PharmaSUG2019 - Paper 311

What's New in the SDTMIG v3.3 and the SDTM v1.7

Fred Wood, Data Standards Consulting Group
A Division of TalentMine, Raleigh, NC

ABSTRACT

The SDTMIG v3.3 and the SDTM v1.7 were released in November of 2018. These versions have been published as HTML documents, rather than the typical PDFs of the past. While the SDTMIG v3.2, which was published in 2013, contained 398 pages, the SDTMIG v3.3 would be more than 600 pages if formatted properly and printed to PDF. A number of new morphology/physiology domains have been added. Other new domains and new concepts have been added, and several domains that had undergone public review in 2014 were subsequently combined with existing domains. The Disease Milestones concept, introduced for the TAUG-Diabetes in 2014, is now included in the latest version of the SDTMIG. A new Section 9 includes Study References, with added models for Device Identifiers, Non-host Organism Identifiers, and Pharmacogenomic/Genetic Biomarker Identifiers.

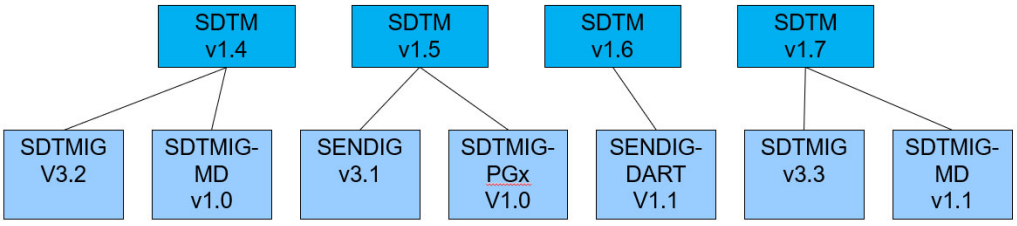
Three versions of the SDTM have been published since the release of v1.4 with the SDTMIG v3.2. These have added not only new variables for general-observation-class domains, but have added new tables including those for as domain-specific variables, SEND reproductive stages, and study references.

The primary focus of this paper will be the changes in SDTMIG v3.3 from SDTMIG v3.2 and to the SDTM versions published in between the publication of these two IGs. When relevant, changes in v3.3 from the batches that went out for public comment will be noted. Changes to the SDTM will be discussed first, followed by changes to the SDTMIG, in approximate order of their appearance in the IG.

INTRODUCTION

CDISC has been providing an accepted standard for the submission of tabulation data in the form of the SDTM and SDTMIG since 2004. The first versions that became part of the FDA's Study Data Specifications were SDTM v.1.0 and SDTMIG v3.1. Prior to the SDTM/SDTMIG, the CDISC Submission Data Standards (SDS) Team had created precursor documents known as Submission Data Standards, which included the 1.x. and 2.x versions of the team's standards-development efforts. For a more detailed history of the SDTM and SDTMIG through 2008, see Wood (1) and Wood and Guinter (2).

The figure below shows the relationship between versions of the SDTM and the SDTM-based implementation guides since the publication of SDTMIG v3.2 in 2013. In order to ensure the alignment of new domains and concepts in each SDTM-based implementation guide (IG) with the SDTM at the time of publication, a new version of the SDTM is released with each IG. In many cases, more than one IG may be associated with a version of the SDTM. The figure shows that two versions of the SDTM were released in between Version 3.2 and 3.3 of the SDTMIG, and SDTM v1.7 with SDTMIG v3.2. This paper will first look at the changes that occurred in Versions 1.5 through 1.7 of the SDTM.



The table below shows additional details around the SDTM and SDTMIG versions shown. It can be seen that number of new domains in the SDTMIG v3.2 reflected the largest single increase between versions. With the current CFAST effort driving the development of therapeutic-area standards, the number of new domains being developed is still increasing, with SDTMIG v3.3 containing at least 50 (see note below). For more information on the impact of CFAST and TransCelerate on the development of SDTM-based standards, see Wood et al. (3).

SDTM Version	Year	SDTMIG Version	Number of Domains*
1.0	2004	3.1	23
1.1	2005	3.1.1	30
1.2	2008	3.1.2	32
1.3	2012	3.1.3	32
1.4	2013	3.2	46
1.5	2016	N/A**	
1.6	2017	N/A***	
1.7	2018	3.3	61

* *Includes all special-purpose, general-observation-class domains, and study references*
** *Created for the SENDIG v3.1*
*** *Created for the SENDIG-DART v1.1*

After the publication of the SDTMIG v3.2 in 2013, the SDS Team had decided that it would be better to release updated material for public review in batches rather than all at once. This was expected to be easier for the Team to manage, and to be easier for public review. Updates that would have been included in the SDTMIG v3.3 were planned to be released in three batches. Batches 1 and 2 had completed public review by the end of 2014. Batch 3 had been scheduled for sometime in Q2 of 2015. What was expected to be in the final version of SDTMIG v3.2 was presented by this author then (4). The Public Review period for Batch 3 occurred in August/September of 2016. The SDS Team received additional comments from FDA in March 2017, which took time to resolve. Handover for publication took place in March 2018, and v3.3 was published in November 2018.

ADDITIONS AND CHANGES TO THE SDTM

There have been a number of changes to the SDTM from Version 1.4, issued with SDTMIG v3.2, through SDTM v1.7, issued with SDTMIG v3.3. These are summarized in the text and tables below. The new tables will be presented first, followed by the new variables.

