

A Relational Understanding of SDTM Tables

John R. Gerlach, MaxisIT, Inc.
Glenn O'Brien, ALTANA Pharma US, Inc.

Abstract

The Study Data Tabulation Model (SDTM) is fast becoming the industry standard for processing data in clinical trials. Although the CDISC standard is well defined, with obvious benefits to those who manage and analyze the data, it is not easy to implement. For example, the design of a Case Report Form (CRF) or how the data are presented in reports affords little help to mapping raw data to SDTM variables and their appropriate domains. Indeed, the data pertaining to one SDTM domain may originate from several pages in a CRF; conversely, the data found on one page of a CRF might map to several SDTM domains, such as demography (DM) and inclusion/exclusion criteria (IE). Thus, the process of creating valid SDTM domain data sets requires a thorough understanding of SDTM domains.

In order to become more learned about SDTM domains, it is important to develop a *relational* understanding of SDTM variables across domains, that is, the variables *without* their domain prefix (e.g., SEQ, not AESEQ). Better still, the relational schema can be class specific, for example, domains pertaining only to the Events class (i.e., Adverse Events, Patient Disposition, Medical History). Even better, the relational schema can indicate the data type and core function (i.e., Required, Expected, Permissible) of each variable, along with its label, across the several domains, assuming the variable exists in the domain. This paper explains a method for learning about SDTM domains by producing class-specific relational schemas.

Introduction

In July 2004 the Clinical Data Interchange Standards Consortium (CDISC) published standards on the design and content of clinical trial tabulation data sets, known as the Study Data Tabulation Model (SDTM). According to the CDISC standard, there are four ways to represent a subject in a clinical study: tabulations, data listings, analysis datasets, and subject profiles. With the implementation of the CDISC standard, trained professionals can use software tools to work more efficiently. Moreover, clinical trials following this standard can be consolidated into a repository for further research.

SDTM domains contain observations about a subject that are topic-specific in a study. The variables in each domain are pre-defined for which there are five major categories, illustrated by the following examples.

Identifier	USUBJID	Subject identifier
Topic	LBTYPE	Type of lab test
Timing	LBDMTC	Date / time of lab measurement
Qualifier	LBORRESU	Units of original lab measurement
Rule	TEDUR	Rule describing the duration of a Trial Element.

Most SDTM variables are distinguished by a two-character identifier that denotes the domain itself. For example, the timing variable AEDTC is found in the AE (Adverse Events) domain and represents the start date/time of an event following the ISO 8601 convention (i.e., *yyyy-mm-ddThh:mm:ss*). In order to produce more meaningful relational schemas, that is, to show the existence of a variable across specific-type domains, the prefix of variables denoting the domain will be ignored.

Despite the fact that even similar variables (e.g., AESEQ, LBSEQ) are not like common variables (e.g., STUDYID) found in a typical relational data model, the CDISC standard has evolved into a general relational model for representing all types of study data, even defining relationships between records in different domains, as well as between so-called supplemental qualifiers and a parent domain. Consequently, it behooves the SAS professional working in clinical trials to develop a relational understanding of the SDTM model.

More About SDTM Domains

SDTM Domains are grouped by classes, which is useful for producing more meaningful relational schemas. Consider the following domain classes and their respective domains.

- **Special Purpose Class** – Pertains to unique domains concerning detailed information about the subjects in a study.
 - Demography (DM)
 - Comments (CM)
- **Findings Class** – Collected information resulting from a planned evaluation to address specific questions about the subject, such as whether a subject is suitable to participate or continue in a study.
 - Electrocardiogram (EG)
 - Inclusion / Exclusion (IE)
 - Lab Results (LB)
 - Physical Examination (PE)
 - Questionnaire (QS)
 - Subject Characteristics (SC)
 - Vital Signs (VS)
- **Events Class** – Incidents independent of the study that happen to the subject during the lifetime of the study.
 - Adverse Events (AE)
 - Patient Disposition (DS)
 - Medical History (MH)
- **Interventions Class** – Treatments and procedures that are intentionally administered to the subject, such as treatment coincident with the study period, per protocol, or self-administered (e.g., alcohol and tobacco use).
 - Concomitant Medications (CM)
 - Exposure to Treatment Drug (EX)
 - Substance Usage (SU)
- **Trial Design Class** – Information about the design of the clinical trial (e.g., crossover trial, treatment arms) including information about the subjects with respect to treatment and visits.
 - Subject Elements (SE)
 - Subject Visits (SV)
 - Trial Arms (TA)
 - Trial Elements (TE)
 - Trial Inclusion / Exclusion Criteria (TI)
 - Trial Visits (TV)

