'TUFF' LOVE: SDTM Pain Points Explained!

Kristin Kelly, ICON
Michael Beers, Pinnacle 21

ABSTRACT

Though the industry is very familiar with preparing SDTM data, there are still many nuances that one can't always be sure! Trial Design, Relative Timing variables, PK, Exposure, validation rules…and the list goes on. In this paper, hot takes and topics that are crucial to 'getting it right' for submission will be discussed.

INTRODUCTION

As of December 17, 2023, it will have been seven years since the requirement to submit data to the FDA in a CDISC-structured format was implemented. For the submission of data that is collected during a clinical trial, the Study Data Tabulation Model (SDTM) is the CDISC foundational standard that must be followed. Over these last years, sponsors and CROs have been diligent in ensuring that staff was properly prepared to implement the standard per guidance in the SDTM Implementation Guide (SDTMIG) and the FDA Technical Conformance Guide (TCG). Though both resources are robust in providing this guidance, there will always be some instances where the path is not initially clear.

This paper will discuss some of the issues that come up often, how implementers have handled them, and the approach that perhaps should have been taken instead with SDTM data. Our journey will begin with Trial Design and proceed all the way through to validation and preparation for submission. And like any journey, we must always plan for inclement weather and bumpy roads ahead.

TRIAL DESIGN DOMAINS – STOP #1

The SDTM Trial Design domains represent the plan for what will be done to subjects, and what data will be collected about them in the course of a clinical trial, to address the trial's objectives. The Trial Design domains consist of the following: Trial Arms (TA), Trial Disease Assessments (TD), Trial Elements (TE), Trial Inclusion/Exclusion (TI), Trial Milestones (TM), Trial Summary (TS), and Trial Visits (TV). All of these domains are required for submission with the exception of TD and TM, which are used as needed depending on the study design.

According to SDTMIG v3.4, the standard Trial Design Datasets will allow reviewers to:

> • clearly and quickly grasp the design of a clinical trial

> • compare the designs of different trials

> • search a data warehouse for clinical trials with certain features

> • compare planned and actual treatments and visits for subjects in a clinical trial.

Typically, these datasets are created before the study starts because some are needed in order to derive other subject-level datasets such as Subject Elements (SE) and Subject Visits (SV). Others, such as TS, contain both planned and actual aspects of the trial that need to be updated during and at the end of the study.

TRIAL ARMS, TRIAL ELEMENTS, AND SUBJECT ELEMENTS

Constructing comprehensive TA and TE domains based on the protocol is instrumental to being able to derive SE at the subject-level to show the trial elements the subjects actually passed through. In the TA dataset, the planned treatment arms for the study and the order of elements (TAETORD, ETCD, and ELEMENT) that a subject must progress through for that particular arm are defined. EPOCH is also defined in TA to be used to assign EPOCH in subject-level general observation class domains. In the TE

dataset, the Elements for the study are listed in ETCD/ELEMENT and the variables, TESTRL and TEENRL, contain the rules for when each Element is to begin and end.

Where Things Can Go Wrong

If care is not taken to create the TE dataset, it can be nearly impossible to derive SE if the TESTRL does not point to a specific date that is collected in the data.

Consider the following example of a poorly constructed TE dataset:

STUDYID	DOMAIN	ETCD	ELEMENT	TESTRL	TEEENRL
KLM-2005	TE	SCREEN	Screening	Informed Consent	Date of Randomization
KLM-2005	TE	PBOT	Placebo Dose Titration	First dose of Treatment Epoch	Four weeks after start of Element
KLM-2005	TE	PBOM	Placebo Dose Maintenance	Four weeks after start of Treatment Epoch	Twelve weeks after start of element
KLM-2005	TE	DRUGAT	Drug A Titration	First dose of Treatment Epoch	Four weeks after start of Element
KLM-2005	TE	DRUGAM	Drug A Maintenance	Four weeks after start of Treatment Epoch	Twelve weeks after start of element
KLM-2005	TE	TAPER	Taper	Sixteen weeks after start of Treatment Epoch	One week after start of element
KLM-2005	TE	FUP	Follow-up	Last visit date before follow-up visit	Two weeks after start of element

