

Therapeutic Area standards and their impact on current SDTM implementations

Johannes Ulander, S-cubed, Copenhagen, Denmark
Niels Both, S-cubed, Copenhagen, Denmark

ABSTRACT

The CDISC SDTM Implementation Guide (SDTMIG) has grown quite considerably with version 3.2, and the volume of CDISC standards is increasing with Therapeutic Area User Guides (TAUGs) being released at a rapid rate and shaping new versions of the SDTMIG, with version 3.3 scheduled for release at the end of 2015. More than 12 Therapeutic Area Standards have been released since 2011 from Alzheimer's to Virology.

We have performed a detailed analysis of the impact of these TAUGs on current implementations of SDTM with focus on the different implementations of Medical History.

Our findings illustrate variations in the use of Medical History across therapeutic areas: therapeutic areas are using the same variables, but in different ways. The variations have an impact on metadata and mapping across therapeutic areas, requiring changes in some cases where SDTM has already been implemented. Although some of the variation may be necessary due to specific needs of a given therapeutic area, many of the variations should be possible to eliminate. We recommend a more consistent usage of Medical History in future versions of SDTM and TAUGs.

Introduction

Since the first introduction of the CDISC SDTM, the model has been refined and extended considerably. For several years CDISC SEND has worked as an extension to the overall SDTM model, enabling SDTM to be applied to non-clinical trials. Over the last few years more extensions have been added at a steady pace, such as "Associated Persons Implementation Guide" and "SDTM Implementation Guide for Medical Devices."

Recently several therapeutic area user guides (TAUG) have also been added, providing information about how to utilize the CDISC Models for particular therapeutic areas. The addition of the TAUGs to the models has been carried out at a relatively high pace.

Pipeline of TAUG releases or updates						
	2011	2012	2013	2014	2015	2016
#TAUGs	1	4	7	9	12	10

Numbers compiled from http://cdisc.org/system/files/all/CFAST_ProjectPipeline20150108.pdf

The CDISC organization is implementing an electronic library named SHARE to hold the metadata about the models in an electronic form. Until the SHARE library is fully developed, the TAUGs are published in pdf format, which makes it difficult to maintain a detailed precise overview of the variations or extensions introduced to the CDISC Models.

This paper is a result of work done in the spring of 2015 to analyse all current TAUGs, with focus on the SDTM domains only. The medical history domain MH is used as an example to determine the amount of variation within a specific domain.

Background

In the spring 2015, all of the (at that time) 12 current TAUGs were examined. An overview was created of which domains were used across the 12 TAUGs. The 12 TAUGs examined are listed below:

- Alzheimer's version 2
- Asthma version 1
- Cardiovascular version 1
- Diabetes version 1
- Dyslipidaemia version 1 (Draft)

PhUSE 2015

- Multiple Sclerosis version 1
- Pain version 1
- Parkinson's Disease version 1
- Polycystic Kidney Disease version 1
- Schizophrenia version 1(Draft)
- Tuberculosis version 1
- Virology version 1

The pdf documents for each TAUG are available via CDISC.ORG. The naming convention of the pdf files is not consistent across the TAUGs, but all can be retrieved using the links on CDISC.ORG.

Origin of domains used in Therapeutic Area User Guides

The various TAUGs listed utilize a number of different SDTM domains. The origin of the 53 domains used in TAUGS are distributed across:

- SDTMIG v3.2
- SDTMIG Medical Devices v1.0 (SDTMIG-MD v1.0)
- SDTMIG Associated Persons v1.0 (SDTMIG-AP v1.0)
- Draft domains proposed (TAUG Draft domains not yet in any SDTMIG)

This distribution of origin of the 53 domains used in TAUGs is illustrated in the below tables:

SDTMIG 3.2			SDTMIG-MD v1.0		TAUG Draft domains not yet in any SDTMIG			
AE	IS	RP	DE	Device Events	AG	Procedure Agents	PB	Pgx Biological State
CE	LB	SC	DI	Device Identifier	BE	Biospecimen Events	PF	Pharmacogenomics Findings
CM	MB	SR	DO	Device Options	BS	Biospecimen	PG	Pharmacogenomics
DD	MH	SU	DR	Device Subject Relationship	ED	Education Details	QT	ECG QT Correction Model Data
DM	MI	TA	DU	Device In-Use	ML	Meals	SB	Subject Biological State
DS	MO	TE	SDTMIG-AP v1.0		NV	Nervous Systems Findings	UR	Urinary System
EG	MS	TR	APBS		OE	Ophthalmic Exams	VR	Viral Resistance
EX	PR	TS	APDD					
FA	QS	TU	APDM					
HO	RE	VS	APMH					

Table 1. Origin of TAUG domains

Most frequently used domains in Therapeutic Area User Guides

Our initial intention was to go through all domains, but as the number of domains exceeded our expectations, and because some domains have only been used in one TAUG, we decided to focus our analysis on one of the domains used most frequently in the TAUGs.