Besides the name and data type of each variable in a domain dataset, the relational schema includes another very important piece of information, the Core function of each variable. This attribute ensures CDISC compliance and provides guidance for those creating the domains. The Core function of a variable falls into three categories:

- **Required** – A variable that is fundamental or pertinent to the identification of the domain. These variables are always included in the domain data set and cannot contain null values.
- **Expected** – A variable that makes a record meaningful in the context of its domain. These variables should exist, but may contain null values.
- **Permissible** – A variable that should exist if it is appropriate, either collected or derived. All timing variables are exemplary of this core function.

The Elements of the Report

In order to produce a meaningful schema, it is necessary to ignore the prefix component of variables denoting the domain (e.g., AESEQ). For example, the ubiquitous variable denoting Sequence Number (<domain>SEQ) becomes SEQ in the schema. Without eliminating the prefix denoting the domain, this very important variable would not be listed in a single row, across domains, indicating existence (data type and core function), which is the primary objective of the report. Again, the purpose is to produce a relational schema of class-specific domains. Also, keep in mind that there are variable names that do not use the domain as part of the variable identifier, such as: STUDYID, DOMAIN, USUBJID, SUBJID; found in most domains, as well as others that are specific to one or more domains (e.g., SEXCD, ARMCD).

The CDISC SDTM Implementation Guide (SDS Version 3.1) contains a list of keywords along with their respective variable ID component, called a fragment, that are used to name variables. For example, the fragment DUR denotes duration of an event, which is found in the Adverse Events (AE) domain. Even though the relational schema contains the label of the variables, the following list should help to understand the naming convention for CDISC variables.

Keyword	Fragment	Keyword	Fragment	Keyword	Fragment
ACTION	ACN	BASELINE	BL	BODY	BOD
CANCER	CAN	CONDITION	CND	CODE	CD
COMPLIANCE	CP	CONGENITAL	CONG	DECODE	DECOD
DISABILITY	DISAB	DISPOSITION	DS	DURATION	DUR
ELAPSED	EL	ELEMENT	ET	EMERGENT	EM
FLAG	FL	GROUP	GRP	HOSPITALIZATION	HOSP
INDICATION	INDC	INDICATOR	IND	LOCATION	LOC
LOINC CODE	LOINC	LOWER_LIMIT	LO	NAME	NAM
NOT DONE	ND	NUMBER	NUM	NUMERIC	N
ONGOING	ONGO	ORIGIN	ORIG	OUTCOME	OUT
POSITION	POS	REASON	REAS	REGIMENT	RGM
RESULT	RES	RULE	RL	SEQUENCE	SEQ
SERIOUS	S, SER	SEVERITY	SEV	SPONSOR	SP
START	ST	STATUS	STAT	SUBCATEGORY	SUBCAT
SUBJECT	SUBJ	TIME	TM	TOTAL	TOT
TREATMENT	TRT	UNIT	U	VALUE	VAL

The report (relational schema) lists the set of variables (excluding the domain prefix) for a specific class juxtaposed with those domains indicating the data type (Character | Numeric) and core function (Requested | Expected | Permissible) for each variable, that is, if the variable exists in that domain; otherwise, a dash is written to indicate that the variable does not exist in that domain. Consider the illustration below that shows the general layout of the schema. Notice the sub-title that indicates the domain class, such as Findings, such that only those domains (e.g., Physical Exams, Vital Signs, etc.) contribute their variables to the report, most importantly, the data type and core function of each variable. Consequently, the reader can develop a relational understanding of the several domains in the context of the respective domain class.