A programmer tasked with creating an SE dataset using the rules defined in TESTRL above will have a difficult task ahead. For example, on the third row, TESTRL = 'Four weeks after start of Treatment Epoch', does not point to a specific date in the collected data. It implies that the SESTDTC for the 'Placebo Dose Maintenance' Element be calculated which should NOT be done in SDTM. Also, per the SDTMIG, by definition, there are no gaps between Elements so the end of one element is always the start of the next. In this TE dataset, some of the rules do not follow this concept. For the 'Screening' Element, TEERNL = 'Date of Randomization' and the TESTRL for the 'Placebo Dose Titration' Element is 'First Dose of Treatment Epoch'. These may not be the same date so there appears to be a gap between Elements.

How to Get Back on Track

In order to ensure that an accurate SE domain for a particular trial is created, the trial design author will need to take into account where the specific dates would be found in the subject-level SDTM domains when the TESTRLs and TEENRLs are defined in TE. When a TESTRL for an Element is defined, it is important to point to a specific date that is planned to be collected since SESTDTC is the date that subjects actually entered an Element and thus, needs to be present in the data. This is often a challenge when the trial design domains are being created prospectively from the protocol before any data is collected for the study and any SDTM domains exist. If it is found during the development of the SE domain that any of the rules in TE do not point to a collected date for the majority of subjects, the rule may need to be revised. For example, if the informed consent date is not collected on the CRF, the TESTRL for ELEMENT = 'Screening' should not be 'Informed consent obtained' since this particular date will not be available in the data for any subject. A more appropriate TESTRL might be 'Earliest date of screening assessments'. And because the start of one element is always the end of the previous, the rule

for TEENRL is less important than TESTRL because the date referenced in TESTRL can be applied to the SEENDTC for the previous element.

In our example, a more appropriate TE dataset might be the following:

STUDYID	DOMAIN	ETCD	ELEMENT	TESTRL	TEEENRL
KLM-2005	TE	SCREEN	Screening	Informed Consent	Within 28 days after start of Element
KLM-2005	TE	PBOT	Placebo Dose Titration	First Dose of Placebo	Four weeks after start of Element
KLM-2005	TE	PBOM	Placebo Dose Maintenance	First Dose of Placebo at Maintenance Visit Week 4	Twelve weeks after start of element
KLM-2005	TE	DRUGAT	Drug A Titration	First Dose of Drug A	Four weeks after start of Element
KLM-2005	TE	DRUGAM	Drug A Maintenance	First Dose of Drug A at Maintenance Visit Week 4	Twelve weeks after start of element
KLM-2005	TE	TAPER	Taper	First Dose of Treatment at Taper Visit Week 16	One week after start of element
KLM-2005	TE	FUP	Follow-up	Follow-up Visit	Two weeks after start of element

Each TESTRL now points to a specific collected date that can be found in the data making it much easier to create the SE dataset for this study.

Below is an example of how the SE dataset may look for this study. USUBJID = KLM-2005-003 on the first row was a screen failure so there is only one record for this subject. USUBJID = KLM-2005-004 was treated and completed the study successfully. Please note that the SEENDTC for an element is always the same as the SESTDTC for the next ELEMENT except for the last record.

STUDYID	DOMAIN	USUBJID	SESEQ	ETCD	ELEMENT	SESTDTC	SEENDTC
KLM-2005	SE	KLM-2005-003	1	SCREEN	Screening	2021-08-03	2021-09-03
KLM-2005	SE	KLM-2005-004	1	SCREEN	Screening	2021-08-15	2021-09-09
KLM-2005	SE	KLM-2005-004	2	DRUGAT	Drug A Titration	2021-09-09	2021-10-12
KLM-2005	SE	KLM-2005-004	3	DRUGAM	Drug A Maintenance	2021-10-12	2021-12-13
KLM-2005	SE	KLM-2005-004	4	TAPER	Taper	2021-12-13	2021-12-20
KLM-2005	SE	KLM-2005-004	5	FUP	Follow-up	2021-12-20	2022-01-15

TRIAL SUMMARY

The Trial Summary (TS) domain is required for all SDTM and SEND submissions to the FDA even for studies that started prior to December 17, 2016. Industry is very familiar with this requirement. One thing

they may not be as informed about is the list of TS parameters in the FDA Technical Conformance Guide (TCG) and how to populate certain parameters such as PCLAS.