NEW SDTM TABLES

Version 1.5

SDTM Version 1.5 saw the addition of two tables for Disease Milestones (Subject Disease Milestones and Trial Disease Milestones), a concept that will be discussed later in this paper, and a table for Domain-Specific variables, shown below. They are shown with their two-letter domain codes. This was initially Table 2.2.11, but was changed to Table 2.2.12 in v1.6. This table shows a limited set of variables approved for use in only in a single domain. The scope of these variables may change in future versions of the SDTM, based upon SDTMIG implementation needs.

Variable Name	Variable Label
MHEVDTYP	Medical History Event Date Type
EXMETHOD	Method of Administration
EGBEATNO	ECG Beat Number
ICIMPLBL	Implantation Site Label
MSAGENT	Agent Name
MSCONC	Agent Concentration
MSCONCU	Agent Concentration Units

Version 1.6

Version 1.6 primarily added new tables that were limited to supporting the SEND Implementation Guide for Developmental and Reproductive Toxicology (SENDIG-DART). These included Subject Repro Stages, Trial Repro Stages, and Trial Repro Paths.

Version 1.7

Changes to v1.7 included the following:

- Table numbers changed. All v1.6 and earlier table numbers have a ".1" added.
- A Study Reference Section 5 was added, and Associated Persons modeling was moved to Section 6.

Additions in v1.7 included the Device-Subject Relationships (DR, Table 4.1.5.1) the Device Identifiers (DI, Table 5.1.1.1) datasets, both previously only in the SDTMIG for Medical Devices (SDTMIG-MD) Also added was the Non-Host Organism Identifiers Dataset (Table 5.1.2.1).

NEW SDTM VARIABLES

Since the publication of SDTM v.14, 42 new variables have been added to the SDTM. These are shown by version in the following tables. Many, however, are not intended for use in human clinical trials. These are shown with an single asterisk (*). Others are accompanied by a statement that have not been evaluated for use in human clinical trials and must therefore be used with extreme caution. These are identified with a double asterisk (**).

SDTM Version 1.5

There are eleven variables in the list below that can be used in human clinical trials.

SDTM Table	Variable Name	Variable Label
Table 2.2.1 Interventions	--USCHFL	Unscheduled Flag *
Table 2.2.2 Events	--USCHFL	Unscheduled Flag *
Table 2.2.3 Findings	--ORREF	Reference Result in Original Units
	--STREFC	Reference Result in Standard Format
	--STREFN	Numeric Reference Result in Std Units
	--IMPLBL	Implantation Site Label *
	--CHRON	Chronicity of Finding **
	--DISTR	Distribution Pattern of Finding **
	--LOBXFL	Last Observation Before Exposure Flag
	--USCHFL	Unscheduled Flag *
	--REPNUM	Repetition Number
Table 2.2.4 Identifiers	APID	Associated Persons Identifier
	FETUSID	Fetus Identifier *
	FOCID	Focus of Study Specific Interest
	--RECID	Invariant Record Identifier
Table 2.2.5 Timing Variables	--NOMDY	Nominal Study Day for Tabulations *
	--NOMLBL	Label for Nominal Study Day *
	MIDS	Disease Milestone Instance Name
	RELMIDS	Temporal Relation to Milestone Instance
	MIDSDTC	Disease Milestone Instance Date/Time

* *Not intended for use in human clinical trials.*

** *Not evaluated for use in human clinical trials and must therefore be used with extreme caution.*

Probably the most significant of these for most implementers is the Identifier, FOCID. FOCID has no domain prefix since it's used to describe a focus of specific interest (e.g., body location) the same way across all domains. The first use case was in the Ophthalmologic Examinations (OE) domain. In this example, FOCID was represented in EX to show the treatment each eye received. Observations on the eyes were then reported by FOCID in OE. Using the initial example from OE, the right eye might be treated (data in EX) and then evaluated, with results in OE. The OE domain has this assumption:

> *Whether or not FOCID is used in a study, --LOC and --LAT should be populated in records related to the eyes. The value in OELOC may be "EYE" but may also be a part of the eye, such as "RETINA", "CORNEA", etc.*

SDTM Version 1.6

In this version of the SDTM, only the variable NHOID is intended for use in human clinical trials. It's used to represent a sponsor-defined, intuitive name of the non-host organism being tested. Further details of the taxonomic classification are represented in the Non-host Organism Identifiers (OI) dataset (SDTM Table 5.1.2, SDTMIG Table 9.2). NHOID should be populated only with values representing what's known about the identity of the organism before the results of any tests are determined. It should never be used as a qualifier of any result.

SDTM Table	Variable Name	Variable Label
Table 2.2.3 Findings	--RESLOC	Result Location of Finding
Table 2.2.4 Identifiers	NHOID	Non-Host Organism Identifier
Table 2.2.5 Timing Variables	RPHASE	Repro Phase *
	RPPLDY	Planned Repro Phase Day of Observation *
	RPPLSTDY	Planned Repro Phase Day of Obs Start *
	RPPLENDY	Planned Repro Phase Day of Obs End *
	--RPDY	Actual Repro Phase Day of Observation *
	--RPSTDY	Actual Repro Phase Day of Obs Start *
	--RPENDY	Actual Repro Phase Day of Obs End *
Table 2.2.6 Demographics	RPATHCD	Planned Repro Path Code *

* *Not intended for use in human clinical trials.*

SDTM Version 1.7

The additions to v1.7 included two new variables for the Demographics dataset, and will be discussed under the following section on Additions and Changes to the SDTMIG. The other three additions are self-explanatory.