Relational Schema of Standard SDTM Domains
(<Domain Class>)

Name	Domain1	Domain2	Domain3	...
Variable1	<C N>/<R E P>	-	-	
Variable2	<C N>/<R E P>	<C N>/<R E P>	-	
Variable3	-	-	<C N>/<R E P>	
	:	:	:	:
Varn	*	-	*	

Obtaining the Relevant Data

In order to produce schemas on class-specific domains, it is necessary to associate a domain identifier to its class, which is easily accomplished by the user-defined format below.

```
proc format;
  value $classf 'AE' = 'Events'           'CM' = 'Interventions'
               'CO' = 'Special Purpose'   'DM' = 'Special Purpose'
               'DS' = 'Events'           'EG' = 'Findings'
               'EX' = 'Interventions'     'IE' = 'Findings'
               'LB' = 'Findings'         'MH' = 'Events'
               'PE' = 'Findings'         'QS' = 'Findings'
               'SC' = 'Findings'         'SE' = 'Trial Design'
               'SU' = 'Interventions'     'SV' = 'Trial Design'
               'TA' = 'Trial Design'     'TE' = 'Trial Design'
               'TI' = 'Trial Design'     'TV' = 'Trial Design'
               'VS' = 'Findings';
run;
```

Besides the \$classf format, the following metadata is required to produce the intended report: variable identifier, variable label, data type, and core function. Because the order of SDTM variables is considered as part of the CDISC standard, another variable, called *order*, will be used to produce the report. In fact, the utility produces a variable call *group* so that identically named variables will be listed first, followed by the other ‘common’ variables.

Given a data file that contains metadata about CDISC domains, the following SAS code reads the relevant metadata about the standard domains and imputes the other variables germane to the report. For this example, assume that the variable NAME does not contain the domain prefix, such as SEQ, DUR, DTC, rather than AESEQ, AEDUR, CODTC.

```
filename sdtm 'SDTM_Vars.txt';

data sdtm_vars;
  length domain name $8 label $60 type $4 core t_c $3;
  infile sdtm;
  input order domain name label type core;
  class = put(domain,$classf.);
  t_c = upcase(substr(type,1,1)) || '/' || upcase(substr(core,1,1));
  select(name);
    when('STUDYID') do; group=1; order = 1; end;
    when('DOMAIN') do; group=1; order = 2; end;
    when('USUBJID') do; group=1; order = 3; end;
    when('SUBJID') do; group=1; order = 4; end;
    otherwise do; group=2; end;
  end;
  keep domain group order name t_c label;
run;
```

The Utility

The utility, a SAS macro, processes a data set that contains the pertinent information (domain, variable ID, data-type, core-function, label, and class) and produces a class-specific relational schema. Initially, it is necessary to obtain only those observations that are relevant to a given class (e.g., Findings), which the SORT procedure accomplishes easily.

```
proc sort data=sdtm_vars out=class;
  by domain name;
  where upcase(class) eq "%upcase(&class.)";
run;
```

Next it is necessary to determine the number of domains for that class. For the Special Purpose class, there are only two domains; whereas, for the Findings class, there seven domains. In any case, the number of domains is determined by the completeness of the data source from which the metadata originated. The SQL step below creates a macro variable denoting the number of class-specific domains, and the following SQL step creates *n*-macro variables denoting the several domains.

```
proc sql noprint;
  select left(put(count(distinct domain),best.)) into :ndomains
  from class;
```

```
quit;
```

```
proc sql noprint;
  select distinct(domain) into :domain1 - :domain&ndomains.
  from class;
quit;
```

Now comes the conceptual hard part. How do you create the appropriate data set that produces the aforementioned report, a relational schema, listing the relevant variables and, in juxtaposition, denoting their respective existence (data type and core function) or non-existence for a class-specific domain? The structure of the acquired metadata is normalized, whose unit of analysis is the domain and variable; whereas, the intended report lists variables across domains, that is, the domains are column headers. Consider the partial listing of the metadata data set found in the AE domain.