TS Parameter Requirements

Beginning in SDTMIG v3.1.3, a minimum set of parameters to be present in TS were listed in Appendix C1. This included guidance regarding whether the parameter should be added, format of the values in TSVAL and applicable CDISC Controlled Terminology (CT). Now the FDA TCG contains similar appendices that list TS parameters for both SDTM and SEND. For SDTM, some of those listed do not appear in Appendix C1 but all can be found in CDISC CT. Any parameter that is flagged as 'FDA Desired = Y' should be present in the Trial Summary domain.

Pinnacle 21 validation rules check that these FDA parameters are included. Though many sponsors have been adding these 'FDA Desired' parameters to TS, there are still some that haven't. Explanations for these rules are being added to the Clinical Study Data Reviewer's Guide (cSDRG) stating that the parameter is not listed in the SDTMIG thus, they are optional to include. These TS parameters in the FDA TCG are NOT optional! They should be added to TS so that the rules no longer fire and they should NOT be explained in the cSDRG. Even if the specific parameter isn't relevant for a study, TS is designed to handle those instances where TSVAL will be null by populating the TSVALNF variable instead.

Further, Appendix C1 has now been removed beginning with SDTMIG v3.4 as these are not CDISC requirements but regulatory requirements instead. Bottom Line: INCLUDE THEM!

Pharmacologic Class (TSPARMCD = PCLAS)

The TS parameter, PCLAS (Pharmacologic Class), is a 'Conditionally Required' parameter when STYPE = 'INTERVENTIONAL' and TSVAL for INTTYPE is of a value for which pharmacological class is applicable, e.g. 'DRUG', 'BIOLOGIC'.

Pharmacological class is a complex concept that consists of three components of a substance: mechanism of action (MOA), physiologic effect (PE), and chemical structure (CS). Per the FDA TCG, TSVAL should be populated with the Established Pharmacological Class (EPC) term from the Medications Reference Terminology (MED-RT) dictionary based on the unique ingredient (active moiety) of the study treatment. The EPC term is the MOA, PE, or CS term that is considered 'the most scientifically valid and clinically meaningful' and is defined by the FDA. It is also part of an NDA's Structured Product Labeling (SPL).

Sometimes PCLAS can be difficult to populate if only the MED-RT is referenced. There is a link that is needed between the active moiety and the EPC term/code to assign the correct values in TSVAL and TSVALCD that is not included in the MED-RT. Simply searching on the NCI browser will yield a term/code from MED-RT but there is no way to determine what category the term is (EPC, MOA, CS, or PE). There are also other categories for a drug in MED-RT that are not applicable to the pharmacologic class.

The FDA maintains a list of active moieties and their corresponding FDA text phrase that matches the text of the EPC term from the MED-RT called 'FDA EPC Text Phrases for Highlights Indications and Usage heading': https://www.fda.gov/media/144963/download. In order to use this document, search for the active moiety and if an EPC term is defined, it will be listed in the next column. Then search MED-RT with this text phrase to obtain the corresponding code. In TS, TSVAL is populated with the EPC term and the code is populated in TSVALCD.

In the TS snippet below, the active moiety for the drug is populated in TSVAL for the TRT parameter and the pharmacologic class is populated in TSVAL for PCLAS.

TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVALNF	TSVALCD	TSVCDREF
1		PCLAS	Pharmacologic Class	Antimalarial		N0000175482	MED–RT
1		TRT	Investigational Therapy or Treatment	QUININE		A7V27PHC7A	UNII

For active moieties for which the EPC term has not yet been assigned by FDA, then the sponsor should discuss the appropriate MOA, PE, or CS term to assign with the review division. For unapproved investigational active moieties where the pharmacologic class is unknown, the TSVAL/TSVALCD cannot be populated. In this case, TSVALNF should be set to 'NAV' (Not Available) from the ISO 21090 standard. Once the class has been added to the MED-RT, the TS domain should be updated to include the EPC and the corresponding MED-RT concept code.

