PharmaSUG 2024 - Paper DS- 274

Guidance Beyond the SDTM Implementation Guide

Kristin Kelly and Michael Beers, Pinnacle 21 by Certara

ABSTRACT

A common misconception among preparers of SDTM data seems to be that it is sufficient to just follow the SDTM Implementation Guide when creating the datasets. The truth is that it is more complicated than that. A preparer of SDTM datasets needs to be aware of all the industry guidance available when preparing for regulatory submission, from CDISC and the regulatory agencies, but also from other organizations as well. This presentation will discuss some of the lesser-known guidance in the industry and why they should be referenced, as well as some of the impacts of not using these documents in the creation of SDTM datasets.

INTRODUCTION

When preparing SDTM datasets for submission, further guidance may be needed to fill in some of the gaps in the standards that are not outlined in the SDTM Model or the SDTM Implementation Guide (SDTMIG). There are also regulatory requirements that must be followed that may differ from what is in the CDISC standards. Clarifying information on what is expected or how to handle certain types of data becomes necessary. But it can be difficult to determine if such documentation exists and if so, where to find that information. Over time, there has been substantial guidance published from many different sources to aid implementers in preparing submission deliverables.

This paper will focus on guidance that is available aside from the SDTMIG that can help facilitate SDTM dataset mapping/creation but there will also be discussion on guidance for SDTM submissions in general from CDISC, Phuse, regulatory authorities and others.

CDISC GUIDANCE

In addition to referencing the SDTMIG when preparing datasets for submission, CDISC has published other materials that may be relevant. The following sections are CDISC documents, though not exhaustive, that should be reviewed if applicable for a particular study.

CDISC THERAPEUTIC AREA USER GUIDES

Though the SDTMIG covers much of the data commonly collected in clinical trials, there are some gaps, mostly due to data specific to certain therapeutic areas (TA). CDISC Therapeutic Area User Guides (TAUGs) were developed to fill some of these gaps.

Each TAUG contains information specific to a therapeutic area as well as examples of the type of data that might be collected in a study. All TAUGs contain SDTM examples and most also contain CDASH-compliant CRF representations as well as ADaM guidance. Over time, much of the new standards developed, such as new variables and SDTM domains, have first appeared in TAUGs. Like the IGs, each TAUG's content is consensus-based, meaning that each must go through both Internal and Public Review before it can be published.

It is important to note that when a TAUG is published by CDISC, it is considered 'Provisional' because there may be new variables and domains that have not yet been added to the SDTM Model and/or the SDTMIG. For this reason, SDTM examples in TAUGs are considered 'informative' and not 'normative' content. Normative content is standard and should be followed as closely as possible. Informative content supports the normative content in the form of examples and best practices. It is recommended that when using new variables in SDTM from a TAUG, the variable is non-standard and should be mapped to SUPPQUAL until it appears in a future version of the SDTM Model. New domains based on one of the three general observation class domains (Events, Interventions, or Findings) will be considered custom domains until they are added to an SDTMIG. Also, if a new version of a TAUG is published for a particular TA, it supersedes the previous version. Any new variable and domain guidance that has been added, changed or removed in the more recent version, should take precedence.

If a TAUG is referenced when preparing the SDTM data, it should be listed in the Trial Summary (TS) domain where TSPARMCD/TSPARM = 'CTAUG'/'CDISC Therapeutic Area User Guide'. This parameter is included in the FDA Technical Conformance Guide (TCG), Appendix B as a 'Conditional' parameter in TS. The table below indicates the condition under which this parameter should be included as well as controlled terminology to be used for TSVAL.

Conditional	CTAUG	CDISC Therapeutic Area User Guide	If applicable. the value should be the exact listing as in section 5.2 of the Technical Conformance Guide. Use as many rows as needed.

Section 5.2 of the TCG lists those TAUGs that have been evaluated and supported by the FDA. It is stated that if sponsors use a TAUG that is not listed, the rationale should be included in the Clinical Study Data Reviewer's Guide (cSDRG).

The following is a list of TAUGs published thus far and available on the CDISC website (https://www.cdisc.org/standards/therapeutic-areas).

