

Adopted Changes for SDTMIG v3.1.3 and 2013 OpenCDISC Upgrades

Yi Liu Celerion Inc., Lincoln, NE (USA)
Stephen Read Celerion Inc., Belfast, Northern Ireland

ABSTRACT

There have been several enhancements and upgrades to CDISC Study Data Tabulations Model (SDTM) data standards and associated implementation guides over recent years. August 2012 saw the release of SDTMIG v3.1.3 and adoption of SDTMIG v3.1.2 Amended 1 recommendations, in advance of the more recent and comprehensive SDTMIG v3.2 upgrade as released in December 2013.

In support of these SDTM enhancements the OpenCDISC community released 2 new SDTM validators in 2013. OpenCDISC Version 1.4 released in March 2013 and OpenCDISC 1.4.1 released in September 2013. Adoption of these enhanced SDTM data standards and associated validation applications has led to the need for SAS programmers or data mapping/submission specialists to consider implementing significant revisions across SDTM related SAS programs, processes and applications.

This paper will give a brief background and overview of some of the key SDTMIG and OpenCDISC updates and offer some potential guidance in support of their implementation with real case examples on potential SDTM programming enhancements for SAS programmers or mapping specialist working with the newer OpenCDISC validators, focusing primarily on some welcome improvements based on the following SDTM recommendations:

1. Interpretations on Required/Expected/Permissible variables including;
 - a) Inclusion of Study day variables such as STDY and ENDY across the majority of SDTM domains
 - b) The addition of EPOCH to all clinical subject-level observation domains
2. Overview of revisions to TS (Trial design) domain
3. Formatting and character variable length/size setting across all SDTM databases
4. Some recommendations on supporting documentation and associated SDTM reviewers guides

SDTM RELEASES BACKGROUND - OVERVIEW

SDTM first evolved from the Submission Data Model (SDM) as developed by CDISCs submission Data Standards (SDS) Team around 10 years ago. The SDS Team developed a set of consistent data standards for submitting tabulation data for human clinical trials. The SDS team is also responsible for maintaining the compatibility, consistency and conformity of SDTM data standards and associated SDTM Implementation Guide (SDTMIG) as part of the long running Clinical Data Interchange Standards Consortium (CDISC) strategy for harmonizing the submission of clinical trials data.

(Table 1) Since its introduction in 2004, the SDTM submission data standards model has grown to become one of the most commonly accepted industry standards for clinical trial submissions to the FDA. In June 2004, the FDA approved the SDTM version 1.0 (the model), and the SDTMIG v3.1 (the implementation guide) format for electronic data submission. With the release of SDTM version 1.1 (April 2005) and SDTMIG version 3.1.1 (Aug 2005), FDA published a proposal (December 2006) indicating the intention to mandate SDTM for electronic clinical data submissions. The next iteration, SDTM version 1.2 and the adjoining implementation guide v3.1.2 was released for public comment in July 2007. By September 2008 the FDA had selected CDISC to provide training to its reviewers of regulatory submissions. By the end of the year, CDISC made available the final version of SDTM version 1.2 and SDTMIG v3.1.2. November 2009, the FDA announced they were accepting SDTM v3.1.2 submissions. The SDTMIG 3.1.2 amendment 1 revision was released by CDISC preceding the subsequent released of SDTM 1.3 and SDTMIG v3.1.1 in July 2012. SDTMIG v3.2 was published in November, 2013, followed by the current SDTM v1.4 in December 2013. CDISC Technical Projects chart (Figure1) gives a general picture for the future planned SDTM and SDTMIG release.

¹ SDTMIG v3.1.3 is inclusive of the amended SDTM v1.2 and SDTMIG v3.1.2

Date	SDTM Released Version
June 2004	<ul style="list-style-type: none"> ➤ June 2004, FDA accepted CDISC SDTM as an allowable format for electronic data submissions SDTM v1.0 (the model), and the SDTMIG v3.1 (the implementation guide)
April 2005 August 2005	<ul style="list-style-type: none"> ➤ SDTM Version 1.1 was published ➤ STMIG v3.1.1 was published
December 2006	<ul style="list-style-type: none"> ➤ The FDA announced their intention to require SDTM as part of all electronic clinical data submissions
July 2007	<ul style="list-style-type: none"> ➤ SDTM v1.2 and SDTMIG v3.1.2 were posted for public comment
September 2008	<ul style="list-style-type: none"> ➤ CDISC announced that the FDA selected CDISC to provide training to FDA reviewers of regulatory submissions
November 2008	<ul style="list-style-type: none"> ➤ CDISC release SDTM 3.1.2 IG
November 2009	<ul style="list-style-type: none"> ➤ FDA announced that they would accept SDTM v3.1.2 submissions.
November 2011	<ul style="list-style-type: none"> ➤ CDISC released SDTM 3.1.2 IG Amendment 1 revisions
July 2012	<ul style="list-style-type: none"> ➤ CDSIC released SDTM 3.1.3 SDTM IG
November 2013 December 2013	<ul style="list-style-type: none"> ➤ SDTMIG v3.2 was published ➤ SDTM v1.4