TSSEQ	TSGRPID	TSPARMCD	TSPARM	TSVAL	TSVALNF	TSVALCD	TSVCDREF
1		PCLAS	Pharmacologic Class		NAV		
1	DRUG A	TRT	Investigational Therapy or Treatment	DRUG A		9B87ACB9B	UNII

PHARMACOKINETICS DOMAINS (PC/PP) – STOP #2

Pharmacokinetics (PK) is the study of the time course of drug absorption, distribution, metabolism, and excretion (ADME) in the body. Most clinical trials, specifically the early-phase studies, include PK assessments. The SDTMIG contains two domains to hold this data:

- Pharmacokinetics Concentrations (PC) - A findings domain that contains concentrations of drugs or metabolites in fluids or tissues as a function of time.

- Pharmacokinetics Parameters (PP) - A findings domain that contains pharmacokinetic parameters derived from pharmacokinetic concentration-time (PC) data.

During the study, blood and/or urine samples are taken from subjects at specific, planned timepoints defined in the protocol. The information regarding whether or not a sample was taken and the date/times are typically collected on the CRF. The actual concentrations of drug in the samples is provided as a separate file that must be merged with the CRF data. This is the data that gets mapped to the PC domain. The concentration data from samples within a planned collection period for a subject are plotted into a time-concentration curve. From this time-concentration curve, analysis is performed using a tool such as WinNonLin which produces results of the analysis of the PK parameters (data about the concentration curve), typically defined in the protocol or the SAP.

PK DATA ABOUT SAMPLES

Historically, data about samples not taken/collected as well as other specimen properties (e.g. Volume) have been mapped to the PC domain. If a sample was not taken and this information was collected, there will be no corresponding concentration record in the raw data and PCSTAT = NOT DONE in PC. With the publication of the Pharmacogenomics IG (PGx-IG) v1.0 (now deprecated), several biospecimen domains were introduced. Some of these domains were added to SDTMIG v3.4. Two of these domains,

Biospecimen Events (BE) and Biospecimen Findings (BS), provide a mechanism for when data is collected about a specimen beyond specimen properties or whether a sample was collected or not.

- Biospecimen Events (BE) - An events domain that documents actions taken that affect or may affect a specimen (e.g., specimen collection, freezing and thawing, aliquoting, transportation).

- Biospecimen Findings (BS) - A findings domain that contains data related to biospecimen characteristics.

In oncology studies where tissue samples are obtained, further details about shipping, freezing, and archiving of samples are commonly collected. Also, samples for genomics testing are typically whittled down to smaller pieces or aliquoted. These are both good use cases for mapping this data to BE and/or BS, if needed.

Some sponsors have started to use these domains for PK data as well. This is not necessary! But why? There's a few reasons:

1. Data about the sample itself is not typically collected for PK samples.

2. Mapping –STAT = NOT DONE records to BS instead of PC makes it difficult to determine which planned timepoints are missing from a given concentration curve.

3. In regards to specimen properties, such as volume, Assumption #1 in the PC section states the following, 'This domain can be used to represent specimen properties (e.g., volume, pH) in addition to drug and metabolite concentration measurements.'

Bottom Line: For PK sampling data, sponsors should continue using the PC domain rather than creating separate BS or BE domains.

PK PARAMETER DATA

Within CDISC, it has been long argued that the PP domain is derived data that then has to be put back into SDTM so that ADPP can be created in ADaM. Typically, ADPP is just a copy of SDTM PP when the real source of data for PK parameters is in PC. Because of this, more and more sponsors have decided against including a PP dataset in SDTM. Though a validation rule will fire if PC is provided but PP isn't, this is easily explained in the cSDRG that the PK parameter data resides in ADaM. This removes the burden of having to map the same data twice and maintain it in two different places. The CDISC ADaM team is currently working on a solution that will replace the need for PP in SDTM at which time, the PP domain will be deprecated and removed from the SDTMIG. This may take some time but until then, whether or not PP is submitted is a sponsor decision and if not part of the SDTM data package, it is typically not an issue for submission.