- Acute Kidney Injury
- Alzheimer's
- Asthma
- Breast Cancer
- Cardiovascular
- CDAD
- Colorectal Cancer
- COPD
- COVID-19
- Crohn's Disease
- Diabetes
- Diabetes Type 1 - Exercise and Nutrition
- Diabetes Type 1 - Pediatrics and Devices
- Diabetes Type 1 - Screening, Staging and Monitoring of Pre-clinical Type 1 Diabetes
- Diabetic Kidney Disease
- Duchenne Muscular Dystrophy
- Dyslipidemia
- Ebola
- Heart Failure
- Hepatitis C
- HIV
- Huntington's Disease
- Influenza
- Kidney Transplant
- Lung Cancer
- Major Depressive Disorder
- Malaria
- Multiple Sclerosis
- Nutrition
- Pain
- Pancreatic Cancer
- Parkinson's Disease
- Pediatrics
- Polycystic Kidney Disease
- Post Traumatic Stress Disorder
- Prostate Cancer
- Psoriasis
- QT Studies
- Rare Diseases
- Rheumatoid Arthritis
- Schizophrenia
- Traditional Chinese Medicine - Acupuncture
- Traditional Chinese Medicine - Coronary Artery Disease-Angina
- Traumatic Brain Injury
- Tuberculosis
- Vaccines
- Virology

Any TAUG in this list that is used should be cross-referenced with Section 5.2 of the TCG to determine if the TAUG is supported by FDA.

SDTM IMPLEMENTATION GUIDE – ASSOCIATED PERSONS

Sometimes in a clinical trial, data are collected about persons that are not study participants. These persons could be associated with the study itself, a study subject, or a device used in the study. The SDTMIG-Associated Persons (SDTMIG-AP) contains guidance for modeling the data collected for associated persons in SDTM.

A common example of this data might be collecting family medical history of the disease under study. Since this data are not about the subject, it needs to be kept separate from the subject data. Using this scenario, the subject's medical history would be found in the Medical History (MH) domain in SDTM. Data collected about the family's medical history would be mapped to the Associated Persons Medical History (APMH) domain. Each Associated Persons domain has the prefix 'AP' appended before the standard domain code, in this case, MH. Except for a few identifier variables specific to AP domains: APID, RSUBJID, SREL and the omission of USUBJID, the structure of the dataset will still use the same SDTM metadata structure and conventions for that domain outlined in the SDTMIG.

apmh.xpt

Row	STUDYID	DOMAIN	APID	MHSEQ	RSUBJID	SREL	MHTERM	MHPRESP	MHOCCUR
1	2011-02-02	APMH	2011-02-02-N120	1	2011-02-02-031	MOTHER, BIOLOGICAL	POMPE DISEASE	Y	N
2	2011-02-02	APMH	2011-02-02-N121	1	2011-02-02-031	FATHER, BIOLOGICAL	POMPE DISEASE	Y	Y
3	2011-02-02	APMH	2011-02-02-NS122	1	2011-02-02-031	SIBLING, FULL	POMPE DISEASE	Y	N
4	2011-02-02	APMH	2011-02-02-NS123	1	2011-02-02-031	COUSIN, BIOLOGICAL	POMPE DISEASE	Y	Y

In the APMH snippet above, the AP identifiers are included as well as the topic variable for MH, MHTERM, and additional SDTM standard Qualifier variables, MHOCCUR and MHPRESP. Please note the DOMAIN variable value in the dataset is 'APMH', the same as the dataset name (apmh.xpt). This is

similar to naming standard SDTM domains/datasets. The SDTMIG-AP also includes guidance on naming SUPPQUAL datasets for AP domains.

It is important to note that the SDTMIG-AP is published as an implementation guide with a status of 'Final' (and not 'Provisional'). Because of this, it is considered standard content as an extension of the SDTMIG.

SDTM IMPLEMENTATION GUIDE – MEDICAL DEVICES

The SDTMIG-Medical Devices (SDTMIG-MD) should be referenced when data is collected about devices in a clinical trial. The device does not have to be the focus of the study. It can be any device used in the study for which data was collected. Like the SDTMIG-AP, it is published as 'Final' and considered standard content as an extension of the SDTMIG. The current version, SDTMIG-MD v1.1, supersedes the previous SDTMIG-MD v1.0.