Metadata - AE Domain					
DOMAIN	NAME	GROUP	ORDER	T_C	LABEL
AE	STUDYID	1	1	C/R	Study Identifier
AE	DOMAIN	1	2	C/R	Domain Abbreviation
AE	USUBJID	1	3	C/R	Unique Subject Identifier
AE	SEQ	2	4	C/R	Sequence Number
AE	GRPID	2	5	C/P	Group ID
AE	REFID	2	6	C/P	Reference ID
AE	SPID	2	7	C/P	Sponsor ID
AE	TERM	2	8	C/R	Reported Term for Adverse Event
AE	MODIFY	2	9	C/P	Modified Reported Term
AE	DECOD	2	10	C/R	Dictionary-Derived Term
AE	CAT	2	11	C/P	Category for Adverse Event
AE	SCAT	2	12	C/P	Subcategory for Adverse Event
AE	OCCUR	2	13	C/P	Adverse Event Occurrence
AE	BODSYS	2	14	C/E	Body System or Organ Class
AE	LOC	2	15	C/P	Location of the Reaction
AE	SEV	2	16	C/P	Severity / Intensity
AE	SER	2	17	C/E	Serious Event
AE	ACN	2	18	C/E	Action Taken with Study Treatment
:	:	:	:	:	:

Since the normalized metadata data set contains all the domains, it is necessary to create subset data sets each representing a particular domain, then to perform a match-merge by the NAME variable. However, because the matrix of common variables and specific domains indicates the data type and core function in its cells, it is necessary to rename the T_C variable, appropriately, to the name of the respective domain. But, how do you do this systematically?

Recall that the previous SQL steps generated macro variables denoting the number of domains and their names for a specific class. The following Data step uses these macro variables to formulate the needed MERGE statement, along with the WHERE data set option to obtain the subset data set representing each domain. Also, notice that the variable T_C is renamed to its respective domain data set name (e.g., CO, DM). It is sufficient to merge the several data sets by the variable NAME. The variable denoting the data type and core function, called by its domain name, contains a blank for those instances where the domain data set does not contribute an observation to the match merge, since the variable does not exist in that domain. The DO loop with an IF statement supplants the missing value with a dash by using the array.

Consider the following Data step that creates the appropriate schema for a specific class of domains.

```
data schema;
  array domains{*}$5 %do i = 1 %to &ndomains.; &&domain&i.. %end; ;
  merge %do i = 1 %to &ndomains.;
    class(where=(domain eq "&&domain&i..") rename=t_c=&&domain&i..)
    %end; ;
  by name;
  do i = 1 to dim(domains);
    if domains{i} eq '' then domains{i} = ' - ';
  end;
  drop i class domain;
```

```
run;
```

The reporting data set contains all the pertinent information needed to produce the schema that represents several domains pertaining to a class and contains the collection of variables along with an appropriate label and an element denoting the data type and core function of the variable. The REPORT procedure generates the desired report.

```
proc report data=schema nowindows headline headskip split='!';
  columns group order name label ('- Domains -' %do i = 1 %to &ndomains.; &&domain&i.. %end;) ;
  define group      / order noprint;
  define order      / order noprint;
  define name       / order id width=8  'Variable!Name';
  define label      / display id width=40 'Label';
  break after group / skip;
  title2 "Relational Schema of Standard SDTM Domains";
  title3 "( %upcase(&class.) )";
run;
```

Keep in mind that the previous steps reside in a macro definition called %schema. This macro contains only one keyword parameter, indicating the default class Special Purpose. Now, consider several invocations of the %schema macro that produces the intended relational schema.

```
%schema();                                %schema(class=findings);                %schema(class=events);
%schema(class=interventions);             %schema(class=trial design);
```

We proceed now to study the several reports with the intent to develop a better understanding of SDTM domains.

Special Purpose Class

Unlike the other classes, there is not much similarity amongst the two Special Purpose domains: Comments (CO) and Demography (DM). In fact, except for the usual set of identifying variables, there is only one common variable, the Date/Time of Collection (DTC), that is, the variables CODTC and DMDTC. Otherwise, the collection of variables is almost mutually exclusive. Also noteworthy is the core function of the DTC variable is consistent (i.e., Permissible), which is surprisingly not always the case.