Table 1. SDTM and SDTMIG Time Table for Releases and Updates

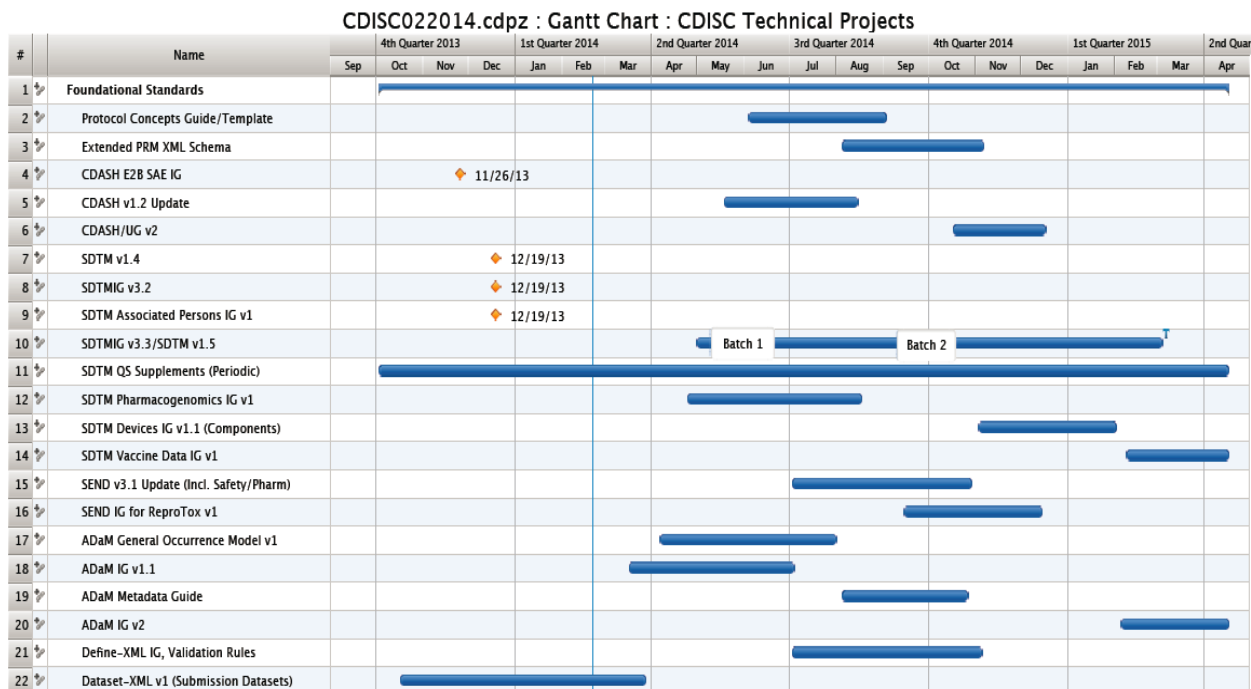


Figure 1. Current CDISC-Technical Projects²

OPENCDISC RELEASES BACKGROUND – OVERVIEW

OpenCDISC validator is a widely accepted open source software application that can be used to assist with the validation of clinical trials data and supporting data definition specifications against various CDISC data submission models (SDTM, ADaM, SEND, Define.xml). The progression of OpenCDISC Validator from v1.0 to v1.4.1 is shown in Table 2.

OpenCDISC Validator Version	Updated Information/Descriptions
February 2010 OpenCDISC Validator 1.0.	<ul style="list-style-type: none"> ➤ Support for SDTM 1.2 model and SDTMIG v3.1.2, 3.1.1 and Define.xml 1.0 standards ➤ Validated both syntactic and semantic constraints using Schematron ➤ Tuned for usability, accessibility, and Section 508
June 2010 OpenCDISC Validator 1.1	<ul style="list-style-type: none"> ➤ Support for SDTM 1.2 model and SDTMIG v3.1.2, 3.1.1 and Define.xml 1.0 standards ➤ Added a check for the presence of DM dataset in a submission
January 2011 OpenCDISC Validator 1.2	<ul style="list-style-type: none"> ➤ Support for SDTM 1.2 model and SDTMIG v3.1.2 ➤ Focused on adding support for ADaM validation ➤ Enhanced to further improve performance and scalability
March 2012 OpenCDISC Validator 1.3	<ul style="list-style-type: none"> ➤ Support for SDTM 1.2 model and SDTMIG v3.1.2 Amendment 1 ➤ Added 43 new SDTM validation rules ➤ Updated ADaM rules to align to the v1.1 release of validation checks from CDISC ADaM team ➤ Added a check for valid MedDRA version in Define.xml

² <http://www.cdisc.org/stuff/contentmgr/files/0/5fbd2b7afd4cc3ae3274d97d44d7a8b0/misc/cdisc022014.pdf>

<p>March 2013 OpenCDISC Validator 1.4</p>	<ul style="list-style-type: none"> ➤ Support for SDTM 1.3 model and SDTMIG v3.1.3 ➤ Compliance with CDER Common Data Standards Issues Document ➤ Improved linking to multiple CDISC Controlled Terminology versions ➤ 100+ new validation rules added for SDTM and ADaM
<p>September 2013 OpenCDISC Validator 1.4.1</p>	<ul style="list-style-type: none"> ➤ New ADaM v1.2 validation checks ➤ Validation of QS controlled terminology ➤ Several key bug fixes form OpenCDISC 1.4 ➤ 543 rules (most new checks are for QS domain and the QS-CT)

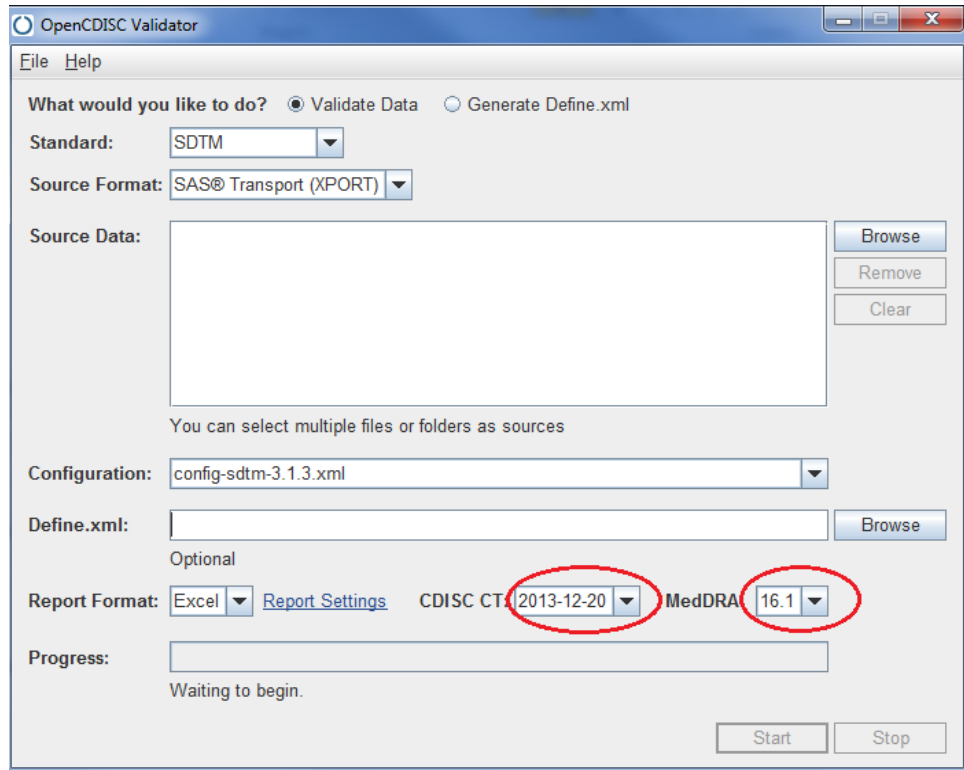
Table 2. OpenCDISC Release/Upgrades

The OpenCDISC Validator is a user friendly application with an intuitive interface that is straightforward to navigate and run across any number of SDTM data submission packages.