There are 7 device domains defined in the SDTMIG-MD:

- Device Identifiers (DI)
- Device-In-Use (DU)
- Device Exposure (DX)
- Device Events (DE)
- Device Tracking and Disposition (DT)
- Device-Subject Relationships (DR)
- Device Properties (DO)

Any of these domains can be used for any study type, if deemed appropriate by sponsors and regulators and if sufficient data was collected about the device to populate them. Though there are several examples of these domains in the SDTMIG-MD v1.1, several TAUGs have these domains modeled in examples as well. It should be noted that even though extensive data about devices is represented in these examples, it does not mean that sponsors need to collect data in their study to the same level of detail. The data to be collected and represented in SDTM would depend on what is outlined in the protocol and the Statistical Analysis Plan (SAP).

QUESTIONNAIRES, RATINGS AND SCALES SUPPLEMENTS

CDISC publishes supplements for Questionnaires, Ratings and Scales (QRS) that are either in the public domain or copyright-approved instruments. These supplements contain information about the instrument as well as guidance and examples for representing the data in SDTM. QRS instrument data are mapped to one of three domains: Questionnaires (QS), Functional Tests (FT), and Clinical Classifications and Disease Response (RS). For preparers of SDTM, these supplements can prove useful in mapping the instrument to the appropriate domain as well as properly structuring the data. QRS instruments that will be administered in a study are outlined in the study protocol.

When preparing to create the SDTM datasets for QS, FT, and/or RS, it is recommended to check the CDISC website (https://www.cdisc.org/standards/foundational/qrs#qrs__supplements) to determine if there is a QRS supplement created for a specific instrument. Also, each CDISC TAUG lists QRS instruments that may apply for a specific TA and lists the QRS supplement status, e.g. 'In progress', 'Final', etc. This information helps guide the person preparing the SDTM datasets to the QRS supplements.

Like TAUGs, QRS supplements also go through Internal and Public Review prior to being finalized. These supplements are considered extensions of the standards outlined in the SDTMIG. When the supplement is authored, the latest version of the SDTMIG is used, but can be applied for any SDTMIG version and modified appropriately, if necessary.

SDTMIG ERRATA AND ERRORS THAT AFFECT CONFORMANCE

Occasionally, errors are identified after the SDTMIG is published. To correct these mistakes, CDISC publishes 'Errata' to rectify them. Errata are located on the CDISC website and are kept on the page for that specific version of the SDTMIG under the 'Errors' tab. The screenshot below shows the page for SDTMIG v3.3. (https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-v3-3)

SDTMIG v3.3

Release Information | Files & Links | Related Standards | Errors

Errata

An erratum (plural: errata) is a correction to a mistake made in a published standard, which was identified after it was published on the CDISC website, and which would have been corrected in the current version had it been identified prior to publication. Updates, revisions, **substantive** corrections, and other changes of enough significance to require a cycle through the Standards Development Process are not errata, and thus are not included on the pages below.

SDTMIG v3.3

Errata typically correct text in the SDTMIG but the update does not affect conformance to SDTM. One example is when the codelist name for DM.COUNTRY was corrected in SDTMIG v3.3:

Section 5.2, Demographics

- DM – Specification: The value in the "Controlled Terms, Codelist or Format" column for COUNTRY should be changed from "ISO 3166-1 Alpha-3" to "ISO 3166-1 alpha-3".

When the error affects conformance and in turn, validation, the issue is not published as an 'Errata' but as an 'Error that Affects Conformance'. These would also be located on the 'Errors' tab for the SDTMIG version on the CDISC website. These 'Errors that Affect Conformance' outline the issue with the standard as well as coping strategies for how to handle the issue that may include explaining the issue in the cSDRG for the specific validation rule that is triggered.