SDTM Table	Variable Name	Variable Label
Table 2.2.1.1 Interventions	--RSDISC	Reason for Treatment Discontinuation
Table 2.2.6.1 Demographics	ARMNRS	Reason Arm and/or Actual Arm is Null
	ACTARMUD	Description of Unplanned Actual Arm
Table 2.2.7.1 Comments	COEVALID	Evaluator Identifier
	CODY	Study Day of Comment

ADDITIONS AND CHANGES TO THE SDTMIG

PUBLICATION FORMAT

The biggest change, no doubt, is the publication format. Version 3.3 is published only as an HTML document. When saved as an HTML file, this author's version lost table formatting and the ability to horizontally scroll through wide tables. While the CDISC website states that implementers can print in a different format; however, when printed as PDF (even as a landscape orientation), this author's version fails to show the right side of wide tables.

This version is considerably longer then v3.2, but this is no surprise, since each version of the SDTMIG has been longer than the previous one. While v3.2 was 398 pages for the non-portfolio version that most people prefer, Version 3.2 can be almost 800 pages when printed as a landscape PDF, and more than 600 pages when printed as a portrait PDF. These numbers are for versions minus the full tables displaying, as noted in the previous paragraph.

Minor changes include the following:

- All values cited as examples in text and domain tables (CDISC Notes) now appear in quotation marks.
- Codelists in the domain specification tables now link to the specific codelist in the NCI-EVS website. While getting to the HTML file is relatively fast, getting to the specific codelist is very slow. It's actually faster to access the file (https://evs.nci.nih.gov/ftp1/CDISC/SDTM/SDTM%20Terminology.html) and then search for the codelist name.
- Hyperlinking to other sections within the document is more extensive.
- Previous references to "define.xml" or "define.xml file" have been changed to "Define-XML Document".
- Section numbering in Section 4 has been changed to remove the unnecessary second-level numbering.
- References to supplemental implementation guides (e.g., SDTMIG-MD, SDTMIG-PGx, SDTMIG-AP) have been added where appropriate.

UPDATES TO SECTIONS 1 (INTRODUCTION) AND 2 (FUNDAMENTALS OF THE SDTM)

- A new section (Section 1.4.1, How to Read a Domain Specification), provides clear descriptions of the columns in the domain specifications, and how these relate to the metadata described in the SDTM.
- Section 2.5, the table listing all the domains in v3.2 and previous versions was removed, as it was considered redundant.
- While not foremost in the minds of most sponsors, the policy on domain versioning was added in the new Section 2.5. Starting with SDTMIG v3.3, any new domain will be Version 1.0. An existing version that has changed since the last published version of the SDTMIG will be up-versioned. Note that some domains in SDTMIG v3.3 are domain versions 3.2 (unchanged) and others are v3.3 (up-versioned).

UPDATES TO SECTION 4, ASSUMPTIONS FOR DOMAIN MODELS

Section 4.5.3.2 Text Values Greater than 200

Clarified section on text values greater than 200 characters, and added a new table, a condensation of which is shown below.

General Observation Class & Supplemental Qualifier Variables	CO.COVAL and TS.TSVAL	TI.IETEST and IE.IETEST
The first 200 characters should be represented in the variable. Each additional 200 characters should be represented as a record in a SUPP-- dataset.	The first 200 characters should be represented in COVAL or TSVAL. Each additional 200 characters should be represented in COVAL1 (or TSVAL1) to COVALn (or TSVALn).	Put meaningful text in IETEST and describe the full text in the study metadata.
When splitting a text string, text should be split between words to improve readability.		Not applicable.
The value for QLABEL should be the original domain variable label.	The variable labels for COVAL1 to COVALn should be "Comment". The variable labels for TSVAL1 to TSVALn should be "Parameter Value".	Not applicable.

Section 4.1.3.1 Added Guidance on Populating the Epoch Variable

The text in Section 4.1.3.1 can be summarized as follows:

- Sponsors should not impute EPOCH values.
- The EPOCH value should be null if it's not possible to determine the EPOCH of an observation.
- Methods for assigning EPOCH values can be described in the Define-XML document.
- Since EPOCH is a study-design construct, it is not applicable observations prior to study start.
- EPOCH values may be determined as follows:

Most Findings	--DTC
Specimen Collection with End Dates	--ENDTC may be more appropriate
Events and Interventions	--STDTC

Section 4.1.5 Updated Guidance on Permissible Variables.

The following key points were added:

- Domain assumptions that say a variable is "generally not used" don't prohibit use of the variable.
- If a study includes a data item that would be represented in a Permissible variable, then that variable must be included in the SDTM dataset, even if null. Indicate no data were available for that variable in the Define-XML document.
- If a study did not include a data item that would be represented in a Permissible variable, then that variable should not be included in the SDTM dataset and should not be declared in the Define-XML document.

Section 4.4.7 Values for Relative Timing Variables, --STRF and --ENRF

The advice in various versions of the SDTMIG for the values used for Relative Timing Variables has varied. For some unknown reason, the text in v3.2 was changed to allow the two additional values (COINCIDENT and ONGOING) that were added to the codelist to support the use of --STRTPT and --ENRTPT to be used for --STRF and --ENRF. This was never the intention, so in v3.3 (Section 4.4.7) this was corrected and reverted back to the original values. A comparison of allowable values is shown in the following table.

