PharmaSUG2015 - Paper DS14

What to Expect in SDTMIG v3.3

Fred Wood, Accenture Accelerated R&D Services, Berwyn, PA

ABSTRACT

The next version of the SDTMIG (version 3.3) is expected in Q3 of this year. The content has been posted for review in three batches. The rational for this posting process for versions of the SDTMIG and the SDTM will be described. A number of new domains, variables, and concepts that will be included in SDTMIG v3.3 will be discussed. Included are the physiology-based Findings domains, (Nervous System Findings (NV), Ophthalmic Examinations (OE), Respiratory Measurements (RE), and Cardiovascular Findings (CV); and the new Interventions domains, Procedure Agents (AG) and Meals (ML). Functional Tests (FT) and Clinical Classifications (CC) are two new Findings domains that will be managed in the same way the Questionnaires (QS) domain is, with the Category variable indicating the type of measurements or tests. The definitions of the Tumor domains (TU, TR) have been expanded to include non-tumor lesions.

The paper will present an overview of several new concepts. Included will be Disease Milestones, which includes new variables and two new special-purpose domains, first introduced in the Diabetes Therapeutic-Area User Guide (TAUG). The option of including non-standard variables in parent domains, thus avoiding the need for SUPP--datasets, was sent out for public comment, and the status of that will be discussed. A new variable for the focus of interest within a subject (FOCID), which has the same meaning across all domains, was first used in the OE domain, but also has applicability to SEND data. Additional potential efforts by the SDS Team for v3.3 that will be covered here include methods for managing multiple enrollments of the same USUBJID, clarification on the use of Findings About, an approach for the standardization of lab units, and developing a standard method for representing multiple ATC codes for concomitant medications.

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 Data Standards 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 development of the SDTM and SDTMIG, with a version history and number of domains modeled, is summarized in the table below. It is apparent from the table that the number of new domains in the SDTMIG v1.4 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	SDTMIG Version	Year	Number of Domains*
1.0	3.1	2004	23
1.1	3.1.1	2005	30
1.2	3.1.2	2008	32
1.3	3.1.3	2012	32
1.4	3.2	2013	42
TBD**	3.3	2015	50***

* Includes all special-purpose and general-observation-class domains
** The SDTM supports a number of implementation guides. As each is updated, a new version of the SDTM is also created.
*** Planned as of April 1, 2015.

Since SDTMIG v3.2, the SDS Team has been releasing SDTMIG updates for public comment in batches. This is easier for the Team to manage, and easier for public review. Updates to the SDTMIG v3.3 have been planned to be released in three batches. Batches 1 and 2 have already been through public review, and the comments from those

resolved. Batch 3 is scheduled for some time in Q2 of 2015. A summary of the new domains, new variables, and new concepts appears in the table below, each with the batch number in which it appeared or will appear.

Table 1. New Domains, Concepts, and Variables Planned for SDTMIG v3.3

NEW DOMAINS	Batch
Cardiovascular System Findings (CV)	1
Clinical Classifications (CC)	2
Functional Tests (FT)	1
Meals (ML)	3
Nervous System Findings (NV)	2
Ophthalmic Examinations (OE)	2
Procedure Agents (AG)	1
Respiratory System Findings (RE)	1
Tumor Identification and Tumor Measurements Update (TU, TR)	2
Subject Milestones (SM)	1
Trial Milestones (TM)	1
NEW CONCEPTS	
Disease Milestones	1
Submitting Non-Standard Variables in Main Domains	2
EMERGING CONCEPTS (Potential Inclusion)	
Methods for Managing Multiple Enrollments of the Same USUBJID	3
Clarification on Use of FA	3
Standardization of Lab Units	3
A Standard Method for Submitting CM ATC Codes	3

NEW VARIABLES		
Observation Class or Role	**Name**	**Batch**
Events	--EVAL	2
	--EVALID	2
	--ACPTFL	2
Findings	--NSPCES	2
	--NSTRN	2
	--ORREF	1
	--STREFC	1
	--STREFN	1
	--CHRON	2
	--DISTR	2
	--REPNUM	2
	--CLMETH	2
	--LOBXFL	3
Identifiers	APID	2
	FOCID	2
	--RECID	2
Timing	MIDS	1
	RELMIDS	1
	MIDSDTC	1

HIGHLIGHTS OF NEW DOMAINS PLANNED FOR SDTMIG V3.3

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.

CLINICAL CLASSIFICATIONS (CC)

Clinical Classifications, along with Functional Tests (below), will be managed in the same way that the Questionnaires (QS) domain is, with the Category variable indicating the type of measurements or tests.