The following table shows which domains are most commonly used across TAUGs:

#TAUG	Domain	Alz	Asth	CV	Diab	Infl	MS	Pain	PD	PKD	QT	TB	Vir.
9	DI	X	X	X	X	X	X		X	X	X		
9	FA		X	X	X	X	X	X	X	X		X	
9	MH	X	X	X	X		X	X	X	X		X	
7	CE			X	X	X	X		X	X		X	
7	LB	X	X	X	X	X				X		X	
6	CM		X		X	X			X	X		X	
6	MO	X		X			X		X	X		X	
5	DU	X	X				X		X			X	
5	PR	X		X		X			X	X			

PhUSE 2015

#TAUG	Domain	Alz	Asth	CV	Diab	Infl	MS	Pain	PD	PKD	QT	TB	Vir.
4	HO		X	X		X				X			
4	PF	X				X				X			X

Table 2. List of domains used in at least 4 TAUGs

As shown in table 2 above, the three domains used most across TAUGs are:

- DI (Device Identifiers)
- FA (Findings About)
- MH (Medical History)

DI was considered for further analysis, but such an analysis should be performed in connection with the other Device domains.

The FA domain is used to store findings about observations in a variety of other domains. In order to make a detailed analysis of FA, it would be most meaningful to do so as part of an analysis of each of the domains related to FA.

MH was chosen for the more detailed analysis as it is a domain which is part of SDTM for the vast majority of clinical trials, and is a domain which all regular users of the SDTMIG are familiar with.

Analysis – Medical History

According to the SDTMIG, the medical history dataset includes the subject's prior history at the start of the trial. It can be either disease specific medical history or general medical history.

Categorisation using MHCAT and MHSCAT

Generally the SDTM variables --CAT and --SCAT are used to categorise the information on separate records. Traditionally SDTM could use MHCAT to distinguish between general MH and disease specific MH. Across the various TAUGs analysed, MHCAT and MHSCAT are used in a number of different ways.

The terms used for categorisation in the TAUGS are outlined in table 3 below.

General medical history

The method to identify general medical history using MHCAT also varies across the TAUGs, which can be seen in table 3. Most of the TAUGs suggest to use MHCAT="GENERAL" to capture the information. Some of them do not give any recommendation. Of special interest is that the TAUG for Parkinson's disease recommends using MHCAT=null for general medical history.

Primary diagnosis

In general there are three different methods in use for identifying the primary diagnosis using MHCAT:

1. MHCAT = PRIMARY DIAGNOSIS
2. MHCAT = STUDY INDICATION
3. MHCAT = Name of Disease
4. FA approach, FAOBJ = MHCAT/MHTERM

Table 3 indicates which TAUG uses which approach.

	MHCAT	MHSCAT	Approach
Alzheimer	PRIMARY DIAGNOSIS		1
	GENERAL		
Asthma	<i>Asthma Related</i>		4
Diabetes	DIABETES	COMPLICATIONS OF DIABETES	3
Dyslipidemia	Hypercholesterolemia		1
	Cardiovascular Risk Factors		

PhUSE 2015

	MHCAT	MHSCAT	Approach
Multiple Sclerosis	PRIMARY DIAGNOSIS	ONSET COURSE	1
	GENERAL	INITIAL SYMPTOMS	
Pain	PRIMARY DIAGNOSIS		1
Parkinson's Disease	NULL	DIAGNOSTIC CRITERIA	1
	PRIMARY DIAGNOSIS	INITIAL MOTOR SYMPTOMS	
	PARKINSON'S DISEASE (for symptoms)	DEVELOPING MOTOR SYMPTOMS	
	INFECTION HISTORY		
Polycystic Kidney Disease	PRIMARY DIAGNOSIS		1
	PKD (for symptoms)		
Schizophrenia	GENERAL		3
	SCHIZOPHRENIA		
Tuberculosis	STUDY INDICATION		2
	RISK FACTORS		
	COMORBID CONDITION		
	GENERAL		
	TB SYMPTOMS		