Version	Allowable Values for –STRF and --ENRF
Versions 3.1.2 and 3.1.3	BEFORE, DURING, DURING/AFTER, AFTER, and U
Version 3.2	BEFORE, DURING, DURING/AFTER, AFTER, COINCIDENT, ONGOING, and U
Version 3.3	BEFORE, DURING, DURING/AFTER, AFTER, and U

NEW DOMAINS IN SECTIONS 5-9

Section	Domain Type	Domain Name and Code
5	Special Purpose	Subject Milestones (SM)
6.1	Interventions	Meal Data (ML)
		Procedure Agents (AG)
6.3	Findings	Cardiovascular Findings (CV)
		Musculoskeletal Findings
		Nervous System Findings (NV)
		Ophthalmic Examinations (OE)
		Respiratory System Findings (RE)
		Urinary System Findings
		Functional Tests (FT)
7	Trial Design	Trial Milestones (TM)
8	Relationships	RELSUB
9	Study References	Device Identifiers (DI)
		Non-host Organism Identifiers (OI)
		Pharmacogenomic/Genetic Biomarker Identifiers (PB)

CHANGES TO DEMOGRAPHICS

Removal of Population Flags

In a major change from previous versions of the SDTMIG, an assumption in the DM domain of v3.3 states that the Population Flags (COMPLT, FULLSET, ITT, PPROT, and SAFETY) should not be included in SDTM data. Values for these in the SUPPDM dataset often came from ADaM datasets, and as was the case with –BLFL would violate the traceability principle that all data in ADaM datasets must be traceable back to the SDTM datasets.

In addition to the change in the DM assumption, the corresponding example in Section 8 was removed, as were the QNAMs from Appendix C2. This change came slowly, again as a joint effort of the SDS and ADaM Teams over many years.

Population of ARMCD/ARM for Screen Failures and Subjects Not Assigned

Text in v3.2 stated that ARMCD (ARM) should be SCRNFAIL (Screen Failure) and NOTASSGN (Not Assigned). In v3.3, this changed, largely as a result of text in the FDA Study Data Technical Conformance Guide that states the following:

> *"Screen failures, when provided, should be included as a record in DM with the ARM, ARMCD, ACTARM, and ACTARMCD field left blank. For subjects who are randomized into a treatment group but not treated, the planned arm variables (ARM and ARMCD) should be populated, but actual treatment arm variables (ACTARM and ACTARMCD) should be left blank."*

Since the ARM and ARMCD values will not be populated in these two cases, the information that subjects were screen failures or not assigned will be represented in two new Expected variables added to v3.3. These are summarized below.

ARMNRS	Reason Arm and/or Actual Arm is Null	The reason that Arm variables (ARM and ARMCD) and/or actual Arm variables (ACTARM and ACTARMCD) are null.
ACTARMUD	Description of Unplanned Actual Arm	A description of actual treatment for a subject who did not receive treatment described in one of the planned trial Arms.

CHANGES IN THE FINDINGS GENERAL OBSERVATION CLASS

Approach to Morphology and Physiology

There are a number of changes in v3.3 that will affect the submission of morphology and physiology data. Prior to the publication of v3.2, there was considerable discussion by the SDS Team around the modeling philosophy of this

type of data. There were at least three team-wide polls that revolved around grouping vs. splitting data by body system. At the time when v3.2 was published, it was decided that morphology findings would be represented in the MO domain, and physiology findings would be in separate body-system-based domains. Since then, there were a number of TAUGs in which separating morphology and physiology endpoints became difficult. After additional discussion, it was decided that it would be easier if morphology data was also separated by body system. The SDTMIG v3.2 contains additional implementation advice around the use of the MO domain, which will be deprecated in a future version of the SDTMIG.

New body-system domains, intended for both morphology and physiology data, appearing in v3.3 include the following:

- Cardiovascular System Findings
- Musculoskeletal System Findings
- Nervous System Findings
- Ophthalmic Examinations
- Respiratory System Findings
- Urinary System Findings

These are grouped together in v3.3 with the Reproductive Findings domain that was published in v3.2. Preceding these domains in the IG is a new section called Generic Morphology/Physiology Specification that shows Required and Expected variables common to all morphology physiology domains. Although not stated in the text, it can be inferred that this would be a base for the creation of any custom (not yet modeled) body-system-based morphology/physiology domain.

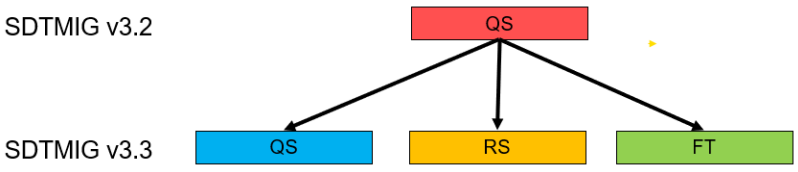
Findings Domains with Expanded Scope

Two domains that were in v3.2 had their scope expanded in v3.3. The Disease Response (RS) domain was expanded to include data that would have been in submitted in the Clinical Classifications (CC) domain, and was renamed Disease Response and Clin Classification. The CC domain that was part of the Batch 2 Public Review was not implemented.

The scope of the TU and TR domains has been broadened to include other types of lesions, rather than being limited to tumors. The domain names have been updated to include "Tumor/Lesion" to reflect the broadened usage. The documentation mentions the very broad definition of "lesion", which "can be almost any abnormal change involving any tissue or organ, usually due to disease or injury." Text and examples have been updated to include examples for Cardiovascular Lesions data and the representation of cysts for the Polycystic Kidney Disease.

Questionnaires, Rating, and Scales

For v3.3, it was decided by the SDS QRS Subteam that data previously represented in the QS domain should be split into three domains to more adequately represent the data that was collected. This is shown in the following diagram.



Many instruments and assessments that had been categorized as questionnaires were really not (since some had no questions). The categories of data included the following:

- QS – True questionnaire data (questions with responses)
- FT – Functional Tests (the subject is asked to perform a task and is evaluated
- RS – Response and Clinical Classifications (subject and objective assessments of a subject's status.