Clinical Classifications 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 is shown below.

Row	STUDYID	DOMAIN	USUBJID	CCSEQ	CCLNKID	CCTESTCD
1	2014-0987	CC	2014-0987-0010	1	CPLB1	CPS0101
2	2014-0987	CC	2014-0987-0010	2	CPLB2	CPS0102
3	2014-0987	CC	2014-0987-0010	3	CPLB3	CPS0103
4	2014-0987	CC	2014-0987-0010	4	CPPE1	CPS0104
5	2014-0987	CC	2014-0987-0010	5		CPS0105
6	2014-0987	CC	2014-0987-0010	6		CPS0106
7	2014-0987	CC	2014-0987-0010	7		CPS0107

Row	CCTEST	CCCAT	CCORRES	CCORRESU	CCSTRESC	CCSTRESN
1	CPS01-Bilirubin	CHILD-PUGH SCORE	>3	mg/dL	3	3
2	CPS01-Serum Albumin	CHILD-PUGH SCORE	<2.8	g/dL	3	3
3	CPS01-PT INR	CHILD-PUGH SCORE	1.7-2.30		2	2
4	CPS01-Ascites	CHILD-PUGH SCORE	Moderate to Severe		3	3
5	CPS01-WH Hepatic Encephalopathy	CHILD-PUGH SCORE	Grade I-II (or suppressed with medication)		2	2
6	CPS01-Child-Pugh Score	CHILD-PUGH SCORE	10		13	13
7	CPS01-Child-Pugh Class	CHILD-PUGH SCORE	C		SEVERE	

A few things worthy of note in this typical modeling:

- CCTESTCD, CCTEST, and CCCAT all have CDISC Controlled Terminology.
- CCORRES contains the classification range or result of the assessment. In Row 1, CCORRES 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 CC and LB records are related via the --LNKID variable, along with a record in RELREC.
- CCSTRESC contains the actual score for the range or result described in CCORRES. When this is numeric, it is copied into CCSTRESN.
- Row 6 contains the composite score of Rows 1-5, and Row 7 contains the classification (CCORRES) and standardization of that (CCSTRESC).

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

FUNCTIONAL TESTS (FT)

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 timing and quantity of ingested meals. Data about meal 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. This domain has yet to go out for public review, but is planned for Batch 3.

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) will appear in SDTM v1.5 as a result of this domain, as well as several domains expected for the SENDIG. 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."

Other implementations of FOCID are expected to use protocol-defined terminology. More information on the FOCID variable can be found in the section below, Variables Appearing in the Next Version of SDTM v1.5

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.

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). More information on these variables can be found in the section below, Variables Appearing in the Next Version of SDTM v1.5.

TU AND TR DOMAINS – BROADENED SCOPE

With SDTMIG v3.3, 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.

As before, the Tumor/Lesion Identification domain (TU) represents data that identifies the anatomical location of tumor(s) or lesion(s), and the Tumor/Lesion Results (TR) domain is used for representing the quantitative or qualitative assessments of the identified tumor(s) or lesion(s). A third domain, Disease Response (RS), is used to provide a determination of response to therapy and has a practical application in oncology studies when used in conjunction with the TU and TR domains.

NEW CONCEPTS APPEARING IN SDTMIG V3.3

REPRESENTATION OF NON-STANDARD VARIABLES IN MAIN DOMAINS

The SDTM Supplemental Qualifiers structure is a method for representing non-standard variables (NSVs) in a standard way within the SDTM. The format for representing NSVs (those variables not found in SDTM Tables 2.2.1-2.2.5) has been as separate SUPP-- datasets that are associated with the corresponding "parent" general-observation-class and Demographics datasets.

Batch 2 of the material planned for SDTMIG v3.3 included the proposal for alternative to submitting NSVs in separate SUPP-- datasets. Text has been written to update Section 8 of the SDTMIG, and pending final agreement from the FDA, the alternative will be documented in v3.3. The alternative method consists of representing NSVs in the main ("parent") domains. Aside from the obvious benefit of eliminating the work involved in splitting off NSVs into separate datasets, this method provides some additional benefits. These include the following:

- Permitting direct viewing of standard variables and NSVs within the same structure, eliminating the need for tools or the writing of programs to display the data together.
- Eliminating some current SUPP-- structural limitations by allowing:
  - Numeric NSVs to be represented in a numeric data type
  - Character NSVs to be defined with an appropriate length for each variable, rather than the typical default of $200 for QVAL
- Allow metadata for NSVs (including Controlled Terminology) to be applied at the variable level instead of the value-level.