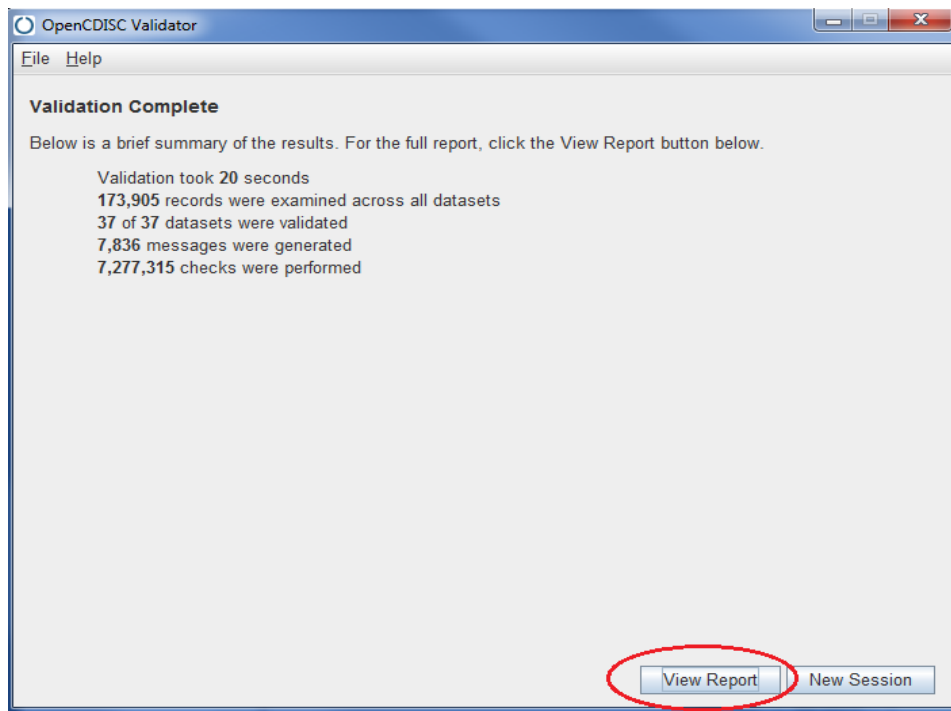
To run OpenCDISC (Display 1 and Display 2) users first choose a desired source database format (SAS Transport XPORT or Delimited Text), and browse to the applicable source database location (desktop folder, shared network drive etc.). Second, select the configuration file as config-sdtm-3.1.3.xml or the SDTM version you want to validate your databases against. Third, choose an appropriate output report format (Excel or CSV), for any given version of CDISC Control Terminology (CT) and MedDRA (as configured³). Users can also choose to validate database with or without a corresponding define.xml file or indeed choose to validate a define.xml individually. Once these options are selected, click “Start” to begin validating your chosen databases. Depending on the size of the databases execution can be very prompt and efficient. After execution users will be presented with a compliance summary/validation report. The output reports details errors and warning logs for any validation findings associated with the selected submission package. The validation reports are very detailed and contain nicely formatted error/warning/notice issue summary tables with specific examples of individual compliance findings with information on where to modify datasets or define xml.

Over the last 4 years OpenCDISC has become an essential tool for any programmer or reviewer working with SDTM submissions. As data, mapping specialists or SDTM reviewers across the industry become more familiar with the validation requirements SDTM review has become a much more streamlined and efficient process.

³ <http://www.opencdisc.org/projects/validator/configuring-opencdisc-validator-meddra>



Display 1. OpenCDISC Validator Main Page



Display 2. OpenCDISC Validator Results Screen

BACKGROUND ON REQUIRED/EXPECTED/PERMISSIBLE VARIABLES

The SDTM implementation guide categorizes SDTM variables as required, expected, or permissible. A required variable must be included in the dataset and cannot be blank. For example: STUDYID, DOMAIN, USUBJID, and -SEQ are all required in each SDTM domain. An expected variable must be included in the dataset, but it can be missing or might have null values. A permissible variable should be used in a domain as appropriate when collected or derived and these variables are not required as part of the structure domain. Empty permissible variables may be dropped in the datasets if needed.

2011 Common Data Standards Issues Document (CDER) and most recent FDA submission guidance⁴ suggests that all permissible variables for which data were collected or for which derivations are possible should be submitted for traceability. The current 3.1.3/3.2 SDTM IGs and 1.4.1 OpenCDISC accommodate these latest FDA recommendations. Prior to 2013's SDTM and OpenCDISC releases variables such as --DY, --STDY, --ENDY and EPOCH may not have been regarded as a 'requirement' across all SDTM database for all sponsors.

-DY -STDY & -ENDY

Study day variables such as -STDY and -ENDY are detailed as 'permissible' in the SDTM IG yet they can and do apply to the majority of SDTM domains. The 2013 OpenCDISC releases include additional checks in relation to their presence in most SDTM domains. By presenting study day variable (--DY, --STDY, and --ENDY), reviewers can see and understand the relationship between records and the study's start or end date and timing of event. This can eliminate the need to rely on VISIT information to determine timing. Therefore, these variables should really be included in the any SDTM domain when the relative --DTC or --STDTC/--ENDTC variables are also populated. Study day variables describe relationship between the study day (start/end day) relative to the subject's reference start date/time (RFSTDTC); and can be calculated with --STDTC and DM.RFSTDTC which have a defined year, month, and day. RFSTDTC is pulled from the demographic domain and normally refers to the date of first exposure to study treatment, ex. Study Day 1. If the collection date is greater than or equal to the date of first exposure, then Study Day (--DY) equal to collection date --DTC minus RFSTDTC+1 since SDTM does not allow for study day 0; if the collection date is less to the date of first exposure, then Study Day (--DY) equal to collection date --DTC minors RFSTDTC.

When Start Study Date/Time (--STDTC) variable is contained in the dataset, the Study Day of Start (--STDY) variable should also be included, the same as requirement exists for the Study Day of End (--ENDY) variables. If --DTC is present, --DY should also be populated in the dataset.

Older version of OpenCDISC didn't flag warnings when these variables were missing. However, OpenCDISC 1.4 or 1.4.1 gives the warnings "Missing -STDY/-ENDY variable, when XX-STDTC/-ENDTC variable is present (Display 3 and Table 3).