More details on the last two categories are provided in the New Domains section of this document.

Changes in Representing Baseline Flags

In previous versions of the SDTMIG, the variable –BLFL was used for representing baseline values. It's use, however, had two major disadvantages: 1) it was sponsor defined, and 2) it was sometimes populated with baseline values from the ADaM datasets. The latter, of course, resulted in a violation of the fundamental traceability principle that all data in the ADaM datasets must be traceable back to the SDTM datasets.

To address these issues, the SDS and ADaM Teams proposed a new variable, --LOBXFL (Last Observation Before Exposure Flag). This proposal was actually developed prior to the publication of SDTMIG v3.2, but getting approval from all the necessary parties went slowly. Success was achieved in time for v3.3. The new variable is defined as the last non-missing value prior to RFXSTDTC. This solves the two problems associated with –BLFL described above.

--LOBXFL has a Core value of Expected in the domains that have baseline values. In domains where --BLFL was present and Expected in v3.2, its Core value has been changed to Permissible in v3.3. Section 4.5.9 was added to explain the use of the new variable, showing a comparison between --LOBXFL, --BLFL, and the ADaM variable ABLFL.

Changes to Physical Exam

Text has been added to describe what constitutes a physical exam. For examinations of targeted body systems, findings should be represented in the appropriate body-system domain. Readers are told to refer to the CDASH Implementation Guide for additional collection guidance.

Changes to Microbiology Domains

A considerable number of changes have been made in the examples in this section. The most significant changes were made to MS, particularly the addition of the variables MSAGENT, MSCONC, and MSCONCU to represent predefined (as part of the testing) concentrations of a drug or agent. The examples SDTMIG v3.2 showed only data for when concentrations such as the minimal inhibitory were obtained as a result. MSAGENT is also used to represent drugs whose resistance is conferred through a genetic mutation. Data on the specific mutation can be represented in the PF domain, modeled in the SDTMIG for Pharmacogenomics (SDTMIG-PGx).

The example below shows testing for the presence of *Mycobacterium tuberculosis* in a subject (MB), and then the susceptibility testing of the microorganism to two different drugs, shown in MSAGENT. Only those columns needed to clearly illustrate the example are shown. The two datasets are linked via MBREFID and MSREFID.

mb.xpt

MBREFID	MBTEST	MBTSTDTL	MBORRES	MBORRESU	MBSTRESC
100.2	Mycobacterium Tuberculosis Complex	DETECTION	PRESENT		PRESENT

ms.xpt

MSREFID	MSTESTCD	MSTEST	MSAGENT	MSORRES	MSSTRESC
100.2	MICROSUS	Microbial Susceptibility	Rifampicin	RESISTANT	RESISTANT
100.2	MICROSUS	Microbial Susceptibility	Isoniazid	SUSCEPTIBLE	SUSCEPTIBLE

relrec.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
ABC	MB		MBREFID		ONE	C
ABC	MS		MSREFID		MANY	C

CHANGES IN THE EVENTS GENERAL OBSERVATION CLASS

Assumptions in the Disposition Domain

Some of the assumptions in v3.2 and previous version were too restrictive, and sponsors ended up not following them. An example was that EPOCH should not be populated when DSCAT = "PROTOCOL MILESTONE". Following this prevented a sponsor from accurately representing multiple informed consents, which are Protocol Milestones. In 3.3, this restriction was removed. Version 3.3 expanded the assumptions to include the use of DSSCAT to distinguish between study treatment and study participation. Many sponsors had been including both types of events, and now there's a standard way to do so.

ADDITION OF SECTION 9, STUDY REFERENCES

A Study Reference section (Section 9) was added in v3.3. It includes the following datasets, the first two of which are also present in the SDTM v1.7.

- Device Identifiers (DI). This was previously published in the SDTMIG-MD as a Special-Purpose domain in v1.0 and as a Study Reference in v1.1.

- The Non-host Organism Identifiers (OI). This domain is essentially a lookup table, created to provides further taxonomic classification data for the NHOID variable in the Findings general observation class. Since parameters of interest may vary by the type of organism (e.g., viruses, bacteria, fungi), sponsors can include as many as needed for clear identification.
- Pharmacogenomic/Genetic Biomarker Identifiers (PB). PB was first Introduced as part of the SDTMIG-PGx v1.0 as a Special-Purpose domain. It serves as a reference to associate observed genetic mutations with medical conclusions (e.g. disease diagnosis, drug resistance.

HIGHLIGHTS OF NEW DOMAINS

Almost all of the domains described below came about as the result of the development of Therapeutic-Area User Guides (TAUGs). Domains are arranged alphabetically by domain name (and not domain code).

Cardiovascular System Findings (CV)

This is a physiology Findings domain which was developed as part of the Therapeutic Area User Guide for Cardiovascular Disease (TAUG-CV). Cited examples of assessments submitted in this domain include coronary artery dominance, ischemic myocardium percentage, coronary artery dissection grade, and degree of stenosis. CV follows the traditional Findings data structure, and contains no new variables. Considerable controlled terminology for cardiac endpoints has been developed by CDISC.

Disease Response and Clin Classifications (RS)

As noted above, the clinical classification data is under the QRS umbrella. RS data will be managed in the same way as the QS and FT domains with the Category (--CAT) variable indicating the type of measurements or tests. RSCAT is required for clinical classifications other than oncology response criteria.