RELATIVE TIMING VARIABLES – STOP #3

Now onto to everyone's favorite topic: Relative Timing Variables! In SDTM v1.1, implementers were limited to only two relative timing variables, --STRF and –ENRF. These variables are bound by the study reference period which is the entire span of time between RFSTDTC and RFENDTC in DM. Because these variables are relative to a span of time, only a limited set of values from the STENRF codelist can be used to populate them. For –STRF, these are BEFORE, DURING, AFTER, and UNKNOWN. The same are allowed in –ENRF with the additional value of DURING/AFTER. These two variables are often misused in that, the concept of timing relative to the study reference period is not considered. Another issue is that these variables are erroneously used for screen failures because RFSTDTC and RFENDTC are typically null for these subjects and would not have dates in DM to which to compare.

Due to these limitations, the paired variables, --STRTPT/--STTPT and –ENRTPT/--ENTPT were added in SDTM v1.2. These variables are used for timing relative to a point in time rather than the study reference period. Because the anchor is a point in time, values of COINCIDENT and ONGOING are used to populate –STRTPT and --ENRTPT instead of DURING and DURING/AFTER from the STENRF codelist. –STTPT and –ENTPT is populated with the point in time, such as 'SCREENING', 'VISIT 12', etc. Dates in ISO 8601 format can also be populated in these variables.

It is important to note that any of the relative timing variables should only be populated in lieu of a date and only if the data to support it is explicitly collected. These variables should never be derived. An example of when this data is collected is an 'Ongoing' or 'Continuing' checkbox on an Adverse Event CRF that would be checked when the AE has not yet resolved. Because these variables should not be derived, it means that when a date is collected, even a partial date, the relative timing variables should be null. Date variables and relative timing variables should not be populated at the same time. It also means that when no date nor any data to support populating relative timing variables was collected, they should be null.

WHEN THINGS GO OFF THE ROAD…INTO A DITCH

The following example is a cautionary tale that does not only apply to relative timing variables. A third party vendor performed an SDTM review for a large CRO. One of the review comments was about the following P21 validation rule:

Pinnacle 21 ID	Publisher ID	Message	Description
SD0021	FDAB038	Missing End Time-Point value	One End Time-Point variable is expected to be populated when an event or an intervention occurred. (E.g., End Date/Time of Event or Intervention (--ENDTC), or End Relative to Reference Period (--ENRF), or End Relative to Reference Period (--ENRTPT) should not be missing, or Occurrence (--OCCUR) = 'N' or Completion Status (--STAT) != '') or Duration (--DUR) != '').

SD0021 checks that at least an end date (--ENDTC) or a relative timing variable (--ENRF or –ENRTPT) is populated for a record in an Events or Interventions domain. The rule had fired for a handful of records in the AE domain where AEENDTC, AEENRF, and AEENRTPT were all null. The programming team that prepared the data package had properly explained this issue in the cSDRG.

Rule ID	Rule Message	Explanation
SD0021	Missing End Time-Point value	For these records, AEENDTC was not collected. The 'Ongoing' box on the CRF was also not checked. End timepoint is not able to be populated.

The vendor performing the review commented that when no date was collected then –ENRTPT must be populated with ONGOING so that SD0021 will no longer fire. In fact, several of the review comments directed to impute/derive values in the SDTM data to avoid validation issues. This is absolutely the wrong thing to do! The advice the vendor provided is vastly different from fixing a programming mistake or adding required TS parameters to resolve validation issues. It is directing to change or add data that was not collected. Very little derivation and no imputations whatsoever should be done in SDTM. Bottom Line: In SDTM, MISSING DATA IS MISSING DATA!! Validation is a useful tool that gives sponsors the opportunity to explain outages in the data in the cSDRG. The goal should never be to game the validation process.