For example, in SDTMIG v3.3, the specification for the Disposition (DS) domain lists the variable, DSDY, as 'Expected' meaning it must always be included in the dataset whether it is populated or not. However, the date variable from which DSDY is derived, DSDTC, is 'Permissible', meaning that it can be included if data were collected to populate it. If DSDTC is not included in the dataset and with DSDY having a 'Core' of 'Expected', this will trigger a validation rule that an 'Expected' variable is missing. The screenshot below shows the 'Error that Affects Conformance' that was published for this issue in DS. It shows what version the issue appears, a description of the issue, what the Core value should be as well as the strategy for coping with the issue.

DSDY should be "Permissible", not "Expected"

Short Name	DSDY should be permissible, not expected
Affected Standard	SDTMIG v3.3
Description of Error	In the DS domain specification, the core value of DSDTC is "Perm" but the core value of the corresponding study day variable, DSDY, is "Exp". The study day variable corresponding to a date/time variable that is "Perm" cannot be "Exp". Note that DSDY was not included in the DS doman specification in previous versions of the SDTMIG.
Efforts to Correct Error	The core value for DSDY will be changed to be "Perm" in the next version of the SDTMIG.
JIRA Issue	SDS-1417

Concerned Published Element	Concerned Published Attribute	Published Attribute Value	Revised Attribute Value
DS.DSDY	Core	Exp	Perm

Impact of Issue	Coping Strategy
If DSDTC was not collected, so that DSDY is always null, failure to include DSDY in the DS dataset may result in errors or warnings.	Explain any validation errors or warnings in the Clinical Study Data Reviewers Guide (cSDRG).

Please note that with 'Errors that Affect Conformance', the error is not formally corrected in that version of the SDTMIG, i.e., the standard is not changed, and this is why coping strategies are provided. In this example, the issue should be explained in the cSDRG.

OTHER GUIDANCE FROM CDISC

CDISC also publishes other guidance that are helpful for submission deliverables though not directly related to preparing SDTM datasets. Some of these resources include:

- Knowledge Base (KB)
  - CDISC now maintains a KB on the CDISC website: https://www.cdisc.org/kb
  - Implementers are encouraged to visit the KB for useful articles, information on Known Issues, examples that are not found in the SDTMIG as well as example CRFs in the eCRF Portal.
- Define-XML v2.0 and v2.1
  - The define.xml file is required for all SDTM submissions and should be structured as outlined in the Define-XML standards. Currently, FDA accepts both Define-XML v2.0 and v2.1.
- SDTM Metadata Submission Guidelines (MSG) v2.0
  - SDTM-MSG v2.0 provides guidance for preparing the different components required for a submission: Annotated Case Report Form (aCRF), SAS v5 XPT files, define.xml, and the cSDRG.
  - The information in this document is considered to be informative rather than normative and should be used in conjunction with CDISC standards.
  - An example submission package is included with the MSG document.

PHUSE GUIDANCE

Like CDISC, Phuse is an organization that's driven by volunteers. The main difference is that Phuse does not develop new standards but instead publishes deliverables and guidance that support the CDISC standards and regulatory requirements. (https://phuse.global/Deliverables/1)

BEST PRACTICES FOR SUBMISSION OF EVENT ADJUDICATION DATA

The whitepaper, 'Best Practices for Submission of Event Adjudication Data', provides a standardized approach for handling adjudication data that may be collected in a clinical trial, regardless of the specific therapeutic area. Adjudication is a committee's blinded evaluation of specific endpoints/events in a clinical trial. The events that are reviewed are typically subjects' adverse event data collected in a study. Though some of the CDISC TAUGs provide some guidance, there is not much on how to handle this type of data in SDTM. This whitepaper outlines a proposal to standardize how adjudication is submitted so that sponsors can follow one approach that would help create consistency across industry.

The Phuse team proposed the domain, Event Adjudication (EA), to house this data. It is a Findings About-structured domain where EAOBJ is populated with the event that is being adjudicated. This approach also moves this data out of the Findings About (FA) domain which can become quite large in some trials. The CDISC SDS Team plans to add the EA domain to the next version of the SDTMIG v4.0. By doing so, it becomes normative content as a standard to be applied across the industry.

OTHER GUIDANCE FROM PHUSE

There is a substantial amount of guidance from Phuse that aids in preparing submission deliverables such as define.xml and the cSDRG as well. A few useful ones that are not industry standard templates, such as for the cSDRG, that should be referenced are listed below.