The clinical classification part of RS is for named instruments that serve to classify, rank, or grade the status of a disease status or other physiological or biological status. The output may be either an ordinal or categorical score. Classifications are based on observable findings by an investigator or other health professional. Some consist of composite scores based on multiple findings that may be found in other SDTM domains such as Laboratory Test Results (LB), Vital Signs (VS), or Clinical Events CE). Examples include the Child-Pugh Score, APACHE, and NYHA Class. An example of the Child Pugh Score has been published previously (4).

Row	STUDYID	DOMAIN	USUBJID	RSSEQ	RSLNKID	RSTESTCD	RSTEST
1	2014-0987	RS	2014-0987-0010	1	CPLB1	CPS0103	CPS01-Bilirubin
2	2014-0987	RS	2014-0987-0010	2	CPLB2	CPS0104	CPS01-Serum Albumin
3	2014-0987	RS	2014-0987-0010	3	CPLB3	CPS0105B	CPS01-PT, INR
4	2014-0987	RS	2014-0987-0010	4	CPPE1	CPS0102	CPS01-Ascites
5	2014-0987	RS	2014-0987-0010	5		CPS0101	CPS01-Encephalopathy
6	2014-0987	RS	2014-0987-0010	6		CPS0106	CPS01-Total Score
7	2014-0987	RS	2014-0987-0010	7		CPS0107	CPS01-Grade

Row	RSCAT	RSORRES	RSORRESU	RSSTRESC	RSSTRESN
1	CHILD-PUGH CLASSIFICATION	>3	mg/dL	3	3
2	CHILD-PUGH CLASSIFICATION	<2.8	g/dL	3	3
3	CHILD-PUGH CLASSIFICATION	1.7-2.30		2	2
4	CHILD-PUGH CLASSIFICATION	Moderate to Severe		3	3
5	CHILD-PUGH CLASSIFICATION	Grade I-II (or suppressed with medication)		2	2
6	CHILD-PUGH CLASSIFICATION	10		13	13
7	CHILD-PUGH CLASSIFICATION	C		SEVERE	

A few things worthy of note in this typical modeling:

- RSTESTCD, RSTEST, and RSCAT all have CDISC Controlled Terminology.
- RSORRES contains the classification range or result of the assessment. In Row 1, RSORRES contains the range into which the subject's serum bilirubin fell. The actual value for the serum bilirubin would be in the LB domain. The RS and LB records are related via the --LNKID variable, along with a record in RELREC.
- RSSTRESC contains the actual score for the range or result described in RSORRES. When this is numeric, it is copied into RSSTRESN.

- Row 6 contains the composite score of Rows 1-5, and Row 7 contains the classification (RSORRES) and standardization of that (RSSTRESC).

Clinical Classifications for which CDISC Controlled Terminology exists can be found at cdisc.org.

Functional Tests (FT)

As noted above, the functional-test data is under the QRS umbrella. This domain is intended for named tests that evaluate a subject's functional capacity. Included are tests for mobility, dexterity, and cognitive ability. Characteristics of functional tests:

- They have documented methods for administration and analysis, and require a subject to perform specific activities that are evaluated and recorded.
- They are an objective measurement of the performance of the task by the subject in a specific instance. Most often, they are quantitative measurements.
- As with questionnaires in QS, they may be documented in the public domain or be owned by a copyright holder. Examples of functional tests include the 25-Foot Walk Test, the 9-Hole Peg Test, and the Rey Auditory Verbal Learning Test (AVLT).

The modeling of the FT domain is consistent with that for questionnaires represented in the QS domain. Functional tests for which CDISC Controlled Terminology exists can be found at cdisc.org.

Meals (ML)

This domain debuted in the Diabetes TAUG. It is used for the submission of meal data. Because this is an Interventions domain, the focus is on the type and timing of ingested meals. Data about meal amounts and composition would need to be submitted in either TS (if the same for all subjects) or in a Findings About domain if unique to each subject.

Nervous System Findings (NV)

This domain was first introduced as part of the TAUG for Multiple Sclerosis. It is a physiology domain intended for the representation of results from active neurological processes. Some evaluations may occur as the results of a procedure. If the information about the procedure is important, it can be represented in the PR domain. One example of data that would be included in this domain is glucose metabolism by various regions of the brain, assessed by using radiotracers and PET scans. Another is the Visual evoked potential (VEP), which assesses a subject's response to a visual stimulus via an EEG. The modeling of this data does not differ from that of most Findings domains, and no new variables were needed.

Ophthalmic Examinations (OE)

The OE domain is a Findings domain used for tests that measure a person's ocular health and visual status. Included are tests of visual acuity, color vision, ocular comfort (e.g., dryness, itching), and intraocular pressure. Excluded are morphological measurements such as pupil diameter and macula thickness.

The variable FOCID (Focus of Study-Specific Interest) appeared SDTM v1.5 as a result of OE, as well as four domains in the SENDIG v3.1. FOCID is an Identifier variable which has no domain prefix. This is because the focus of specific interest within the subject should be identified the same way across all domains. For example, the right eye (FOCID = OD) might be treated (EX) and then evaluated, with results in OE. This domain uses controlled terminology for FOCID: OD (Oculus Dexter, Right Eye), OS (Oculus Sinister, Left Eye), and OU (Oculus Uterque, Both Eyes).The Findings variables --LOC (e.g., EYE) and --LAT (e.g., RIGHT), and to a lesser extent, --DIR, and --PORTOT may also be used. In fact, the Assumptions state that "the variables --LOC and --LAT are recommended to be populated in all cases for ophthalmic findings, since the benefits of facilitating grouping and data aggregation for other needs are recognized."

