# **Adopted Changes for SDTMIG v3.1.3 and 2013 OpenCDISC Upgrades** Liu, Y.<sup>1</sup> and Read, S.<sup>2</sup> Celerion, Lincoln, NE, USA and <sup>2</sup> Celerion Inc., Belfast, Northern Ireland

## INTRODUCTION

There have been several enhancements and upgrades to the Clinical Data Interchange Standards Consortium (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<sup>1</sup> recommendations, in advance of the more recent and comprehensive SDTMIG v3.2 upgrade 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. The 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 poster will give a brief background and overview of some of the key SDTMIG and OpenCDISC updates and detail some potential options in support of their implementation for SAS programmers or mapping specialists working with the newer OpenCDISC validators.

## **SDTM RELEASES BACKGROUND - OVERVIEW**

SDTM first evolved from the Submission Data Model (SDM) as developed by CDISC Submission Data Standards (SDS) team around 10 years ago. The SDS team developed a set of consistent data standards for submitting the tabulation data for human clinical studies.

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 study submissions to the FDA (Table 1).

Date	SDTM Released Version
Jun 2004	In June 2004, the FDA accepted CDISC SDTM as an allowable format for electronic data submissions for SDTM v1.0 (the model), and the SDTMIG v3.1 (the implementation guide)
Apr 2005 Aug 2005	<ul> <li>SDTM Version 1.1 was published</li> <li>STMIG v3.1.1 was published</li> </ul>
Dec 2006	The FDA announced their intention to require SDTM as part of all electronic clinical data submissions
Jul 2007	SDTM v1.2 and SDTMIG v3.1.2 were posted for public comment
Sep 2008	CDISC announced that the FDA selected CDISC to provide training to FDA reviewers of regulatory submissions
Nov 2008	CDISC release of SDTM 3.1.2 IG
Nov 2009	FDA announced that they would accept SDTM v3.1.2 submissions.
Nov 2011	CDISC released SDTM 3.1.2 IG Amendment 1 revisions
Jul 2012	CDSIC released SDTM 3.1.3 SDTM IG
Nov 2013 Dec 2013	<ul> <li>SDTMIG v3.2 was published</li> <li>SDTM v1.4</li> </ul>

### Table 1. SDTM and SDTMIG Timeline for Releases and Updates

## **OPENCDISC RELEASES BACKGROUND – OVERVIEW**

The OpenCDISC validator is a widely accepted open source software application that can be used to assist with the validation of clinical studies 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		ι
Feb 2010 OpenCDISC Validator 1.0.	> V	upport for SDTM alidated both syn uned for usability
Jun 2010 OpenCDISC Validator 1.1		upport for SDTM dded a check for
Jan 2011 OpenCDISC Validator 1.2	> F	upport for SDTM ocused on adding nhanced to furthe
Mar 2012 OpenCDISC Validator 1.3	<ul> <li>A</li> <li>U</li> <li>A</li> </ul>	upport for SDTM dded 43 new SD pdated ADAM ru DaM team dded a check for
Mar 2013 OpenCDISC Validator 1.4	➢ C ➢ In	upport for SDTM ompliance with C nproved linking to 00+ new validatio
Sep 2013 OpenCDISC Validator 1.4.1	▷ V ▷ S	ew ADaM v1.2 va alidation of QS ca everal key bug fiz 43 rules (most ne

 
 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.

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.

### **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. The 2011 Common Data Standards Issues Document (CDER) and most recent FDA submission guidance<sup>4</sup> 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 Study Day (--DY), Study Day of Start (--STDY), Study Day of End (--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. Study day variables describe the 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 is pulled from the demographic domain and normally refers to the date of the first exposure to study treatment.

When the --STDTC variable is contained in the dataset, the --STDY variable should also be included, the same requirement exists for the --ENDY variables. If --DTC is present, --DY should also be populated in the dataset.

Older versions of OpenCDISC did not flag warnings when these variables are missing.

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\_stdt and

### ted Information/Description

del and SDTMIG v3.1.2, 3.1.1 and Define.xml 1.0 standar ectic and semantic constraints using Schematron

2 model and SDTMIG v3.1.2, 3.1.1 and Define.xml 1.0 standards the presence of DM dataset in a submission

#### 2 model and SDTMIG v3.1.2 support for ADaM validation nprove performance and scalability

ccessibility, and Section 508

1.2 model and SDTMIG v3.1.2 Amendment 1 TM validation rules iles to align to the v1.1 release of validation checks from CDIS

#### A 1.3 model and SDTMIG v3.1.3 CDER Common Data Standards Issues Document nultiple CDISC Controlled Terminology versions n rules added for SDTM and ADaM

valid MedDRA version in Define.xml

alidation checks ntrolled terminolog

xes form OpenCDISC 1.4 ew checks are for QS domain and the QS-CT)

ae endt 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;

#### aestdy, ae stdt, fmeddt 1); \*\* first med date equivalent to RFSTDTC; %day(aeendy, ae\_endt, fmeddt\_1);

## **EPOCH**

The CDER Common Data Standards Issues document has recommended the inclusion of an EPOCH variable since 2011 version 1.1 guidance. CDER also recommended the inclusion of more granular study element information across domains although this has yet to be formally adopted due to the difficulties in implementation.