Specific implementation rules exist in order to ensure that the number and nature of NSVs is no different than it would have been under the existing method of using SUPP-- datasets.

DISEASE MILESTONES

The concept of Disease Milestones arose in the context of representing information collected around a hypoglycemic event in diabetes trials, 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 MIDS). This 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 last year (3), 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):

- Data about the event and prespecified symptoms (Clinical Events, CE)
- 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)The mapping of example data for this approach is shown in Figure 1. A larger, landscape version can be found in Appendix 1. Recall that in this model, everything is based upon the following:
- A reference event, referred to as a disease milestone (MIDS, shaded in yellow)
- The relationship of other observations relative to that event (RELMIDS, shaded in blue).

In some cases the date of the Disease Milestone (MIDSDTC) may also be collected. More information on these variables can be found in the section below, Variables Appearing in the Next Version of SDTM v1.5.

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

Figure 1. Disease Milestones Concept

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
ABC	FA	ABC-1001	2	POSSCAUS	Possible cause identified	HYPOGLYCEMIA	Y	HYPO 1	PRIOR TO EVENT
ABC	FA	ABC-1001	3	MEALCAUS	Missed or delayed meal a possible cause	HYPOGLYCEMIA	Y	HYPO 1	PRIOR TO EVENT
ABC	FA	ABC-1001	4	PACAUS	Physical activity a possible cause	HYPOGLYCEMIA	N	HYPO 1	PRIOR TO EVENT
ABC	FA	ABC-1001	5	ALCCAUS	Alcohol a possible cause	HYPOGLYCEMIA	N	HYPO 1	PRIOR TO EVENT

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

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
ABC	CM	ABC-1001	1	HYPOGLYCEMIC TREATMENTS	HYPOGLYCEMIC TREATMENTS		Y	Y	HYPO 1	IMMEDIATLEY AFTER
ABC	CM	ABC-1001	4	GLUCOSE TABLETS	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	Y	HYPO 1	IMMEDIATLEY AFTER
ABC	CM	ABC-1001	5	GLUCAGON INJECTION	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	N	HYPO 1	IMMEDIATLEY AFTER
ABC	CM	ABC-1001	6	INTRAVENOUS GLUCOSE	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	N	HYPO 1	IMMEDIATLEY AFTER

Another part of the Disease Milestone concept is the idea that these can be defined at the trial level, as shown in the following example. The last column, MIDSRPT, indicates whether the Milestone repeats.

Trial Milestones (tm.xpt)

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

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, 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.

Subject Milestones (sm.xpt)

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

EMERGING CONCEPTS POTENTIALLY APPEARING IN SDTMIG V3.3

METHODS FOR MANAGING MULTIPLE ENROLLMENTS OF THE SAME USUBJID

A PhUSE Working Group on Optimizing Data Standards, Data Standards Implementation has developed a proposal for managing data for unique subjects (single USUBJID) with multiple enrollments in the same trial. This may include where 1) screen failures are allowed to be rescreened, and the sponsor has decide to submit screen-failure data, and 2) subjects are allowed to complete a trial and then re-enroll after some waiting period. More information can be

found at http://www.phusewiki.org/wiki/index.php?title=USUBJID. The proposal recommends the use of multiple Demographics records for the same USUBJID, each unique by SUBJID. In addition, SUBJID would, at a minimum, need to appear on any domain records where data from the same USUBJID was represented more than once (e.g., a subject with two Visit 1 serum glucose records, two months apart from each other). While this proposal is still under discussion, the next step will be to determine what changes might need to be made in the wording of the SDTMIG for its implementation.

CLARIFICATION ON USE OF FINDINGS ABOUT

The concept of Findings About was introduced in SDTMIG v3.1.2. Since that time, there has been much confusion as to when it should be implemented. The SDS Team is working on a decision tree that is expected to help sponsors decide when to use Findings About, when to use a custom Findings domain, and when to use Supplemental Qualifiers.

STANDARDIZATION OF LAB UNITS

When the SDTM was first created, it was envisioned that standardization would be required to a single set of units for each test, either to conventional units or SI units. As a result, there is only one set of standardized-result variables, --STRESC, --STRESN, and --STRESU. As implementation of the SDTM has grown, and sponsors are 1) preparing data for potential submission to regulatory bodies other than the FDA, and 2) are attempting to meet the needs of specific reviewers or Review Divisions, the need to provide lab units that are standardized to both conventional and SI units has arisen. A recommendation for how to do this, taking into account the potential of FDA review tools to create alternate displays of the data, may be necessary.