Procedure Agents (AG)

This Interventions domain is used to represent agents administered to the subject as part of a procedure, as opposed to drugs, medications and, therapies administered with therapeutic intent. Examples so far have included a short-acting bronchodilator administered as part of a reversibility assessment for asthma, glucose or meals administered as part of a tolerance test in subjects with diabetes, and contrast agents and radio-labeled substances used in imaging studies.

This domain has also been used to represent challenge-agent exposure, modeled in the SENDIG-AR (SEND Implementation Guide for Animal Rules studies. For more information on these types of studies, see Wood (4).

Respiratory System Findings (RE)

The Respiratory Systems Findings domain was first created for the Asthma TAUG. It is used for data related to physiological findings related to the respiratory system, including the organs that are involved in breathing such as the nose, throat, larynx, trachea, bronchi and lungs. Examples of data collected in this domain include forced expiratory volume in one second (FEV1) and forced vital capacity (FVC).

This domain introduced the concept of a reference value, as opposed to a reference range defined by the pairs --ORNRLO and --ORNRHI, and by --STNRLO and STNRHI. This is because FEV1 (forced expiratory volume in 1 second), FVC (forced vital capacity), and the FEV1/FVC ratio have reference values (not ranges) that are based upon age, race, sex, and height. Three reference value variables were created: --ORREF (Reference Result in Original Units), --STREFC (Reference Result in Standard Format), and --STREFN (Numeric Reference Result in Std Units). These were added to SDTM v1.5, as noted previously.

DISEASE MILESTONES

The concept of Disease Milestones arose in the context of representing information collected around a hypoglycemic event in diabetes trials (Diabetes TAUG), but has applicability to any type of event-driven data collection. This is in contrast to collection driven by a schedule (e.g., visits, daily diaries). The event driving the data is referred to as a Disease Milestone (variable name MIDSTYPE), and is described at the trial level in the Trial Milestones dataset (tm.xpt), shown below. The last column, MIDSRPT, indicates whether the Milestone repeats (there could be more than one per subject in the trial).

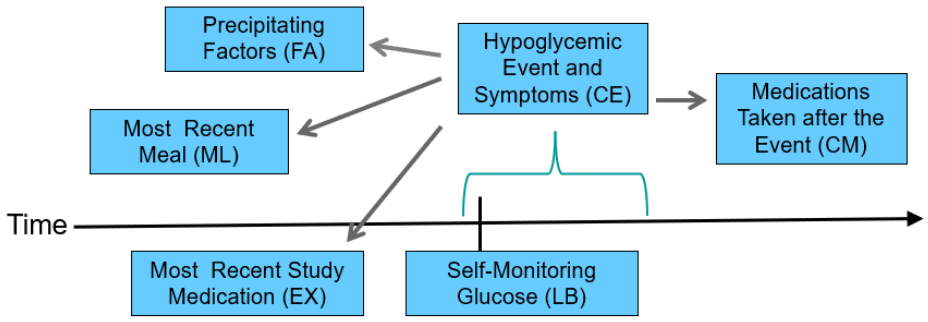
STUDYID	DOMAIN	MIDSTYPE	TMDEF	TMRPT
ABC	TM	HYPOGLYCEMIC EVENT	Hypoglycemic Event, the occurrence of a blood glucose concentration below the specified (by study) level of hypoglycemia	Y

The Milestone serves as an anchor around which information leading up to the event, information about the event, assessments made during the event, and information subsequent to the event revolve. The Timing variable, RELMIDS, contains information of the timing relative to the event, with typical values of BEFORE, DURING, and AFTER, but other values may be used to convey additional meaning. This topic was covered in this author's paper and presentation (5), when the concept had not yet undergone public review. Much of what appears has been condensed from that paper.

To better understand this Disease Milestones concept, an example follows, using hypoglycemic events as the Disease Milestone. The information typically collected around hypoglycemic events includes the following (with the domains shown in parentheses), also illustrated in Figure 1:

- Data about the event and prespecified symptoms (Clinical Events, CE)
- Questions around any precipitating factors (Findings About, FA)
- Blood glucose (self-monitored) at the time of the event (Labs, LB)
- Last dose of study medication (e.g., insulin or an analog) prior to the event (Exposure, EX)
- Last meal prior to the event (Meals, ML)
- Whether any hypoglycemic medications were taken after the event, along with a pre-specified list (Concomitant Medications, CM).

Figure 1. Data Collection for a Hypoglycemic Event



Example data for these domains is shown in Appendix 1.

The use of the Disease Milestones concept allows for consistent mapping across all domains that have data related to each hypoglycemic event, as shown by the yellow (horizontal lines) and blue (vertical lines) shading. Because of this consistency, RELREC is not needed.

In much in the same way that the Trial Elements and Trial Visits have corresponding subject-level domains, Trial Milestones has a corresponding subject-level domain, Subject Milestones (sm.xpt), an example of which is shown below. Although the focus in this section has been on a single hypoglycemic event, the subject milestones table will contain a record for every hypoglycemic event experienced by the subject in the trial. For illustrative purposes, the subject milestones table below shows that the subject had two hypoglycemic events.

STUDYID	DOMAIN	USUBJID	SMSEQ	MIDS	MIDSTYPE	SMSTDTC	SMENDTC	SMSTDY	SMENDY
ABC	SM	ABC-1001	2	HYPO 1	HYPOGLYCEMIC EVENT	2013-09-01T11:00	2013-09-01T11:00	25	25
ABC	SM	ABC-1001	3	HYPO 2	HYPOGLYCEMIC EVENT	2013-09-24T08:48	2013-09-24T08:48	50	50

CONCLUSIONS

The development of SDTM-based standards is occurring at a relatively rapid rate. This paper has presented a summary of the new domains, new concepts, and new variables that have been developed for SDTMIG v3.3 and the SDTM versions created since v1.4. Nine new general-observation class domains have been added since SDTMIG v3.2. Seventeen new variables have been added for use in SDTMIG domains. The Disease Milestones concept resulted in the creation of three new Timing variables and two new domains to more clearly represent event-driven observations of interest.