Relational Schema of Standard SDTM Domains
(SPECIAL PURPOSE)

Variable Name	Label	- Domains --	
		CO	DM
STUDYID	Study Identifier	C/R	C/R
DOMAIN	Domain Abbreviation	C/R	C/R
USUBJID	Unique Subject Identifier	C/R	C/R
SUBJID	Subject Identifier for the Study	-	C/R
AGE	Age in AGEU at Reference Date/Time	-	N/E
AGEU	Age Units	-	C/E
ARM	Description of Arm	-	C/R
ARMCD	Arm Code	-	C/R
BRTHDTC	Date/Time of Birth	-	C/P
COUNTRY	Country	-	C/R
DTC	Date/Time of Collection	C/P	C/P
DY	Study Day of Collection	-	N/P
ETHNIC	Ethnicity	-	C/P
EVAL	Evaluator	C/P	-
IDVAR	Identifier Variable Name	C/P	-
IDVARVAL	Identifier Variable Value	C/P	-
INVID	Investigator Identifier	-	C/P
INVNAM	Investigator Name	-	C/P
RACE	Race	-	C/E
RDOMAIN	Related Domain Abbreviation	C/E	-
REF	Comment Reference	C/P	-
RFENDTC	Subject Reference End Date/Time	-	C/R
RFSTDTCT	Subject Reference Start Date/Time	-	C/R
SEQ	Sequence Number	N/R	-
SEX	Sex	-	C/R

SITEID	Study Site Identifier	-	C/R
VAL	Comment	C/R	-

Findings Class

The Findings class pertains to information about a subject related to some kind of evaluation or assessment, which includes: Electrocardiogram (EG), Inclusion / Exclusion (IE), Laboratory Results (LB), Physical Examinations (PE), Questionnaire (QS), Subject Characteristics (SC), and Vital Signs (VS). The relational schema below shows that the primary common variables (Study ID, Domain, and Subject ID) exist in all the domains for that class, as well as date-related variables. Notice that the variables concerning visits are found in all except Subject Characteristic (SC). Also, the variable ORRES is 'Expected' in all domains except for the Inclusion / Exclusion (IE) domain, where it is 'Required,' which makes sense. It is left to the reader to consider other common traits or differences with respect to existence, data type, or core function.

Relational Schema of Standard SDTM Domains
(FINDINGS)

Variable Name	Label	Domains						
		EG	IE	LB	PE	QS	SC	VS
STUDYID	Study Identifier	C/R	C/R	C/R	C/R	C/R	C/R	C/R
DOMAIN	Domain Abbreviation	C/R	C/R	C/R	C/R	C/R	C/R	C/R
USUBJID	Unique Subject Identifier	C/R	C/R	C/R	C/R	C/R	C/R	C/R
BLFL	Baseline Flag	C/E	-	C/E	C/E	C/E	-	C/E
BODSYS	Body System or Organ Class	-	-	-	C/P	-	-	-
CAT	Category for Vital Signs	C/P	C/R	C/E	C/P	C/R	C/P	C/P
DRVFL	Derived Flag	C/P	-	C/P	-	C/P	-	C/P
DTC	Date/Time of Measurements	C/E	C/E	C/E	C/E	C/E	C/E	C/E
DY	Study Day of Vital Signs	N/P	N/P	N/P	N/P	N/P	N/P	N/P
ELTM	Elapsed Time from Reference Point	C/P	-	C/P	-	C/P	-	C/P
EVAL	Evaluator	C/E	-	-	C/P	-	-	-
FAST	Fasting Status	-	-	C/P	-	-	-	-
GRPID	Group ID	C/P	-	C/P	C/P	C/P	C/P	C/P
LOC	Location of Vital Signs Measurement	-	-	-	C/P	-	-	C/P
LOINC	LOINC Code	C/P	-	C/P	-	-	-	C/P
METHOD	Method of Test or Examination	C/P	-	C/P	-	-	-	-
MODIFY	Modified Reported Term	-	-	-	C/P	-	-	-
NAM	Vendor Name	C/P	-	C/P	-	-	-	-
NRIND	Reference Range Indicator	C/	-	C/E	-	-	-	-
ORNRHI	Normal Range Upper Limit in Orig Units	-	-	C/E	-	-	-	-
ORNRLO	Normal Range Lower Limit in Orig Units	-	-	C/E	-	-	-	-
ORRES	Result or Finding in Original Units	C/E	C/R	C/E	C/E	C/E	C/E	C/E
ORRESU	Original Units	C/P	-	C/E	C/P	C/P	C/P	C/E
POS	Vital Signs Position of Subject	C/E	-	-	-	-	-	C/E
REASND	Reason Not Performed	C/P	-	C/P	C/P	C/P	C/P	C/P
REFID	Specimen ID	C/P	-	C/P	-	-	-	-
SCAT	Subcategory for Vital Signs	C/P	C/P	C/P	C/P	C/P	C/P	C/P
SEQ	Sequence Number	N/R	N/R	N/R	N/R	N/R	N/R	N/R
SPCCND	Specimen Condition	-	-	C/P	-	-	-	-
SPEC	Specimen Type	-	-	C/P	-	-	-	-
SPID	Sponsor ID	C/P	C/P	C/P	C/P	C/P	C/P	C/P
STAT	Vitals Status	C/P	-	C/P	C/P	C/P	C/P	C/P
STNRC	Reference Range for Char Rslt-Std Units	-	-	C/P	-	-	-	-
STNRHI	Normal Range Upper Limit-Standard Units	-	-	N/E	-	-	-	-
STNRLO	Normal Range Lower Limit-Standard Units	-	-	N/E	-	-	-	-
STRESC	Character Result/Finding in Std Format	C/E	C/R	C/E	C/E	C/E	C/E	C/E
STRESN	Numeric Result/Finding in Standard Units	N/P	-	N/E	N/E	N/P	N/P	N/E
STRESU	Standard Units	C/P	-	C/E	C/E	C/P	C/P	C/E
TEST	Vital Signs Test Name	C/R	C/R	C/R	C/R	C/R	C/R	C/R
TESTCD	Vital Signs Test Short Name	C/R	C/R	C/R	C/R	C/R	C/R	C/R
TOX	Toxicity	-	-	C/P	-	-	-	-
TOXGR	Standard Toxicity Grade	-	-	C/P	-	-	-	-
TPT	Planned Time Point Name	C/P	-	C/P	-	C/P	-	C/P
TPTNUM	Planned Time Point Number	N/P	-	N/P	-	N/P	-	N/P
TPTREF	Time Point Reference	C/P	-	C/P	-	C/P	-	C/P
VISIT	Visit Name	C/P	C/P	C/P	C/P	C/P	-	C/P
VISITDY	Planned Study Day of Visit	N/P	N/P	N/P	N/P	N/P	-	N/P
VISITNUM	Visit Number	N/R	N/P	N/R	N/E	N/E	-	N/R