Domain	Variables	Values	Rule ID	Message	Category	Severity
AE	VARIABLE	AESTDY	SD1087	Missing AESTDY variable, when AESTDTC variable is present	Presence	Warning
AE	VARIABLE	AEENDY	SD1091	Missing AEENDY variable, when AEENDTC variable is present	Presence	Warning
MH	VARIABLE	MHSTDY	SD1087	Missing MHSTDY variable, when MHSTDTC variable is present	Presence	Warning
MH	VARIABLE	MHENDY	SD1091	Missing MHENDY variable, when MHENDTC variable is present	Presence	Warning
SE	VARIABLE	MHSTDY	SD1087	Missing SESTDY variable, when SESTDTC variable is present	Presence	Warning
SE	VARIABLE	MHENDY	SD1091	Missing SEENDY variable, when SEENDTC variable is present	Presence	Warning

Display 3. Associated OpenCDISC 1.4.1 Warnings

⁴ CDER Common Data standards Issues document Version 1.1/December 2011. Available at <http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM254113.pdf>

Rule ID	Message	Description	Category	Severity
SD1087	Missing --STDY variable, when --STDTC variable is present	Study Day of Start (--STDY) variable should be included into dataset, when Start Study Date/Time (--STDTC) variable is present	Presence	Warning
SD1091	Missing --ENDY variable, when -ENDTC variable is present	Study Day of End (--ENDY) variable should be included into dataset, when End Study Date/Time (--DTC) variable is present	Presence	Warning

Table 3. New Rule for OpenCDISC v1.4 Timing Variables

The consistent use of a basic core SAS macro such as the example below across SDTM build or data mapping can facilitate the derivation of DY across most applicable SDTM databases and thus avoid any associated OpenCDISC warning where these variables may have previously been omitted from a submission. The example code below shows a simple 'day' core macro call for AEs, where ae_stdtd and ae_endtd are source AE start and end date variable names and corresponding fmedddat_1 (overall first medication date) has been appropriately merged in against all source AEs

```

%macro day(var, date1, date2);
    if &date1 ge &date2 then &var = ((&date1 - &date2) + 1);
    else &var = &date1 - &date2;
%mend;

%day(aestdy, ae_stdtd, fmeddt_1); ** first med date equivalent to RFSTDTC;
%day(aeendy, ae_endtd, fmeddt_1);
    
```

EPOCH

“Please include the variables EPOCH for every clinical subject-level observation (e.g., adverse events, laboratory, concomitant, medications, exposure, and vital signs)”³. CDER Common Data standards Issues document have recommended the inclusion of an EPOCH variable since 2011 version 1.1 guidance. CDER also recommended the inclusion of more granular trial element information across domains although this has yet to be formally adopted due to the difficulties in implementation. Associating specific actual treatments across all individual events for SDTM may even begin to blur the lines between Analysis Data (ADaM) and SDTM data expectations.

SDTM guidance describes Epoch: as “an interval of time in the planned conduct of a study during which treatment is constant.” The period of subjects’ participation in the trial is divided into epochs. The worked example through section 7.2 of the SDTM submission guidance offer more detail on potential definitions of EPOCH across various common clinical trial designs. Normally epoch starts with Screening Epoch, Treatment Epoch and then follow up Epoch (Figure 2). Additional terms like run-in and rest can be added as needed, these values must also exist in Trial Arm (TA) domain.



Figure 2. Screen Epoch-Treatment Epoch-Follow up Epoch

Domain	Variables	Values	Rule ID	Message	Category	Severity
AE	VARIABLE,	EPOCH, AE	SD1077	FDA Expected variable not found	Metadata	Warning
CM	VARIABLE, DATASET	EPOCH, CM	SD1077	FDA Expected variable not found	Metadata	Warning
DS	VARIABLE, DATASET	EPOCH, DS	SD1077	FDA Expected variable not found	Metadata	Warning
EG	VARIABLE, DATASET	EPOCH, EG	SD1077	FDA Expected variable not found	Metadata	Warning
EX	VARIABLE, DATASET	EPOCH, EX	SD1077	FDA Expected variable not found	Metadata	Warning
LB	VARIABLE, DATASET	EPOCH, LB	SD1077	FDA Expected variable not found	Metadata	Warning
MH	VARIABLE, DATASET	EPOCH, MH	SD1077	FDA Expected variable not found	Metadata	Warning
PE	VARIABLE, DATASET	EPOCH, PE	SD1077	FDA Expected variable not found	Metadata	Warning
SU	VARIABLE, DATASET	EPOCH, SU	SD1077	FDA Expected variable not found	Metadata	Warning
VS	VARIABLE, DATASET	EPOCH, VS	SD1077	FDA Expected Message variable not found	Metadata	Warning

Display 4. OpenCDISC Validator Results for Missing EPOCH Variables

Rule ID	Message	Description	Category	Severity
SD1077	FDA Expected variable not found	Variables requested by FDA in CDER Common Data Issues document should be included in the dataset	Metadata	Warning

Table 4. New Rule for the OpenCDISC Validator v1.4/v1.4.1 EPOCH Variables

Although the CDER 2011 guidance on EPOCH was fairly clear, prior to 2013 many sponsors may not have regarded it's inclusion as a submission requirement as this was not clearly detailed in the SDTM 3.1.2 (and amendment 1) IG. The 2013 OpenCDISC releases moved to apply additional checks for the inclusion and compliance of EPOCH across most SDTM domains. OpenCDISC v1.4.1 rule SD1077 (Table 4) followed the recommendation. Warning message will be populated (Display 4) if epoch variables are missing in certain domains. Table 5 provides examples of domains with required epoch variables that will be flagged if missing only in the current iteration of the OpenCDISC 1.4/1.4.1 validator.

CLASSES	DOMAIN						
Events	AE	CE	DS	DV	MH		
Findings	EG	LB	PC	PE	PP	QS	VS
Interventions	CM	EX	SU				

Table 5. Example of Require Epoch Variable Domains

One approach for defining EPOCH across a batch of SDTM databases may be to associate EPOCH with corresponding study visit macros or metadata; essentially associating an Epoch with the original source/CDASH/CRF visit or timing variables. Table 6 shows an example Visit/Epoch meta-data file that could be used to assist in the population. However special care should be taken in the definition of EPOCH across different types of SDTM domains for any given study design.