SDTM guidance describes EPOCH: as "an interval of time in the planned conduct of a study during which treatment is constant." Normally EPOCH starts with "Screening EPOCH", "Treatment EPOCH" and then "Follow up EPOCH". Additional terms like run-in and rest can be added as needed, these values must also exist in Trial Arm (TA) domain.

Although the CDER 2011 guidance on EPOCH was fairly clear, prior to 2013 many sponsors may not have regarded its 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 followed the recommendation. Warning message will be populated if epoch variables are missing in certain domains. Table 3 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	СМ	EX	SU				

 
 Table 3. Example of Required 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 4 shows an example of a 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 participant doses on SDTM Visit Day 1, it would obviously makes 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. See example high level EPOCH Macro extract below, leveraging appropriate 'DY' variable, (based on AE start date/time, AE STTM and first medication date/time fmedtm\_1) to associate an appropriate EPOCH with Adverse events.

#### %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';

	- ,						
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	А	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 4. 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 study title, study design, study interventions, assessments, study 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.

**TSVALCD, TSVDREF and TSVCDVER are newer variables** required since SDTMIG v3.1.3. TSVALCD is the code of the term in TSVAL. TSVDREF 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. These new variables 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 list of newer TS parameters that sponsors may not have regarded as a requirement prior to 2013 OpenCDISC upgrades are included in Display 1 highlighted in bold/red. There are five additional parameters conditionally required if the 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).

DOMAIN	TSPARMCD	TSPARM	TSVAL	TS¥AL1	TSYALCD	TSYCDREF	TSYCDYER
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	DCUTDTC	Data Cutoff Date	DBLK date			ISO 8601	
TS	DOSE	Dose per Administration	e.g. 800				
TS	DOSFRQ	Dosing Frequency	e.g ONCE				
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	Υ1				
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	OPENLABEL		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 II/III Trial		C15694	CDISC	12/20/2013
TS	TRT	Investigational Therapy or Treatment	Treatment name				

### **Display 1. Example TS Parameters for New OpenCDISC Required**

To assist in the build, population and review of the TS database, programming teams should consider maintaining 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 worth consideration. Any programming to convert such configurable TS data files to a corresponding SAS xpt file is straight forward.

## **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 charter variable to correspond to the length of the longest term or string contained within the variable.

## **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. 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.

## METHODS FOR MINIMIZING LENGTH-INTRODUCTION TO MINCVARS MACRO

While there are several useful published SAS macros that automatically set the length of selected character variable to match the length of the longest term within the variable. Celerion's 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,

Example of a call to the Mincvars macro in production %mincvars(vit sdtm,rai=vstestcd);

%mend mincvars ;



th of all character e longest actual length. is for the following list of by spaces, do not change their tandards.

ocal macro variable NUMCVARS, character variables

there are zero character 'L') ;

of character variables.;

UMCVARS,'L') ;

minimum lengths of

32;

character variable. ;

VARS[&i])) ;

contains minimum character

#### MPCVAR) ; R) ;

tain As Is parameter. ;  $(R) = \mathbf{0} \text{ THEN}$ " || STRIP(TEMPCVAR) || ' \$ ' &i.,BEST.)) ; able " TEMPCVAR " was not

\_') ;

variables.;

## **REVIEWERS GUIDE**

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 team<sup>5</sup>. 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 and essential requirement for SDTM programmers to consider.

## CONCLUSION

This paper only covers some of the key recent changes and additions that SAS programmers may have encountered in 2013 while 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 specialists interested well into the future as the industry works to continuously improve and streamline the data reporting process.

## REFERENCES

- 1. CDSIC SDTM Implementation Guide Available at: http://www.cdisc.org/sdtm
- 2. CDISC Website. http://www.cdisc.org/
- 3. Open CDISC Website. http://www.opencdisc.org/
- 4. CDER Common Data standards Issues document Version 1.1. December 2011. Available at: http://www.fda.gov/downloads/
- Drugs/DevelopmentApprovalProcess/ FormsSubmissionRequirements/ElectronicSubmi ssions/UCM254113.pdf
- 5. 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, including Larry Gregg, Jenny Erskine and Anthony (Miao) Zheng.

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 Lincoln, NE 68502 Phone: 402-437-4778

www.celenion.com

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