A STANDARD METHOD FOR REPRESENTING MULTIPLE ATC CODES FOR CONCOMITANT MEDICATIONS

The SDTMIG provides some advice on how to manage the submission of multiple ATC codes per drug. It states, “If using a dictionary and coding to multiple classes, then follow Section 4: 4.1.2.8.3, Multiple Values for a Non-Result Qualifier Variable or omit CMCLASCD.” This results in sponsors either populating CMCLASCD (and CMCLAS) with “MULTIPLE” or leaving them null and, in many cases, using SUPPCM to represent the ATC Codes and the respective levels within those codes. There has been no standard method for how to represent the data in SUPPCM, with the main issues being 1) choosing the values in QNAM, and 2) determining how many levels of the ATC Codes are needed. A proposal was drafted by representatives at the Uppsala Monitoring Center in 2011, but it has never been finalized. Progress seems to halt, rightfully so, with a number of important questions:

- One question that continually arises, and that seems to halt progress, is whether all this information needs to be submitted at the subject level.
- Could a file containing all the concomitant medications taken by all subjects in a trial and the ATC codes be submitted? This file could be merged with the CM dataset?
- What is the optimal structure for the submission dataset when ATC Codes are included? For example, could it be one record per concomitant medication or one record per ATC code? We may not get universal agreement on this since some sponsors choose a single ATC code and some sponsors analyze data for all ATC codes a medication may have.

Because of the complexity of this topic, it is likely that it may not be resolved by the time the SDS Team is ready to finalize SDTMIG v3.3. It may benefit everyone, however, if some progress can be made.

NEW VARIABLES APPEARING IN THE SDTM V1.5

EVENTS

--EVAL (Previously only in Findings)	Evaluator	Char	Record Qualifier	Role of the person who provided an evaluation. Used only for results that are subjective (e.g., assigned by a person or a group). Examples: ADJUDICATION COMMITTEE, INDEPENDENT ASSESSOR, RADIOLOGIST.
--EVALID (Previously only in Findings)	Evaluator Identifier	Char	Variable Qualifier of --EVAL	Used to distinguish multiple evaluators with the same role recorded in --EVAL. Examples: RADIOLOGIST1, RADIOLOGIST2.
--ACPTFL	Accepted Record Flag	Char	Record Qualifier	In cases where more than one assessor provides an evaluation of a result or response, this flag identifies the record that is considered to be the accepted evaluation. Expected values can include Y, N or null. This is not intended to be an analysis flag to indicate acceptability for a given analysis

FINDINGS

--NSPCES (Added for the Pharmacogenomics IG)	Non-Host Organism Species	Char	Record Qualifier	Biological classification for a non-host organism. Examples: STAPHYLOCOCCUS AUREAS, HCV, HIV, PLASMODIUM FALCIPARUM.
--NSTRN Added for the Pharmacogenomics IG)	Non-Host Organism Strain	Char	Record Qualifier	A subtype of a non-host organism. Examples: 1a, 1b (when -- NSPCES=HCV), HxB2 (when --NSPCES=HIV-1).
--ORREF	Reference Result in Original Units	Char	Variable Qualifier of --ORRES	Reference value for the result or finding as originally received or collected. --ORREF uses the same units as --ORRES, if applicable. Examples: value from predicted normal value in spirometry tests.
--STREFC	Reference Result in Standard Format	Char	Variable Qualifier of --STRESC	Reference value for the result or finding copied or derived from --ORREF in a standard format.
--STREFN	Numeric Reference Result in Std Units	Num	Variable Qualifier of --STRESN	Reference value for continuous or numeric results or findings in standard format or in standard units. --STREFN uses the same units as --STRESN, if applicable.
--CHRON (Added for SEND Microscopic Findings domain)	Chronicity of Finding	Char	Variable Qualifier of --STRESC	Characterization of the duration of a biological process resulting in a particular finding. Multiple terms are not allowed for this variable. Examples: ACUTE, CHRONIC, SUBACUTE.
--DISTR Added for SEND Microscopic Findings domain)	Distribution Pattern of Finding	Char	Variable Qualifier of --STRESC	Description of the distribution pattern of a finding within the examined area. Examples: FOCAL, MULTIFOCAL, DIFFUSE, FOCAL MULTIFOCAL.
--REPNUM	Repetition Number	Num	Record Qualifier	The incidence number of a test that is repeated within a given timeframe for the same test. The level of granularity can vary, e.g., within a time point or within a visit. For example, multiple measurements of blood pressure or multiple analyses of a sample.
--CLMETH	Sample Collection Method	Char	Record Qualifier	Method of sample collection. *Additional wording TBD.*
--LOBXFL	Last Observation Before Exposure Flag	Char	Record Qualifier	Last observation prior to exposure. *Additional wording TBD.*