For example if on a basic single dose study design where a Subject doses on SDTM Visit Day 1, it would obviously make sense to associate most Day 1 events with a 'Treatment' EPOCH, but pre-dose events on Day 1 would still be considered as RUN-IN. Therefore we need to allow for the same visit definition on SDTM to be associated with multiple EPOCHs. In the example below if participants remain confined in a clinic on Days 2-4, these subsequent visits could be considered for TREATMENT or WASHOUT EPOCH. For adverse events it is important to associate EPOCH with the start date/time of the AE and not end date etc. The Define.xml can be used to detail any domain specific computational methods or algorithms associated with EPOCH definitions in each instance. DY Macro shown as example below:


```

%macro epoch(day);
  if &day. lt -1 then epoch='SCREENING';
  else if &day. in (-1) and ae_sttm lt fmedtm_1 then epoch='RUN-IN';
  else if &day. = 1 and ae_sttm gt fmedtm_1 then epoch='TREATMENT';
%mend;

```

CDASH/CRF		SDTM VISITS			SDTM EPOCH/ELEMENT (all Domains)		
Period	Day	VISITNUM	VISIT	VISITDY	EPOCH	ETCD	ELEMENT
Screen	.	1	SCREENING	.	SCREENING	SCRN	SCREENING
1	-1	2	DAY -1	-1	RUN-IN	BASE	BASELINE
1	1	3	DAY 1	1	TREATMENT	A	Treatment ARM name
1	2	4	DAY 2	2	WASHOUT	WASH	WASHOUT
1	3	5	DAY 3	3	WASHOUT	WASH	WASHOUT
1	4	6	DAY 4	4	WASHOUT	WASH	WASHOUT
Post	.	99	FINAL VISIT	.		FU	FOLLOW-UP

Table 6. Visit/Epoch Meta-data

TRIAL SUMMARY DOMAIN

The Trial Summary (TS) dataset in SDTM provides basic high-level information on information about the general study protocol, and study conduct such as such as study title, study design, study interventions, assessments, trial objectives etc.

The 2013 OpenCDISC releases were the first to check for several of the newer TS classification parameters that were expected for SDTM IG v3.1.2 and now required or conditionally required in current SDTMIG 3.1.3 and 3.2. OpenCDISC v1.4.1 now reflects these regulations and returns errors message for several additional missing parameters. Display 5 shows an example of the type of OpenCDISC error message where two 'newer' TS parameters have been omitted from TS.

REGID is the Trial register identifier and can be located through searches of the ClinicalTrials.gov⁵ database to get the identifier. OUTMSPRI is primary outcome measure as detailed in protocol.

TS		
SD2221	Missing REGID Trial Summary Parameter	Error
SD2222	Missing OUTMSPRI Trial Summary Parameter	Error

Display 5. Example on TS Error Message for Missing Required Parameters.

TSVALCD, TSVDFREF and TSVCDVER are newer variables required since SDTMIG v3.1.3. TSVALCD is the code of the term in TSVAL. TSVDFREF is null favor of the value of TSPARM, it should be populated if TSVAL is null. TSVCDVER is the name or version number of the Reference Terminology from which TSVALCD is taken. The description on Null flavor logic in the TS domain in the latest SDTM IG is worth a read for anyone involved with set up or reviews of this domain as these new variables do allow for the avoidance of OpenCDISC errors in TS in the event of not being able to populate an appropriate value for one of the required TS parameters.

A complete list of newer TS parameters that sponsors may not have regarded as a 'requirement prior to 2013 OpenCDISC upgrades are included in Display 6 highlighted in bold/red. There are five additional parameters conditionally required if study type is 'INTERVENTIONAL'; Trial Indication Type (TINDTP), Investigational Therapy or Treatment (TRT), Pharmacological Class of Investigational Therapy (PCLAS), Interventional Model (INTMODEL) and Interventional Type (INTTYPE).

⁵ <http://clinicaltrials.gov/ct/gui>

DOMAIN	TSPARMCD	TSPARM	TSVAL	TSVALI	TSVALCD	TSVCDREF	TSVCDVER
TS	ACTSUB	Actual Number of Subjects	8				
TS	ADAPT	Adaptive Design	Y		C49488	CDISC	12/20/2013
TS	ADDON	Added on to Existing Treatments	N		C49487	CDISC	12/20/2013
TS	AGEMAX	Planned Maximum Age of Subjects	P55Y			ISO 8601	
TS	AGEMIN	Planned Minimum Age of Subjects	P18Y			ISO 8601	
TS	CURTRT	Current Therapy or Treatment	Main Drug/compound name				
TS	DCUTDESC	Data Cutoff Description	DATABASE LOCK				
TS	DCUTDC	Data Cutoff Date	DBLK date			ISO 8601	
TS	DOSE	Dose per Administration	e.g. 800				
TS	DOSFRQ	Dosing Frequency	e.g. QNCE				
TS	DOSU	Dose Units	e.g. mg				
TS	FCNTRY	Planned Country of Investigational Sites	USA		USA	ISO 3166	
TS	HLTSUBJI	Healthy Subject Indicator	N		C49487	CDISC	12/20/2013
TS	INTMODEL	Intervention Model	CROSS-OVER		C82639	CDISC	12/20/2013
TS	INTTYPE	Intervention Type	DRUG		C1909	CDISC	12/20/2013
TS	LENGTH	Trial Length	P9D			ISO 8601	
TS	NARMS	Planned Number of Arms	1				
TS	OBJPRIM	Trial Primary Objective	From protocol				
TS	OBJSEC	Trial Secondary Objective	From protocol				
TS	OUTMSPRI	Primary Outcome Measure	Pharmacokinetics				
TS	PCLAS	Pharmacological Class of Invest. Therapy	NONE		NONE		12/20/2013
TS	PLANSUB	Planned Number of Subjects	8				
TS	RANDOM	Trial is Randomized	U		C17998	CDISC	12/20/2013
TS	REGID	Registry Identifier	NONE		NONE	CT.GOV	
TS	ROUTE	Route of Administration	ORAL				
TS	SENDTC	Study End Date	7/30/2013			ISO 8601	
TS	SEXPOP	Sex of Participants	F		C16576	CDISC	12/20/2013
TS	SPONSOR	Clinical Study Sponsor	XXXXXXXX				
TS	SSTDTC	Study Start Date	7/22/2013			ISO 8601	
TS	STOPRULE	Study Stop Rules	INTERIM ANALYSIS FOR FUTILITY				
TS	STYPE	Study Type	INTERVENTIONAL		C98388	CDISC	12/20/2013
TS	TBLIND	Trial Blinding Schema	OPEN LABEL		C49659	CDISC	12/20/2013
TS	TCNTRL	Control Type	PLACEBO		C49648	CDISC	12/20/2013
TS	TINDTP	Trial Indication Type	TREATMENT		C49656	CDISC	12/20/2013
TS	TITLE	Trial Title	From protocol				
TS	TPHASE	Trial Phase Classification	Phase I/III Trial		C15694	CDISC	12/20/2013
TS	TRT	Investigational Therapy or Treatment	Treatment name				
TS	TTYPE	Trial Type	PHARMACOKINETIC		C49663	CDISC	12/20/2013

Display 6. Example TS Parameters for New OpenCDISC Required

Given the increased amount of content and scope of the TS domain SAS programmers and SDTM specialists should not be afraid to reach out to colleagues in clinical study teams in order to review and confirm all TS parameter definitions are appropriate. For example PCLAS (Pharmacological class of interest) may not be immediately clear to a SDTM programmer if it is not explicitly detailed in a SAP or protocol so programmers should not be afraid to reach out to their appropriate peers in order to clarify potential definitions rather than defaulting to the Null flavor type responses.