- Define-XML v2.0 Completion Guidelines
  - The intent of this document is to further clarify some of the more challenging metadata items in the define.xml file. It not meant to repeat what is already outlined in CDISC's Define-XML standard. Though this is geared towards Define-XML v2.0, most of the guidance could also be applied to Define-XML v2.1.
- Best Practices for Documenting Dataset Metadata: Define-XML Versus Reviewer's Guide
  - This whitepaper provides further guidance on documenting the dataset metadata for a trial and provides recommendations on where to best add this information: define.xml vs cSDRG.
  - It also includes some best practices for explaining validation results in the cSDRG.

REGULATORY GUIDANCE

The regulatory agencies for the U.S. (FDA) and Japan (PMDA) provide numerous resources that a preparer of SDTM data must be aware. These documents are provided at the following locations:

- For the FDA, see the Study Data Standards Resources webpage, located here: https://www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources
- For the PMDA, see the New Drug Review with Electronic Data webpage, located here: https://www.pmda.go.jp/english/review-services/reviews/0002.html

REGULATORY AGENCY DATA STANDARDS CATALOGS

The FDA and PMDA, and it is expected that other regulatory agencies will at some point as well, provide a data standards catalog which lists the versions of the standards and terminologies that are supported/required for use. It is critical that preparers of SDTM data ensure that the versions of the standard and terminologies planned for use are supported by the agency that will be reviewing the application.

REGULATORY AGENCY STUDY DATA TECHNICAL CONFORMANCE GUIDES AND FAQS

The FDA and PMDA each publish a Study Data Technical Conformance Guide (sdTCG), which describes the agencies' preferences, recommendations, guidance, expectations, or even requirements for how they would like (or require) study data to be prepared and submitted. This includes the agencies' considerations regarding the following: exchange format, eCTD format, planning and providing standardized study data (SDSP, cSDRG/nSDRG/ADRG), study data submission format and conventions, data definition files, therapeutic area information, terminology, study data validation and traceability, trial summary parameters, and more.

While the SDTM Implementation Guide instructs how to provide the collected trial data in standard format in a more general way, each agency may have specific preferences for how they expect to see certain data represented. Two example scenarios are provided:

- Example 1: CDISC does not, as of the time of this paper, have published guidance for how to map multiple enrollment data. The FDA, however, has described their preferred approach in section 4.1.1.3 of the Study Data Technical Conformance Guide. Hopefully, when this mapping guidance is provided by CDISC in a new version of the SDTM Implementation Guide, it will align with the agency's expectations.
- Example 2: CDISC provides guidance for how to map standard laboratory units to SDTM variables, but currently does not provide guidance on how to submit multiple (different) sets of standard laboratory units. The FDA does provide their preference for how this is handled in section 4.1.1.3 of the sdTCG.

The PMDA also publishes FAQs on Electronic Study Data Submission, which is used to answer questions from the industry and further clarify PMDA expectations on preparing study data for submission.

Occasionally there are differences in expectations listed in a regulatory agency's Study Data Technical Conformance Guide or FAQs, and guidance provided by CDISC. As a general rule, the recommendation is to follow the regulatory agency expectations, or at least to discuss with the regulatory agency how to handle the discrepancy.

FDA TECHNICAL SPECIFICATION DOCUMENTS

The FDA publishes therapeutic area specifications that describe their expectations for how certain specific data is created and submitted. These are listed in the FDA's Technical Conformance Guide, section 5.3, as shown below:

The FDA tech specs that contain SDTM guidance have been marked with a green check mark. The SDTM-related tech specs are:

- Tech Spec – Vaccines
  - This document contains FDA's expectations/recommendations for how to represent safety and efficacy data specific to vaccine trials.
- Tech Spec - Noncirrhotic Nonalcoholic Steatohepatitits (NASH)
  - Section 1.0 states: *This document provides detailed information and specifications for the content of the tabulated domains and analysis data sets submitted to FDA as part of the sponsor's/applicant's application for drugs intended to treat noncirrhotic NASH.*
- Tech Spec - Submitting Patient-Reported Outcome Data in Cancer Clinical Trials
  - Section 1.0 states: *This document provides technical specifications for submitting patient-reported outcome (PRO) data collected in cancer clinical trials to support a marketing application for a medical product in oncology, where a PRO is a type of clinical outcome assessment (COA) used to collect patient experience data.*
- Tech Spec - Submitting Clinical Trial Datasets and Documentation for Clinical Outcome Assessments Using Item Response Theory
  - Section 1.0 states: *This document provides technical specifications for the submission of clinical outcome assessment (COA) data that use Item Response Theory (IRT) and supplements the FDA Center for Drug Evaluation and Research (CDER) Patient-Focused Drug Development (PFDD) Methodological Guidance Series.*

OTHER GUIDANCE

- Uppsala Monitoring Centre's "How to use WHODrug for compliance with CM domain in the CDISC SDTM standard" is one example of another organization's guidance of which to be aware. This document describes how to map WHODrug data to the CM (and SUPPCM) SDTM domain, specifically focusing on 5 SDTM variables: CMTRT, CMMODIFY, CMDECOD, CMCLAS, and CMCLASCD. This is important because CDISC does not provide much specific guidance on the mapping of this data. There are some important considerations to be aware of in this guidance, such as how to map dictionary values greater than 200 characters, which has an impact on the validation of these datasets, as shown here:

CMDECOD is longer than 200 characters

For drugs with many ingredients, the generic name is longer than 200 characters. The SAS export format has a limitation to 200 characters per field, if this format is used for submission, the supplemental dataset needs to be utilised. Note that the guidelines state that the text should be truncated between words, in the case for long generic names the text should be truncated after the semicolon closest to 200 characters. Illustrations of the ordinary and supplemental datasets are shown in table 1 and 2.

Table 1. Illustration of SDTM dataset where CMDECOD is longer than 200 characters.

USUBJID	CMSEQ	CMTRT	CMMODIFY	CMDECOD	CMCLAS	CMCLASCD
AB-21-01	1	....	...	Ascorbic acid;Biotin;Calcium;Carbohydrates nos; Chloride;Choline;Chromium;Colecalciferol; Copper;Cyanocobalamin;Docosahexaenoic acid; Fats nos;Folic acid;Fructooligosaccharides; Iodine;Iron;Magnesium;	....	....

Table 2. Illustration of supplemental dataset for CM domain where CMDECOD is longer than 200 characters.

USUBJID	RDOMAIN	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL
AB-21-01	CM	CMSEQ	1	CMDECOD1	Standardized Medication Name 1	Manganese;Nicotinic acid;Pantothenic acid;Phosphorus;Phytomenadione; Potassium;Proteins nos;Pyridoxine; Retinol;Riboflavin;Selenium;Sodium; Thiamine;Vitamin e nos;Zinc

IMPACT OF DISREGARDING AVAILABLE GUIDANCE

When preparers of SDTM datasets are unaware of these additional guidance documents beyond the SDTM Implementation Guide, or they are aware but choose not to follow the guidance, there are actual real impacts in disregarding each type of guidance.

IMPACT OF DISREGARDING REGULATORY GUIDANCE

Disregarding the expectations of the regulatory agency, whether it be information in the Study Data Technical Conformance Guide, Technical Specifications documents, FAQs on Electronic Study Data Submission, etc., is a risky proposition. Since the agency is the authority that will be reviewing the submitted application, all attempts should be made to provide the agency with the data in the requested/expected format. Not doing so could potentially result in an information request from the agency, or worse.

Furthermore, if contradictions exist between regulatory guidance and CDISC, Phuse, or other guidance, the strong recommendation is to follow regulatory guidance, or at least to discuss the contradiction with the regulatory agency to determine how they expect it to be handled.

IMPACT OF DISREGARDING SDTMIG ERRATA AND ERRORS THAT AFFECT CONFORMANCE

Preparers of SDTM datasets that are unaware of the SDTMIG Errata and Errors that Affect Conformance risk introducing errors into the SDTM datasets and being out of compliance with CDISC standards. A possible impact of introducing these errors into SDTM datasets would be that validation issues would not be corrected, but instead incorrectly explained as false positive or validation engine error.