IDENTIFIERS FOR ALL CLASSES

APID	Associated Persons Identifier (Overlooked for inclusion in SDTMIG v1.4)	Char	Identifier for a single associated person, a group of associated persons, or a pool of associated persons. If APID identifies a pool, POOLDEF records must exist for each associated person. (See Section 5 for Associated Persons data).
FOCID	Focus of Study Specific Interest	Char	Identification of a focus of study-specific interest on or within a subject or specimen as called out in the protocol for which a measurement, test, or examination was performed, such as a drug application site, e.g., "Injection site 1", "Biopsy site 1", "Treated site 1", or a more specific focus, e.g., "OD" (right eye) or "Upper left quadrant of the back". The value in this variable should have inherent semantic meaning.
--RECID	Invariant Record Identifier	Char	Identifier for a record that is unique within a domain for a study and that remains invariant through subsequent versions of the dataset, even if the content of the record is modified. When a record is deleted, this value must not be reused to identify another record in either the current or future versions of the domain.

TIMING VARIABLES

MIDS	Disease Milestone Instance Name	Char	The name of a specific instance of a Disease Milestone Type (MIDSTYPE) described in the Trial Disease Milestones dataset. This should be unique within a subject. Used only in conjunction with RELMIDS and MIDSDTC.
RELMIDS	Temporal Relation to Milestone Instance	Char	The temporal relationship of the observation to the Disease Milestone Instance Name in MIDS. Examples: IMMEDIATELY BEFORE, AT TIME OF, AFTER.
MIDSDTC	Disease Milestone Instance Date/Time	Char	The start date/time of the Disease Milestone Instance Name in MIDS, in ISO8601 format.

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 (or are being) developed for SDTMIG v3.3. Eight new general-observation class domains have been added since SDTMIG v3.2. Nineteen new variables have been added; three of these were already existing, but just added to a second general observation class. 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. New concepts being considered for inclusion in the upcoming Batch 3 of SDTMIG updates include methods for managing multiple enrollments of the same USUBJID, clarification on use of Findings About, an approach for the standardization of lab units, and developing a standard method for representing multiple ATC codes for concomitant medications.

REFERENCES

1. Wood, F. (2008) The CDISC Study Data Tabulation Model (SDTM): History, Perspective, and Basics. PharmaSUG Proceedings, May 2008.
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.

CONTACT INFORMATION

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

Name: Fred Wood
Enterprise: Accenture Life Sciences, Accelerated R&D Services
Work Phone: 484-881-2297
E-mail: f.wood@accenture.com

Appendix 1
Disease Milestones 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
ABC	FA	ABC-1001	2	POSSCAUS	Possible cause identified	HYPOGLYCEMIA	Y	HYPO 1	PRIOR TO EVENT
ABC	FA	ABC-1001	3	MEALCAUS	Missed or delayed meal a possible cause	HYPOGLYCEMIA	Y	HYPO 1	PRIOR TO EVENT
ABC	FA	ABC-1001	4	PACAUS	Physical activity a possible cause	HYPOGLYCEMIA	N	HYPO 1	PRIOR TO EVENT
ABC	FA	ABC-1001	5	ALCCAUS	Alcohol a possible cause	HYPOGLYCEMIA	N	HYPO 1	PRIOR TO EVENT

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

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
ABC	CM	ABC-1001	1	HYPOGLYCEMIC TREATMENTS	HYPOGLYCEMIC TREATMENTS		Y	Y	HYPO 1	IMMEDIATLEY AFTER
ABC	CM	ABC-1001	4	GLUCOSE TABLETS	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	Y	HYPO 1	IMMEDIATLEY AFTER
ABC	CM	ABC-1001	5	GLUCAGON INJECTION	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	N	HYPO 1	IMMEDIATLEY AFTER
ABC	CM	ABC-1001	6	INTRAVENOUS GLUCOSE	HYPOGLYCEMIC TREATMENTS	MEDICATION	Y	N	HYPO 1	IMMEDIATLEY AFTER

** The shading shows hypoglycemic events with horizontal stippling and the timing relative to the events with vertical stippling.*