To assist in the build, population and review of the TS database, programming teams are likely to find it easier to develop and maintain configurable data files to help better control and define the expanded TS requirements. A configurable excel based TS build spreadsheet that automatically links to drop down lists containing controlled terminology with auto-populated corresponding code list numbers is well worth consideration. Any programming to convert such configurable TS data files to a corresponding SAS xpt file is straight forward. Display 7 shows an example of a CT drop down list for TSVAL parameters.

TSPARMCD	TSPARM	PROTSPEC	TSVAL	TSVALCD	TSVCDREF
ACTSUB	Actual Number of Subjects		40		
ADAPT	Adaptive Design		N	C49487	CDISC
ADDON	Added on to Existing Treatments		Y	C49488	CDISC
AGEMAX	Planned Maximum Age of Subjects		P55Y		ISO 8601
AGEMIN	Planned Minimum Age of Subjects		P18Y		ISO 8601
AGESPAN	Age Span		P37Y		ISO 8601
CURTRT	Current Therapy or Treatment				
DCUTDESC	Data Cutoff Description				
DCUTDTC	Data Cutoff Date				ISO 8601
DOSE	Dose per Administration				
DOSFRQ	Dosing Frequency		BID	4496	CDISC
DOSU	Dose Units		BID	3253	CDISC
FCNTRY	Planned Country of Investigational Sites		2 TIMES PER WEEK	A	ISO 3166
HLTSUBJI	Healthy Subject Indicator		Q10H	9488	CDISC
LENGTH	Trial Length		Q11H		ISO 8601
NARMS	Planned Number of Arms		Q12H		
			Q13H		
OBJPRIM	Trial Primary Objective	This trial is an o	This trial is an open-la		
OBJSEC	Trial Secondary Objective	This trial is an o	This trial is an open-la		
OUTMSPRI	Primary Outcome Measure				
PLANSUB	Planned Number of Subjects		40		
RANDOM	Trial is Randomized		Y	C49488	CDISC
REGID	Registry Identifier				
ROUTE	Route of Administration		ORAL	C38288	CDISC
SENDTC	Study End Date				ISO 8601
SEXPOP	Sex of Participants		BOTH	C49636	CDISC

Display 7. Example of a Controlled Drop down List in a Configurable TS Excel File

COLUMN LENGTH/SIZE

Large SAS transport file size can cause significant issues for the FDA. Many submission datasets may have previously contained a significant amount of wasted space where sponsors defaulted to a character variable length of 200 across SDTM datasets. Due to archiving storage, hardware limitations etc. this can result in an inefficient data review and storage process for FDA staff analyzing sizable datasets.

One simple approach to help address file size limitations is to re-size the maximum length of any given character variable to correspond to the length of the longest term or string contained within the variable. CDER also states: “For both CDISC and non-CDISC datasets, in order to significantly reduce dataset file sizes, the allotted character variable length/size for each column in a dataset should be the maximum length used.”⁶

OPENCDISC CHECK CURRENT UPDATES FOR LENGTH/SIZE

OpenCDISC version 1.4, added checks to ensure SDTM domains adhere to the FDA guidelines on variable size (Display 8 and Table 7). An error now occurs when extra characters are detected; similar error messages are included under SD1080 and SD1081. All “Non-recommended variable length” checks also remain in OpenCDISC v1.4.1. As a result the 2013 versions of OpenCDISC flagged several errors across any SDTM databases that did not adhere to this variable length constraint.

⁶ CDER Common Data standards Issues document Version 1.1/December 2011. Available at <http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM254113.pdf>



LB				
SD1082	LBCAT	variable length is too long for actual data	Error	1
SD1082	LBELTM	variable length is too long for actual data	Error	1
SD1082	LBNAM	variable length is too long for actual data	Error	1
SD1082	LBNRIND	variable length is too long for actual data	Error	1
SD1082	LBORNRLI	variable length is too long for actual data	Error	1
SD1082	LBORNRL0	variable length is too long for actual data	Error	1
SD1082	LBORRES	variable length is too long for actual data	Error	1
SD1082	LBREFID	variable length is too long for actual data	Error	1
SD1082	LBSPEC	variable length is too long for actual data	Error	1
SD1082	LBSTNRC	variable length is too long for actual data	Error	1
SD1082	LBSTRESC	variable length is too long for actual data	Error	1
SD1082	LBTPPTREF	variable length is too long for actual data	Error	1
SD1082	USUBJID	variable length is too long for actual data	Error	1
SD1082	VISIT	variable length is too long for actual data	Error	1

Display 8. OpenCDISC Report Results indicating the excess of Variables Character Length

Rule ID	Message	Description	Category	Severity
SD1080	-- variable length is too long for actual data	Variable length should be assigned based on actual stored data to avoid to minimize file size. Datasets should be resized to the maximum length used prior to splitting.	Metadata	Information
SD1081	-- variable length is too long for actual data	Variable length should be assigned based on actual stored data to avoid to minimize file size. Datasets should be resized to the maximum length used prior to splitting.	Metadata	Warning
SD1082	-- variable length is too long for actual data	Variable length should be assigned based on actual stored data to avoid to minimize file size. Datasets should be resized to the maximum length used prior to splitting.	Metadata	Error

Table 7. New OpenCDISC Rules on Variable Character Length

Some results of re-sizing the dataset variables show a significant decline in the size of the resultant data files. Thereby removing unnecessary clutter and optimizing the allocation of space. Display 9-11 below shows the potential for file size reduction after character variable minimization across a typical lab dataset. This lab dataset size is 10010 KB before adjustment, while the file size drop significantly to 2877 KB after variable length reduction.