Events Class

The Events class pertains to an occurrence or incident independent of the clinical trial, such as an adverse event (AE), occurring during the trial, and medical history (MH), occurring prior to the trial. This class documents protocol milestones such as randomization and patient disposition (DS) (e.g., completed study). Obviously, the AE domain represents most of the data found in this class. Notice that the DS and MH domains have actual visit numbers; whereas, AE does not, which makes sense, since such an event is not planned according to the protocol.

Relational Schema of Standard SDTM Domains
(EVENTS)

Variable Name	Label	----- Domains -----		
		AE	DS	MH
STUDYID	Study Identifier	C/R	C/R	C/R
DOMAIN	Domain Abbreviation	C/R	C/R	C/R
USUBJID	Unique Subject Identifier	C/R	C/R	C/R
ACN	Action Taken with Study Treatment	C/E	-	-
ACNOTH	Other Action Taken	C/P	-	-
BODSYS	Body System or Organ Class	C/E	-	C/E
CAT	Category for Medical History	C/P	C/P	C/P
CONTRT	Concomitant or Additional Trtmnt Given	C/P	-	-
DECOD	Dictionary-Derived Term	C/R	C/R	C/E
DTC	Date/Time of History Collection	-	C/P	C/P
DUR	Duration of Event	C/P	-	-
DY	Study Day of History Collection	-	-	N/P
ENDTC	End Date/Time of Medical History Event	C/E	-	C/P
ENDY	Study Day of End of Event	N/P	-	-
ENRF	End Relative to Reference Period	C/P	-	C/P
EPOCH	Trial Epoch	-	C/P	-
GRPID	Group ID	C/P	C/P	C/P
LOC	Location of the Reaction	C/P	-	-
MODIFY	Modified Reported Term	C/P	-	C/P
OCCUR	Medical History Occurrence	C/P	-	C/P
OUT	Outcome of Adverse Event	C/P	-	-
PATT	Pattern of Event	C/P	-	-
REASND	Reason Medical History Not Collected	-	-	C/P
REFID	Reference ID	C/P	C/P	C/P
REL	Causality	C/E	-	-
RELNST	Relationship to Non-Study Treatment	C/P	-	-
SCAN	Involves Cancer	C/P	-	-
SCAT	Subcategory for Medical History	C/P	C/P	C/P
SCONG	Congenital Anomaly or Birth Defect	C/P	-	-
SDISAB	Persist or Signif Disability/Incapacity	C/P	-	-
SDTH	Results in Death	C/P	-	-
SEQ	Unique Sequence Number	N/R	N/R	N/R
SER	Serious Event	C/E	-	-
SEV	Severity/Intensity	C/P	-	-
SHOSP	Requires or Prolongs Hospitalization	C/P	-	-
SLIFE	Is Life Threatening	C/P	-	-
SMIE	Other Medically Important Serious Event	C/P	-	-
SOD	Occurred with Overdose	C/P	-	-
SPID	Sponsor ID	C/P	C/P	C/P
STAT	Medical History Status	-	-	C/P
STDTC	Start Date/Time of Medical History Event	C/E	C/E	C/P
STDY	Study Day of Start of Disposition Event	N/P	N/P	-
TERM	Reported Term for the Medical History	C/R	C/R	C/R
TOXGR	Standard Toxicity Grade	C/P	-	-
VISIT	Visit Name	-	C/P	C/P
VISITDY	Planned Study Day of Visit	-	N/P	N/P
VISITNUM	Visit Number	-	N/P	N/P