GUIDANCE DOCUMENTS TO REFERENCE – STOP #4

MYTH: FOLLOWING GUIDANCE IN THE SDTM IMPLEMENTATION GUIDE IS SUFFICIENT TO PREPARE SDTM DATA FOR SUBMISSION

Preparers of SDTM datasets for submission may think that if they reference the CDISC SDTM Implementation Guide, they are in the clear. However, that is not the case…

FACT: IT IS MORE COMPLICATED THAN THAT

There are requirements and guidance from other agencies and organizations of which a preparer needs to be aware. Here are some (but not necessarily all) examples:

- Regulatory Agency Study Data Technical Conformance Guides and FAQs

    o The FDA and PMDA each publish a Study Data Technical Conformance Guide, which describes the agency's preferences, recommendations, guidance, expectations, or even requirements for how they would like (or require) study data to be prepared and submitted.

    o 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.

    o 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.

- Regulatory Agency Therapeutic Area Specifications

    o Some regulatory agencies, such as the FDA, publish therapeutic area specifications that describe their expectations for how certain specific data is created and submitted. Examples of these are the FDA's Vaccines Technical Specifications Guidance, QT Studies Technical Specifications Document, HIV Technical Specifications Guidance, etc. Some of these are more relevant to SDTM mapping, while others are more focused on what to represent in the ADaM datasets.

- CDISC Therapeutic Area User Guides

    o While CDISC standards cover much of the data common to clinical trials, there are some gaps, most due to data specific to certain therapeutic areas, but uncommon in most other trials. Therapeutic Area User Guides (TAUGs) are created to fill some of these gaps.

- SDTM Implementation Guide Errata

    o Occasionally mistakes are included in CDISC standards and implementation guides. To correct these mistakes, CDISC sometimes publishes errata.

    o These can be found on the CDISC website. When you click on a version of the SDTM Implementation Guide, one of the tabs to click is 'Errors'. Here CDISC provides an Errata, as well as Errors that Affect Conformance.

    o This is important information to be aware of when preparing CDISC datasets, however it seems much of the industry is unaware of the existence of these corrections.

- Guidance for domain-specific mapping from other organizations

    o 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. 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.

    o The PHUSE Best Practices for Submission of Event Adjudication Data is another example of guidance provided by a different organization. Since CDISC has not provided specific guidance on mapping adjudication data, a preparer would need to be aware of this guidance in order to map adjudication data in a more standardized way than deciding their own choice of mapping.

- Special Circumstances
  - During the COVID-19 pandemic, missed visits, or remote visits replacing in-person visits became more common. In order to provide some guidance to the industry in how to handle pandemic-related complications such as these, the FDA, PMDA, and CDISC released a number of guidance documents, such as:
    - The FDA's 'Conduct of Clinical Trials of Medical Products During the COVID-19 Public Health Emergency
    - The PMDA's 'Principles for the Evaluation of Vaccines Against the Novel Coronavirus SARS-CoV-2'
    - CDISC's 'Guidance for Ongoing Studies Disrupted by COVID-19' and 'Therapeutic Area User Guide for COVID-19'

As should be obvious at this point, the amount of guidance that preparers of CDISC datasets need to be aware of is somewhat extensive, and therefore it would be insufficient to only be aware of one, or just a few, of these guidance documents.

USE OF CUSTOM DOMAINS – STOP #5

Now that six versions of SDTM Implementation Guides have been released over a period of more than 15 years, and more than 40 Therapeutic Area User Guides have been published, the industry has standard domains defined for much of the data typically found in clinical trials, *but not all types of data*.

Custom Domains … the Gaps in Standardized Data

There are sometimes less common types of data collected in clinical trials that have no standard domain, and a custom domain is needed. As is apparent in the name 'custom', these represent gaps in standardized data, and should be used only in cases where a standard domain really does not exist.

While a preparer deciding to use a custom domain may ultimately be the right choice, there are some steps you should follow before rushing to this decision.