Name	Date modified	Type	Size
 Lab Dataset after Limiting Size to the True Size.xpt	3/18/2014 1:45 PM	SAS Xport Transport File	2,877 KB
 Lab Dataset before Limiting Size to the True Size.xpt	3/18/2014 12:18 PM	SAS Xport Transport File	10,010 KB

Display 9. Lab Datasets File Size before and after limiting Size

#	Variable	Type	Length	Format	Informat	Label
1	STUDYID	Character	20			Study Identifier
2	DOMAIN	Character	2			Domain Abbreviation
3	USUBJID	Character	40			Unique Subject Identifier
4	LBSEQ	Numeric	8			Sequence Number
5	LBREFID	Character	40			Specimen ID
6	LBTESTCD	Character	8			Lab Test or Examination Short Name
7	LBTEST	Character	40			Lab Test or Examination Name
8	LBCAT	Character	40			Category for Lab Test
9	LBORRES	Character	200			Result or Finding in Original Units
10	LBORRESU	Character	20			Original Units
11	LBORNRL0	Character	40			Reference Range Lower Limit in Orig Unit
12	LBORNRLHI	Character	40			Reference Range Upper Limit in Orig Unit
13	LBSTRESC	Character	200			Character Result/Finding in Std Format
14	LBSTRESN	Numeric	8			Numeric Result/Finding in Standard Units
15	LBSTRESU	Character	20			Standard Units
16	LBSTNRLO	Numeric	8			Reference Range Lower Limit-Std Units
17	LBSTNRHI	Numeric	8			Reference Range Upper Limit-Std Units
18	LBSTNRC	Character	40			Reference Range for Char Rst-Std Units
19	LBNRIND	Character	40			Reference Range Indicator
20	LBNAM	Character	60			Vendor Name
21	LBSPEC	Character	40			Specimen Type
22	LBBFL	Character	1			Baseline Flag
23	VISITNUM	Numeric	8			Visit Number
24	VISIT	Character	40			Visit Name
25	VISITDY	Numeric	8			Planned Study Day of Visit
26	EPOCH	Character	40			Epoch
27	LBDTC	Character	19			Date/Time of Specimen Collection
28	LBDY	Numeric	8			Study Day of Specimen Collection
29	LBTPPT	Character	40			Planned Time Point Name
30	LBTPPTNUM	Numeric	8			Planned Time Point Number
31	LBELTM	Character	40			Planned Elapsed Time from Time Point Ref
32	LBTPPTREF	Character	40			Time Point Reference

Display 10. Variable Length within a lab dataset prior to Limiting the Variable Size

#	Variable	Type	Length	Format	Informat	Label
1	STUDYID	Character	7			Study Identifier
2	DOMAIN	Character	2			Domain Abbreviation
3	USUBJID	Character	17			Unique Subject Identifier
4	LBSEQ	Numeric	8			Sequence Number
5	LBREFID	Character	9			Specimen ID
6	LBTESTCD	Character	7			Lab Test or Examination Short Name
7	LBTEST	Character	33			Lab Test or Examination Name
8	LBCAT	Character	17			Category for Lab Test
9	LBORRES	Character	12			Result or Finding in Original Units
10	LBORRESU	Character	7			Original Units
11	LBORNRL0	Character	5			Reference Range Lower Limit in Orig Unit
12	LBORNRLHI	Character	5			Reference Range Upper Limit in Orig Unit
13	LBSTRESC	Character	12			Character Result/Finding in Std Format
14	LBSTRESN	Numeric	8			Numeric Result/Finding in Standard Units
15	LBSTRESU	Character	7			Standard Units
16	LBSTNRLO	Numeric	8			Reference Range Lower Limit-Std Units
17	LBSTNRHI	Numeric	8			Reference Range Upper Limit-Std Units
18	LBSTNRC	Character	14			Reference Range for Char Rst-Std Units
19	LBNRIND	Character	8			Reference Range Indicator
20	LBNAM	Character	8			Vendor Name
21	LBSPEC	Character	5			Specimen Type
22	LBBFL	Character	1			Baseline Flag
23	VISITNUM	Numeric	8			Visit Number
24	VISIT	Character	11			Visit Name
25	VISITDY	Numeric	8			Planned Study Day of Visit
26	EPOCH	Character	22			Epoch
27	LBDTC	Character	19			Date/Time of Specimen Collection
28	LBDY	Numeric	8			Study Day of Specimen Collection
29	LBTPPT	Character	21			Planned Time Point Name
30	LBTPPTNUM	Numeric	8			Planned Time Point Number
31	LBELTM	Character	7			Planned Elapsed Time from Time Point Ref
32	LBTPPTREF	Character	17			Time Point Reference

Display 11. Variables Length within a Lab Dataset after Limiting the variable Size

METHODS FOR MINIMIZING LENGTH- INTRODUCTION TO MINCVARS MACRO

While there are several useful published SAS macros that can be leveraged to automatically set the length of selected character variable to match the length of the longest term within the variable. Celerion data programming team decided to develop our own array based character variable minimisation 'MINCVARS' core macro (as provided below).

Adoption of this MINCVARS macro across SDTM submissions has led to a significant file size reduction across SDTM datasets,

```

%macro mincvvars(dsn,rai=) ;
  /* Macro mincvvars will reduce the length of all character
     variables in a SAS dataset to the longest actual length.
     Parameter DSN is the dataset name,
     RAI is Retain As Is, that is for the following list of
     variable names, separated by spaces, do not change their
lengths,
     due perhaps to OPENCDISC standards.
  ;
  %put dsn=&dsn., rai=&rai. ;
  %let rai = %upcase(&rai.) ;
  %put rai=&rai. ;
  * The following data step will return local macro variable NUMCVARS,
     which will contain the number of character variables
     in the dataset.
  ;
  DATA _NULL_ ;
    IF NOB = 0 THEN DO ;
      * With zero observations, there are zero character
variables.;
      CALL SYMPUTX('numcvvars',0,'L') ;
      STOP ;
    END ;
    ELSE DO ;
      SET &dsn. NOBS=NOB ;
      * An array composed only of character variables.;
      ARRAY CVARS _CHARACTER_ ;
      NUMCVARS = DIM(CVARS) ;
      CALL SYMPUTX('numcvvars',NUMCVARS,'L') ;
      STOP ;
    END ;
  RUN ;
  %put numcvvars = &numcvvars ;

  * Create macro variable containing the minimum lengths of
     character variables.
  ;
  DATA _NULL_ ;
    SET &dsn. END=END1 ;
    ARRAY CVARS _CHARACTER_ ;
    LENGTH CLENS $ 32000 TEMPVCVAR $ 32 ;

  * Create maxlength variables for each character variable. ;
  %do i = 1 %to &numcvvars ;
    RETAIN mx&i. 0 ;
    mx&i. = MAX(mx&i.,LENGTH(CVARS[&i])) ;
  %end ;

  IF NOT END1 THEN DELETE ;

  * Create CLENGTHS macro variable, which contains minimum character
lengths.;