REFERENCES

1. Wood, F. (2008) The CDISC Study Data Tabulation Model (SDTM): History, Perspective, and Basics. PharmaSUG Proceedings, May 2008. http://www.lexjansen.com/pharmasug/2008/rs/RS10.pdf
2. Wood, F., and Guinter, T. (2008) Evolution and Implementation of the CDISC Study Data Tabulation Model (SDTM). Pharmaceutical Programming 1 (1): 20-27.
3. Wood, F. Wold, D., Facile, R., and Kubick, W. (2014). Data Standards Development for Therapeutic Areas: A Focus on SDTM-Based Datasets. PharmaSUG Proceedings, June 2014. http://www.pharmasug.org/proceedings/2014/DS/PharmaSUG-2014-DS05.pdf
4. Wood, F. (2017). Considerations in Submitting Standardized Electronic Data Under the Animal Rule: The Use of Domains in the SDTMIG and the SENDIG. PharmaSUG Proceedings, May 2017. http://www.pharmasug.org/proceedings/2017/DS/PharmaSUG-2017-DS14.pdf
5. Wood, F. (2015). What to Expect in SDTMIG v3.3. PharmaSUG Proceedings, June 2015. http://www.pharmasug.org/proceedings/2015/DS/PharmaSUG-2015-DS14.pdf

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Fred Wood
Enterprise: Data Standards Consulting Group, A Division of TalentMine.
Work Phone: 484-881-2297
E-mail: fred.wood@datastandardsconsulting.com

Appendix 1
Disease Milestones Example Data for Hypoglycemic Events *

ce.xpt

STUDYID	DOMAIN	USUBJID	CESEQ	CETERM	CECAT	CEPRESP	CEOCCUR	CESTDTC	MIDS	RELMIDS	MIDSDTC
ABC	CE	ABC-1001	1	HYPOGLYCEMIA	HYPOGLYCEMIA			2013-09-01T11:00	HYPO 1		2013-09-01T11:00
ABC	CE	ABC-1001	2	SWEATING	HYPOGLYCEMIA	Y	Y		HYPO 1	DURING	2013-09-01T11:00

fa.xpt

STUDYID	DOMAIN	USUBJID	FASEQ	FATESTCD	FATEST	FAOBJ	FAORRES	MIDS	MIDSREL	MIDSDTC
ABC	FA	ABC-1001	2	POSSCAUS	Possible cause identified	HYPOGLYCEMIA	Y	HYPO 1	PRIOR TO EVENT	2013-09-01T11:00
ABC	FA	ABC-1001	3	MEALCAUS	Missed or delayed meal a possible cause	HYPOGLYCEMIA	Y	HYPO 1	PRIOR TO EVENT	2013-09-01T11:00
ABC	FA	ABC-1001	4	PACAUS	Physical activity a possible cause	HYPOGLYCEMIA	N	HYPO 1	PRIOR TO EVENT	2013-09-01T11:00
ABC	FA	ABC-1001	5	ALCCAUS	Alcohol a possible cause	HYPOGLYCEMIA	N	HYPO 1	PRIOR TO EVENT	2013-09-01T11:00

lb.xpt

STUDYID	DOMAIN	USUBJID	SPDEVID	LBSEQ	LBTESTCD	LBTEST	LBORRES	LBORRESU	LBSTRESC	LBSTRESN	LBSTRESU	LBSPEC	LBDTC	MIDS	RELMIDS
ABC	LB	ABC-1001	GLUCOMETER	1	GLUC	GLUCOSE	60	mg/dL	3.33	3.33	mmol/L	BLOOD	2013-09-01T11:00	HYPO 1	DURING

MIDSDTC is not shown due to space limitations.

ml.xpt

STUDYID	DOMAIN	USUBJID	MLSEQ	MLTRT	MIDS	RELMIDS	MIDSDTC
ABC	ML	ABC-1001	1	EVENING MEAL	HYPO 1	LAST INTERVENTION PRIOR TO	2013-09-01T11:00

ex.xpt

STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXCAT	EXDOSE	EXDOSU	EXSTDTC	MIDS	RELMIDS	MIDSDTC
ABC	EX	ABC-1001	1	DRUG A	HIGHLIGHTED DOSE	10	mg	2013-09-01T07:00	HYPO 1	LAST INTERVENTION PRIOR TO	2013-09-01T11:00

cm.xpt

STUDYID	DOMAIN	USUBJID	CMSEQ	CMTRT	CMCAT	CMSCAT	CMPRESP	CMOCCUR	MIDS	RELMIDS	MIDSDTC
ABC	CM	ABC-1001	1	HYPOGLYCEMIC TREATMENTS	HYPOGLYCEMIC TREATMENTS		Y	Y	HYPO 1	IMMEDIATLEY AFTER	2013-09-01T11:00
ABC	CM	ABC-1001	4	GLUCOSE TABLETS	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	Y	HYPO 1	IMMEDIATLEY AFTER	2013-09-01T11:00
ABC	CM	ABC-1001	5	GLUCAGON INJECTION	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	N	HYPO 1	IMMEDIATLEY AFTER	2013-09-01T11:00
ABC	CM	ABC-1001	6	INTRAVENOUS GLUCOSE	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	N	HYPO 1	IMMEDIATLEY AFTER	