- Confirm that none of the existing published domains will fit the need…
  - Both FDA and PMDA specify this point in their Study Data Technical Conformance Guides:
    - The FDA's Technical Conformance Guide, October 2022, section 4.1.1.3, states: Prior to creating a custom domain, sponsors should confirm that the data do not fit into an existing domain…
    - The PMDA's Technical Conformance Guide on Electronic Study Data Submissions, April 2022, section 4.1.1.2, states: Depending on the characteristics of the collected data, it may not fit into an existing domain of SDTM. In such a case, it is acceptable for the applicant to create a custom domain. To perform this, the applicant must confirm that the data does not fit into existing domains…
  - CDISC also states this, as the first point to consider before creating a custom domain:
    - The Study Data Tabulation Model Implementation Guide: Human Clinical Trials Version 3.4 (Final), section 2.6, states: 1. Confirm that none of the existing published domains will fit the need…
- CDISC then directs (in section 2.6 of the SDTMIG v3.4) the preparer to check the SDTM Draft Domains area of the CDISC wiki (https://wiki.cdisc.org/display/SDD/SDTM+Draft+Domains+Home) for proposed domains developed since the last published version of the SDTMIG. These proposed domains may be used as custom domains in a submission.

- Finally, CDISC states (in section 2.6 of the SDTMIG v3.4) that the preparer should look for an existing, relevant domain model to serve as a prototype. If no existing model seems appropriate, choose the general observation class (Interventions, Events, or Findings) that best fits the data by considering the topic of the observation…

One point of clarification is that if an existing published domain doesn't exist in the version of the SDTMIG being used for the trial, but has been introduced in a later version of the SDTMIG, the preparer should consider up-versioning to the newer version of the SDTMIG where this domain has been introduced, but even if up-versioning does not happen, this domain introduced in a later version of SDTMIG should still be used instead of creating a custom domain.

The impacts of using a custom domain, instead of an existing published domain are:

- Breakdown of automated processes. Automation, which is one of the important ways to reduce review times and make drugs available to patients as fast as possible, relies on high data quality and high data standardization. Manual review is inefficient and is usually insufficient in evaluating and enforcing data quality. Therefore, the use of custom domains, which is a gap in standardized data, should be kept to a minimum.

- Validation is too general on custom domains. Domain-specific checks are not run and domain-specific codelists are not checked properly. An example of this is if a custom findings domain is created to store some lab data (that the preparer wants to keep separate from the typical safety labs):
    - In SDTMIG v3.3, there are 155 validation rules run on the LB domain, but there are only 144 validation rules run on a custom findings domain. Therefore, not all lab-related validation rules would be run properly on this custom domain.
    - Variables such as –TESTCD and –TEST would not be checked against the LBTESTCD and LBTEST codelists in that custom domain, even though the controlled terminology should have been used.

- Use of custom domains, which is not standardized data, may cause problems if regulatory agencies attempt cross-product analysis using their accumulated data.

STUDY DATA VALIDATION – STOP #6

Validation of study data is a critical step in preparing for a submission. Proper validation processes, specific to each regulatory agency, need to be followed to avoid delays to the application.

Where Things Can Go Wrong

Some preparers of CDISC datasets seem to think that using the same validation engine to validate the data throughout the course of the trial is an acceptable approach. However, this is not the correct way to handle study validation. Each regulatory agency has its own requirements, and the recommendations for both FDA and PMDA are listed as follows.

The FDA routinely publishes a set of FDA Validator Rules, which represent the latest understanding of what best supports regulatory review. These are the rules that are run whenever study data is submitted to the agency, therefore preparers must run these rules as well, to see the same set of validation issues that the agency will see at the time of submission. Therefore, for FDA submissions, preparers of CDISC data should always use the latest available FDA validation engine, for the following reasons:

- Seeing any differences in validation issues from a new validation engine is only good. It is best to view compliance as a moving target, requiring continual adjustments.

- Surfacing new validation issues prevents sponsors from being surprised by FDA reviewers who have the most up-to-date version of the Validation Rules. If the team is facing time or budget constraints, an issue explanation can always be added to such newly flagged Issues.