```

```

%do i = 1 %to &numcvars ;
    CALL VNAME(CVARS[&i.], TEMPCVAR) ;
    TEMPCVAR = UPCASE(TEMPCVAR) ;

* Only include variables not in the Retain As Is parameter. ;
    IF INDEXW("&rai.",TEMPCVAR) = 0 THEN
        CLENS = TRIMN(CLENS) || " " || STRIP(TEMPCVAR) || ' $ '
            || COMPRESS(PUT(MX&i.,BEST.)) ;
    ELSE PUT "Length for variable " TEMPCVAR " was not
changed." ;

%end ;
CALL SYMPUTX('clengths',CLENS,'L') ;
RUN ;
%put clengths = &clengths.;

* Minimize the lengths of the character variables. ;
DATA &dsn. (REPLACE=YES REPEMPTY=YES) ;
    LENGTH &clengths ;
    SET &dsn. ;
*OUTPUT ;
RUN ;

%mend mincvars ;

```

Example of a call to the Mincvars macro in production:

```

*****WRITE FINAL DATASET TO TRANSFER AREA*****;
%mincvars(vit_sdtm,rai=vstestcd);

```

REVIEWERS GUIDE

BACKGROUND

A critical component of any SDTM submissions includes how we document the submission package via a corresponding Study Data Reviewer's Guide (SDRG). Extensive background guidance and recommendation on this supporting specification is currently under development and available for public review from the Association Programming Pharmaceutical Users Software Exchange (PHUSE) Study Data Reviewer's Guide (SDRG) core team⁷. The SDRG gives SDTM specialists the opportunity to formally document and expand upon SDTM data as submitted for any given protocol. A SDRG can help avoid any potential for version control headaches associated with retrospective validation findings being flagged if of older SDTM submission packages are ran against newer OpenCDISC releases that were not available at the time of SDTM build.

Documenting the specific version of SDTM and define.xml that has been submitted, the version of OpenCDISC that was used to validate the submission and detailing justifications for any unavoidable 'false positive' OpenCDISC findings via a conformance summary has now become an essential requirement for SDTM programmers to consider.

Some high-level examples of these key sections of a SDRG are detailed below:

Wonder Drug ABC 123456

Guide to SDTM Tabulations Data

Overview

This SDTM submission package has been prepared in line with CDISC defined guidelines for submission of Tabulations Data, following CDISC Study Data Tabulation Model (SDTM) v1.3 and Study Data Tabulation Model Implementation Guide (SDTMIG) v3.1.3. The SDTM package includes a supporting define.xml (v1.0) and annotated 'blankcrf.pdf'.

⁷ http://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer%27s_Guide#SDRG_Core_Team

Conformance Summary

OpenCDISC version 1.4.1 was used to assist with the preparation and validation of the SDTM databases and supporting define.xml. There are no OpenCDISC errors associated with the SDTM datasets or supporting define.xml. Responses to any non-issue false positive warning that were flagged as a result of OpenCDISC validation are outlined below:

Domain	Finding	Response
MH	Value for MHBODSYS not found in (SOC) CT <u>codelist</u>	A known Open CDISC bug that arose when the CDISC System Organ Class Code list (C66783) was removed from the Controlled Terminology File. Open CDISC was not updated to reflect this removal
MH	Variable appears in dataset, but is not in SDTM standard	Coding variables SHOULD be reported in the core MH domain. Open CDISC bug
LB	Missing value for LBORRESU, when LBORRES is provided	Okay as is – Lab tests such as Specific Gravity and pH that do not have units
TV	Invalid VISIT/VISITNUM	The Unscheduled Visits are not included in the TV domain per the Sponsor requested SDTM standards provided for the Trial Design domains.

Display 12. Reviewers Guide Sample**CONCLUSION**

This paper only covers some of the key recent changes and additions that SAS programmers may have encountered during 2013 when working with the enhanced OpenCDISC releases associated with latest SDTMIG.

SDTM data standards will continue to expand so ongoing review, and maintenance of SDTM programming standards are likely to keep programmers and data mapping specialist interested well into the futures as the industry works to continuously improve and streamline the data reporting process.

REFERENCE

CDSIC SDTM Implementation Guide 3.1.2 amendment
Available at <http://www.cdisc.org/sdtm>

CDSIC SDTM Implementation Guide 3.1.3
Available at <http://www.cdisc.org/sdtm>

CDISC Website. <http://www.cdisc.org/>

OpenCDISC Website. <http://www.opencdisc.org/>

CDER Common Data standards Issues document Version 1.1, December 2011. Available at <http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM254113.pdf>

Pharmaceutical Users Software Exchange (PHUSE), Study Data Reviewer's Guide (SDRG). Available at http://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer%27s_Guide#SDRG_Core_Team

ACKNOWLEDGE

The authors would like to acknowledge Celerion Inc. for providing the opportunity to work on this paper and some of our colleagues in Celerion's Data Programming group for their valued input and assistance across last year's SDTM and OpenCDISC upgrades. Including Larry Gregg (Celerion, Senior Data Programmer II) as the primary author of the mincvars macro, Jenny Erskine (Celerion, Senior Data Programmer I) for her development work on Trial Summary meta data driven speciation's, and Miao Zheng (Celerion, Data Programmer I) for his assistance with our initial testing and development work for SDTM IG 3.1.3/3.2 programming revisions.

We would also like to take this opportunity to acknowledge the many contributors to CDISC data standard initiatives including the SDTM SDS group, the OpenCDISC community and the PHUSE SDRG working group.

CONTACT INFORMATION

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

Name: Yi Liu
Enterprise: Celerion Inc.
Address: 621 Rose Street
City, State ZIP: Lincoln, NE 68502
Work Phone: 402-437-4778
E-mail: yi.liu@celerion.com

Name: Stephen Read
Enterprise: Celerion Inc.
Address: 22-24 Lisburn Road
City, State ZIP: Belfast, Northern Ireland, BT9 6AD
Work Phone: +44 (0)28 90 554091
E-mail: stephen.read@celerion.com

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.

Other brand and product names are trademarks of their respective companies.