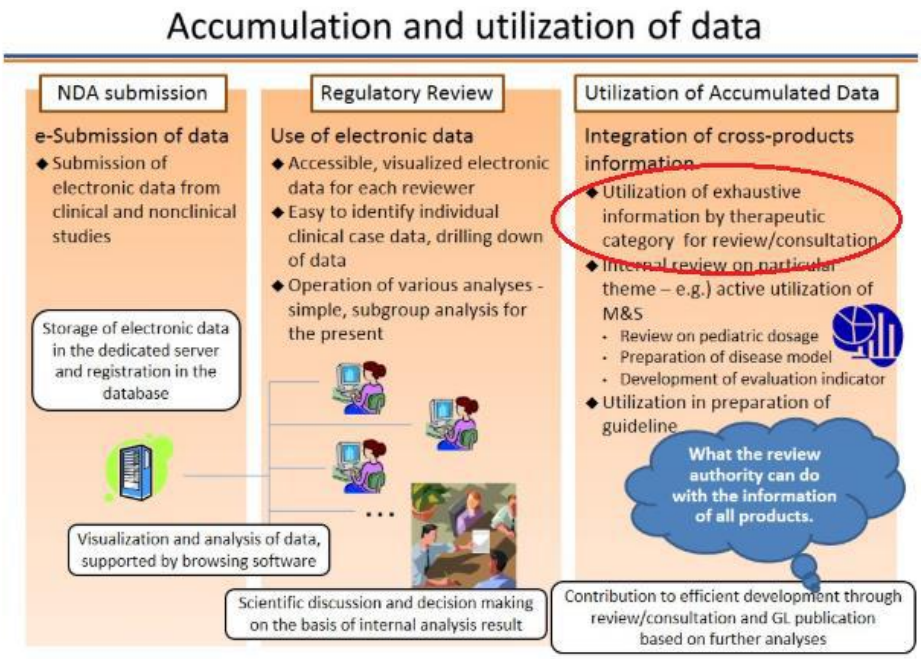
IMPACT OF DISREGARDING OTHER GUIDANCE

It is likely that a preparer of SDTM data might not be aware of the existence of other guidance such as the Uppsala Monitoring Centre's "How to use WHODrug for compliance with CM domain in the CDISC SDTM standard". If this document is not referenced, the preparer would likely choose their own conventions for how to map some of this data to SDTM, and this would result in the same type of data mapped inconsistently across the industry. A more pressing impact, however, would be that these conventions are sometimes used in validation rule algorithms, such as the convention for truncating dictionary values longer than 200 characters.

IMPACT OF DISREGARDING CDISC TAUGS AND PHUSE GUIDANCE

When TAUGs, and guidance for specific types of data like adjudication data (from Phuse), are not referenced when preparing SDTM data, the result is that the same exact types of data are mapped inconsistently across the industry.

An important impact of this is that it would potentially affect the regulatory agencies' ability to utilize the accumulated data that has been received to do cross-product analysis, as is specifically mentioned on the PMDA's New Drug Review with Electronic Data webpage:

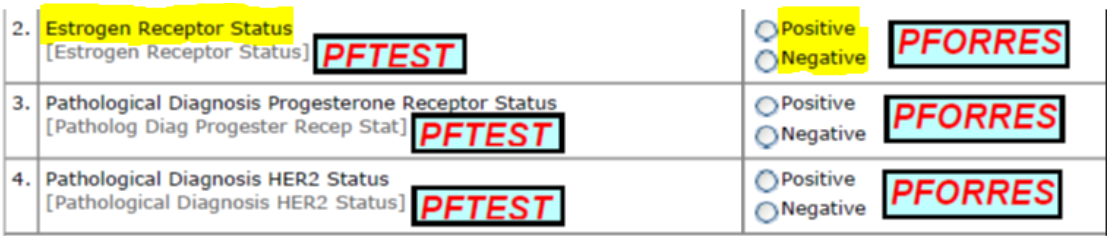


Below is an example of how the exact same data is mapped inconsistently across three sponsors, taken from actual studies. These three studies have an indication of breast cancer, for which Estrogen Receptor Status is typically collected. The Breast Cancer Therapeutic Area User Guide v1.0 does contain an example of how this data should be mapped to SDTM.