- Reducing false positives, which is continually done in new validation engines, allows teams to focus their efforts on actual issues only. This is a key time-saving benefit for users.

For PMDA, the acceptable validation engines, which correspond to versions of Study Data Validation Rules, are listed on the PMDA website (https://www.pmda.go.jp/english/review-services/reviews/0002.html). The Study Data Validation Rules, and applicable validation engines, are determined by application date. Preparers of CDISC data for submission to PMDA should be sure to use an acceptable validation engine. The impact of not using an acceptable validation engine is mentioned in the PMDA's FAQs on Electronic Study Data Submission, Q1-16, which states: *"…if violations of the validation rule corresponding "Reject" or "Error" with no explanation are newly detected by validation at the PMDA, the PMDA will require action such as correction of data.*" and in Q1-18, which states "*If the validation detects a violation of the validation rule corresponding to "Reject", electronic study data that require action and the validation rule will be notified to applicants in writing if the application has not been made yet or as an inquiry if the application has already been made. If the validation detects a violation of the validation rule corresponding to "Error" with no explanation, electronic study data that require action and the validation rule will be notified to applicants as an inquiry after the application is made.*" Therefore, if a preparer sees different validation issues than the PMDA will see, the result may be a delay in processing the application due to the need to correct data or documents.

To summarize, for FDA submissions, always use the latest FDA validation engine, even if an older validation engine had been used previously for the trial's study data, to be sure that new issues are seen and explained (if the issues can't be fixed) for all validation issues that the agency will see. For PMDA, be sure to use an acceptable validation engine per PMDA guidance and follow all applicable processes regarding validation of study data and documents.

CONCLUSION

And with that, our journey through SDTM Land has come to end. We've traveled through Trial Design all the way through to submission with many stops along the way. Though obstacles were encountered, we now have some tools to rechart the course to get back on the road. If some of the course corrections discussed here can be applied, the journey will be much easier moving forward. The final destination is always a successful submission, an accelerated review and eventual regulatory approval in an overall effort to get drugs (and vaccines) to patients faster.

REFERENCES

Study Data Tabulation Model Implementation Guide: Human Clinical Trials. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team. Version 3.4. November 2021

Kelly, Kristin. 2015. 'SDTM TE, TA, and SE domains: Demystifying the Development of SE'. PharmaSUG 2015 https://www.pharmasug.org/proceedings/2015/DS/PharmaSUG-2015-DS02.pdf

Kelly, Kristin. 2016. 'SDTM Trial Summary Domain: Putting Together the TS Puzzle'. PharmaSUG 2016, CDISC Interchange 2016 https://www.pharmasug.org/proceedings/2016/DS/PharmaSUG-2016-DS11.pdf

Salyers, Jerry and Wood, Fred. 2013. 'Considerations in the Use of Timing Variables in Submitting SDTM-Compliant Datasets'. PharmaSUG 2013. https://www.pharmasug.org/proceedings/2013/DS/PharmaSUG-2013-DS14.pdf

U.S. Food & Drug Administration, Study Data Technical Conformance Guide (current version at time of access – v5.0/ October 2022), https://www.fda.gov/media/153632/download

FDA EPC Text Phrases for Highlights Indications and Usage heading. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). January 2021. https://www.fda.gov/media/144963/download

Pharmaceuticals and Medical Devices Agency (PMDA), Technical Conformance Guide on Electronic Study Data Submissions. April, 2022. https://www.pmda.go.jp/files/000247157.pdf

Pharmaceuticals and Medical Devices Agency (PMDA), FAQs on Electronic Study Data Submission. June, 2022. https://www.pmda.go.jp/files/000247170.pdf

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

Uppsala Monitoring Centre, How to use WHODrug for compliance with CM domain in the CDISC SDTM standard. 2017. https://www.who-umc.org/media/2940/how-to-use-whodrug-for-compliance-with-cm-domain-in-the-cdisc-sdtm-standard-march-2017.pdf

CONTACT INFORMATION

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

Kristin Kelly	Michael Beers
ICON	Pinnacle 21
kristin.kelly@iconplc.com	mbeers@pinnacle21.com

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