Interventions Class

The Interventions class pertains to information about treatment either as specified by the protocol or prior to the study period. This class includes the domains Concomitant Medications (CM), Exposure (EX), and Substance Use (SU). Once again, the primary common variables exist in all three domains and have the same attributes. Notice that several of the variables change across domains with respect to core function, such as DOSE (Substance Use Consumption) and ENDTC (End Date / Time of Substance Use). Also, the CM and EX domains are not concerned with the variables denoting visits, since visits are scheduled according to the protocol.

Relational Schema of Standard SDTM Domains
(INTERVENTIONS)

Variable Name	Label	----- Domains -----		
		CM	EX	SU
STUDYID	Study Identifier	C/R	C/R	C/R
DOMAIN	Domain Abbreviation	C/R	C/R	C/R
USUBJID	Unique Subject Identifier	C/R	C/R	C/R
CAT	Category for Substance Use	C/P	C/P	C/P
CLAS	Substance Use Class	C/P	-	C/P
CLASCD	Substance Use Class Code	C/P	-	C/P
DECOD	Standardized Substance Name	C/P	-	C/P
DOSE	Substance Use Consumption	N/P	N/E	N/P
DOSFRM	Dose Form	C/P	C/P	C/P
DOSFRQ	Use Frequency Per Interval	C/P	C/P	C/P
DOSRGM	Intended Dose Regimen	C/P	C/P	-
DOSTOT	Total Daily Consumption using SUDOSU	N/P	N/P	N/P
DOSTXT	Substance Use Consumption Text	C/P	C/P	C/P
DOSU	Consumption Units	C/P	C/E	C/P
DUR	Duration of Substance Use	C/P	C/P	C/P
ELTM	Planned Elapsed Time from Reference Pt	-	C/P	-
ENDTC	End Date/Time of Substance Use	C/P	C/E	C/P
ENDY	Study Day of End of Substance Use	N/P	N/P	N/P
ENRF	End Relative to Reference Period	C/P	-	-
GRPID	Group ID	C/P	C/P	C/P
INDC	Indication	C/P	-	-
LOC	Location of Dose Administration	-	C/P	-
LOT	Lot Number	-	C/P	-
MODIFY	Modified Substance Name	C/P	-	C/P
OCCUR	SU Occurrence	C/P	-	C/P
REASND	Reason Substance Use Not Collected	C/P	-	C/P
ROUTE	Route of Administration	C/P	C/P	C/P
SCAT	Subcategory for Substance Use	C/P	C/P	C/P
SEQ	Sequence Number	N/R	N/R	N/R
SPID	Sponsor ID	C/P	C/P	C/P
STAT	Substance Use Status	C/P	-	C/P
STDTC	Start Date/Time of Substance Use	C/P	C/R	C/P
STDY	Study Day of Start of Substance Use	N/P	N/P	N/P
STRF	Start Relative to Reference Period	C/P	-	-
TAETORD	Order of Element within Arm	-	N/P	-
TPT	Planned Time Point Name	-	C/P	-
TPTNUM	Planned Time Point Number	-	N/P	-
TPTREF	Time Point Reference	-	C/P	-
TRT	Name of Substance	C/R	C/R	C/R
VISIT	Visit Name	-	-	C/P
VISITDY	Planned Study Day of Visit	-	-	N/P
VISITNUM	Visit Number	-	-	N/R