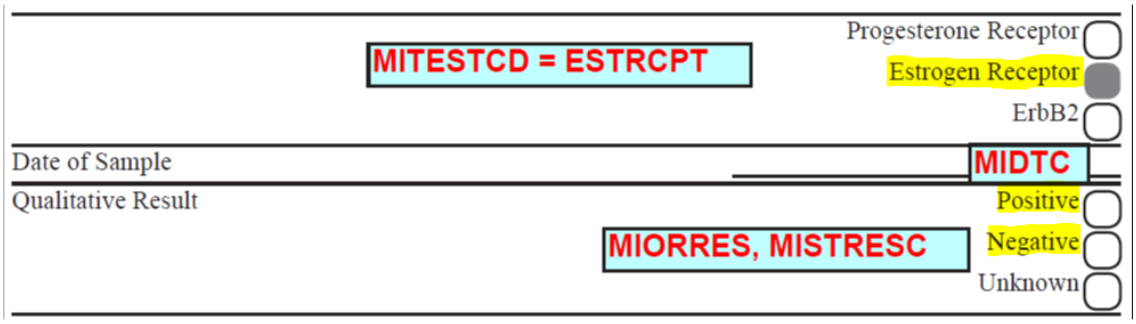
Study 1. This study has the data mapped to the Laboratory Test Results (LB) domain.

Type of receptor	LBSCAT	ER
Test method	LBMETHOD	IHC
		FISH
		CISH
		ISET
		VERIDEX
		NISH
		SISH
		Not Done
Receptor result	LBORRES	Negative
		Positive

Study 2. This study has the data mapped to the Pharmacogenomics/Genetics Findings (PF) domain.



Study 3. This study has the data mapped to the Microscopic Findings (MI) domain. This is the approach that matches the example in the Breast Cancer TAUG for Estrogen Receptor Status data.



If the same data common to certain therapeutic areas is mapped inconsistently across studies and submitted to an agency, it makes this cross-product analysis much less useful. Instead, this cross-product analysis would only be useful for data common to all trials (those mapped in implementation guides instead of therapeutic area user guides), and not other important data commonly collected for that therapeutic area.

CONCLUSION

It has been demonstrated in this paper that relying solely on the SDTM Implementation Guide is not sufficient when preparing SDTM datasets. It can be a confusing process to know which guidance documents exist, where to find them, and when they must be referenced, but they are critical pieces of the puzzle to ensure that a study's SDTM data are conformant with CDISC standards, compliant with regulatory expectations, and harmonized with similar data across the industry.

REFERENCES

Study Data Tabulation Model Implementation Guide: Human Clinical Trials. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team. Version 3.3. November 2018

Study Data Tabulation Model Implementation Guide: Associated Persons. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team. Version 1.0. December 2013

Study Data Tabulation Model Implementation Guide: Medical Devices. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team. Version 1.1. December 2018

Study Data Tabulation Model Implementation Guide: Metadata Submission Guidelines. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team. Version 2.0. March 2021

PHUSE, Best Practices for Submission of Event Adjudication Data. October 2019. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/Best+Practices+for+Submission+of+Event+Adjudication+Data.pdf

PHUSE, Define-XML v2.0 Completion Guidelines. 2019. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/Define-XML+Version+2.0+Completion+Guidelines.pdf

PHUSE, Best Practices for Documenting Dataset Metadata: Define-XML Versus Reviewer's Guide. 2019. https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/Best+Practices+for+Documenting+Dataset+Metadata-Define-XML+Versus+Reviewers+Guide.pdf

FDA. Study Data Technical Conformance Guide. Version 5.6. December, 2023. https://www.fda.gov/media/153632/download

PMDA. Technical Conformance Guide on Electronic Study Data Submissions. April, 2022. https://www.pmda.go.jp/files/000247157.pdf

PMDA. FAQs on Electronic Study Data Submission. October, 2023. https://www.pmda.go.jp/files/000247170.pdf

CONTACT INFORMATION

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

Michael Beers
Pinnacle 21 by Certara
michael.beers@certara.com

Kristin Kelly
Pinnacle 21 by Certara
kristin.kelly@certara.com