Table 3. List of categories and sub-categories (Note: text case is from the respective TAUG)

Note 1: Cardiovascular and Virology TAUG did not specify any categories

Note 2: Parkinson's disease also uses a 2nd sub-category (SUPPMH.MHSSCAT)

As the SDTMIG does not give guidance on how to distinguish between symptoms and diagnosis of a disease, some variation is expected. The most commonly used approach (approach 1) for capturing primary diagnosis is to set MHCAT="PRIMARY DIAGNOSIS".

mh.xpt

Row	STUDYID	DOMAIN	USUBJID	MHSEQ	MHTERM	MHDECOD	MHCAT	MHSTDTC
1	ABC123	MH	AD01-101	1	Mild Cognitive Impairment		PRIMARY DIAGNOSIS	2001-05
2	ABC123	MH	AD01-101	2	Blind Left Eye	Blindness unilateral	GENERAL	1999-07-07
3	ABC123	MH	AD01-101	3	Traumatic Brain Injury	Traumatic brain injury	GENERAL	2005-03-28
4	ABC123	MH	AD01-102	1	Alzheimer's Disease		PRIMARY DIAGNOSIS	2003-05

Example from Alzheimer TAUG

Approach 2 is to set MHCAT=STUDY INDICATION on the disease under study and using categories that explain the relationship of the diagnosed disease to the disease under study.

mh.xpt

Row	STUDYID	DOMAIN	USUBJID	MHSEQ	MHTERM	MHDECOD	MHCAT
1	ABC	MH	ABC-01-101	1	PULMONARY TUBERCULOSIS	Pulmonary tuberculosis	STUDY INDICATION
2	ABC	MH	ABC-01-101	2	COPD	Chronic obstructive pulmonary disease	COMORBID CONDITION
3	ABC	MH	ABC-01-101	3	TYPE II DIABETES	Type 2 diabetes mellitus	RISK FACTORS

Example from Tuberculosis TAUG

Approach 3 is to set MHCAT=Name of Disease

mh.xpt

Row	STUDYID	DOMAIN	USUBJID	MHSEQ	MHTERM	MHCAT	MHDTTC	MHSTDTC
1	XYZ	MH	XYZ-001-001	1	TYPE 1 DIABETES	DIABETES	2010-09-26	2010-03-25
2	XYZ	MH	XYZ-001-002	1	TYPE 2 DIABETES	DIABETES	2010-10-26	2010-04-25

Example from Diabetes TAUG

PhUSE 2015

A rather different approach (approach 4) is used in the Asthma TAUG, as it recommends using FAMH (Findings about Medical History) to store diagnosis information, as the diagnosis may be implicit (i.e. not captured as medical history) and the MHSTDTC does not have a precise definition and might refer to date of diagnosis or the start of some other event related to the disease (e.g. symptoms start date.)

famh.xpt

Row	STUDYID	DOMAIN	USUBJID	FASEQ	FATESTCD	FATEST	FAOBJ	FACAT	FAORRES	FASTRESC	FADTC
1	ABC123	FA	456	1	DIAGDTC	Date of Diagnosis	Asthma	Asthma History	1992	1992	2013-05-24
2	ABC123	FA	456	2	SUBTYPE	Subtype	Asthma	Asthma History	Exercise induced	Exercise induced	2013-05-24

Example from Asthma TAUG

Date of diagnosis vs. date of related symptoms

In all TAUGs, there is a clear need to clearly differentiate between the onset symptoms of a disease, which can occur before the disease have been diagnosed, and the date of diagnosis. Most TAUGs use MHSTDTC for date of diagnosis.

