
- 49) Minor corrections to definitions of Interventions and Findings general classes in Section 2.2.
- 50) Added new qualifier variable --PRESP to list of qualifiers in Table 2.2.1 and 2.2.2.
- 51) Added new qualifier variables --VAMT and --VAMTU in Table 2.2.1.
- 52) Added new qualifier variable --OBJ in table 2.2.3.
- 53) Added new timing variables --STRPT, --ENRPT, --STPT, --ENTPT to Table 2.2.5.
- 54) Reclassified variables RFSTDTC and RFENDTC as Qualifiers in Table 2.2.6.
- 55) Shifted order of variables in Table 2.2.7 (CODTC now appears last).
- 56) Moved Subject Elements and Subject Visits tables from Section 3 to Section 2 (Tables 2.2.8 and 2.2.9).
- 57) Shifted order of variables in Table 3.2.3 (VISITNUM now appears third).
- 58) Added TIVERS to Table 3.3.1.
- 59) Added TSGRPID to table 3.4.1.
- 60) Minor corrections and clarifications to descriptions or examples in Tables for consistency with SDTMIG. In some cases, certain implementation details that were more suited to the implementation guides have been removed.
- 61) Several clarifications to text and typographical errors corrected throughout.