Trial Design Class

Finally, the Trial Design class represents information about the planned sequence of events and the treatment plan for a clinical trial. Also, this class documents events about the subject during the trial. The domains include: Subject Elements (SE), Subject Visits (SV), Treatment Arms (TA), Trial Elements (TE), Trial Inclusion / Exclusion (TI), and Trial Visits. Unlike the IE domain, the TI domain is not subject oriented, since the IE domain contains records only for inclusion and exclusion criteria that a subject did not meet.

Upon inspection of these domains, it is reasonable that the Trial domains do not have a subject identifier, unlike the SE and SV domains. Similarly, date variables are manifest in the Subject domains. Also noteworthy, the variables ARM and ARMCD differ between the TA and TV domains with respect to core function.

Relational Schema of Standard SDTM Domains
(TRIAL DESIGN)

Variable Name	Label	Domains					
		SE	SV	TA	TE	TI	TV
STUDYID	Study Identifier	C/R	C/R	C/R	C/R	C/R	C/R
DOMAIN	Domain Abbreviation	C/R	C/R	C/R	C/R	C/R	C/R
USUBJID	Unique Subject Identifier	C/R	C/R	-	-	-	-
ARM	Description of Arm	-	-	C/R	-	-	C/P
ARMCD	Arm Code	-	-	C/R	-	-	C/E
BRANCH	Branch	-	-	C/E	-	-	-
CAT	Category for Exception Criterion	-	-	-	-	C/R	-
DUR	Planned Duration of Element	-	-	-	C/P	-	-
ELEMENT	Description of Element	C/P	-	C/P	C/R	-	-
ENDTC	End Date/Time of Visit	C/E	C/E	-	-	-	-
ENRL	Visit End Rule	-	-	-	C/R	-	C/P
EPOCH	Trial Epoch	-	-	C/P	-	-	-
ETCD	Element Code	C/R	-	C/R	C/R	-	-
ETORD	Order of Element within Arm	-	-	N/R	-	-	-
RL	Inclusion/Exclusion Criterion Rule	-	-	-	-	C/P	-
STDTC	Start Date/Time of Visit	C/E	C/E	-	-	-	-
STRL	Visit Start Rule	-	-	-	C/R	-	C/P
TEST	Exception Criterion	-	-	-	-	C/R	-
TESTCD	Exception Criterion Short Name	-	-	-	-	C/R	-
TRANS	Transition Rule	-	-	C/E	-	-	-
UPDES	Description of Unplanned Visit	C/P	C/P	-	-	-	-
VISIT	Visit Name	-	C/P	-	-	-	C/R
VISITDY	Planned Study Day of Visit	-	N/P	-	-	-	N/P
VISITNUM	Visit Number	-	N/R	-	-	-	N/R

Conclusion

As with any computer generated report, the outcome is as good as the data. With the growth and development of the CDISC standard, it is obvious that there will be changes, such as variables being added or dropped, attributes being changed, or even the creation of new domains; e.g., the Protocol Deviations domain (DV) found in the CDISC SDTM Implementation Guide (SDS Version 3.1.1). Thus, it is extremely important to have the latest, most complete, version of the metadata in order to produce viable reports.

The Study Data Tabulation Model is becoming the industry standard for clinical trials. Proper implementation of this data model requires an understanding of the rules that map clinical data to their appropriate domains. By using metadata concerning these domains, one can develop a relational understanding of SDTM domains in the context of their specific class with respect to the data type of each variable, if it exists in the domain, and its core function. This SAS solution depicts a clever way to generate relational schemas on class-specific SDTM domains and affords a good opportunity to understand the structure and content of those domains.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.