ROW	USUBJID	MHSEQ	MHTERM	MHCAT	MHSTDTC
1	2324-P0001	1	Fibromyalgia	PRIMARY DIAGNOSIS	2001-05

Example from Pain TAUG. MHSTDTC = Date of diagnosis

But there are examples of alternatives, such as duplicating the information into a supplemental qualifier:

Row	STUDYID	DOMAIN	USUBJID	MHSEQ	MHTERM	MHCAT	MHDTIC	MHSTDTC
1	XYZ	MH	XYZ-001-001	1	TYPE 1 DIABETES	DIABETES	2010-09-26	2010-03-25

suppmh.xpt

Row	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QVAL
1	XYZ	MH	XYZ-001-001	MHSEQ	1	MHDXDTC	Date of Diagnosis	2010-03-25	CRF	INVESTIGATOR

Example from Diabetes TAUG: MHSTDTC = SUPPMH.MHDXDTC = Date of diagnosis

Or using FA, shown above in the example from the Asthma TAUG.

The preferred solution on how to distinguish “symptoms” versus diagnosis might be found in the Schizophrenia TAUG. The Schizophrenia TAUG introduces a new variable called MHEVTYP. So MHEVTYP can be set to “DIAGNOSIS” for diagnosis and to “SYMPTOMS” for symptoms.

Row	STUDYID	DOMAIN	USUBJID	MHSEQ	MHGRPID	MHTERM	MHDECOD	MHCAT	MHEVTYP	MHSTDTC	MHENDTC
3	ABC123	MH	001-001	3	1	AGITATION	Agitation	SCHIZOPHRENIA	SYMPTOMS	2003-11	2004-02-14
4	ABC123	MH	001-001	4	1	DISORGANIZED SCHIZOPHRENIA	Disorganized schizophrenia	SCHIZOPHRENIA	DIAGNOSIS	2004-01	

Example from Schizophrenia TAUG

This approach from the Schizophrenia TAUG is probably going to be implemented in the up-coming SDTM version as the variable was present in the recently publicly reviewed draft of Study Data Tabulation Model Version 1.5, where it is part of a new section for “Domain-Specific Variables for the General Observation Class”.

MH	MHEVTYP	Medical History Event Type	Char	Record Qualifier	Specifies the aspect of the medical condition or event by which MHSTDTC is defined. Examples: DIAGNOSIS, SYMPTOMS, RELAPSE, INFECTION.	Before MHSTAT

New variable in SDTM v1.5 (draft)

Conclusion & recommendations

A great amount of effort has been put into the development of the TAUGs. We are happy to see how the development is aiding in the work to extend and improve future versions of CDISC standards.

The analysis of these 12 first TAUGs shows a level of variation on how Medical History information is stored in SDTM that cannot be justified by the diseases being different, but which may be the result of a very fast paced development.

PhUSE 2015

Reduce variation

We recommend that future versions of the TAUGs and of SDTMIG decrease the amount of variation as much as possible. It is feasible to express some of the principles for how to standardize medical history in a more generic way and apply this to all TAUGs. In the cases where the principles can be made generic they can be described in the SDTMIG, rather than in each individual TAUG. A good example of this is the development of handling date of diagnosis vs. onset of symptoms, which is expected to be included in SDTMIG v1.5 in the way suggested by the Schizophrenia TAUG.

More granular versioning

We foresee the need for more frequent updates of the SDTM and SDTMIG to aid in alignment between the TAUGs. A new option could be considered to version on a domain level, making it possible to develop and align standards at a faster rate; maybe this is not feasible until the SHARE library is more widely adopted.

Standards hierarchy

The relationship of the TAUGs and other CDISC standards needs to be clarified further so that it is clear which standard is applicable in cases of ambiguity, such as:

- Should TAUGs released after Schizophrenia use MHEVTYP even though it has not been implemented in SDTM yet?
- Once the MHEVTYP is implemented in the SDTM, does it then supersede what is written in other TAUGs?
- Is it OK to submit the draft ML domain found in the Diabetes TAUG before it has been implemented in SDTMIG?
- Is it OK to use the ML domain in other Therapeutic Areas?

These are not simple questions because the TAUGs also cover the other CDISC standards. Currently there is no place to find a formally correct answer to these questions or guidance about which standards take highest precedence.

Working with Medical History

In order to support standardised use of Medical History across therapeutic areas for the time being, we at S-cubed will take this work a step further, recommending a best practice for handling the inconsistencies cited in this paper.

ACKNOWLEDGMENTS

The authors would like to thank our colleague Heather Brage for help provided during the editorial process.

Contact Information

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

Johannes Ulander/Niels Both
S-Cubed ApS
Lille Strandstraede 20C, 5
Copenhagen / 1254
Web: <http://www.s-cubed.dk>

Brand and product names are trademarks of their respective companies.