PhUSE 2012

Paper CD09

CDISC Advisory Board Validation Project: Best Practices for CDISC Validation Rules

Lauren Shinaberry, Business & Decision Life Sciences, Brussels, Belgium

ABSTRACT

In 2010 the CDISC Advisory Board (CAB) began the Validation Project. This working group was given several different tasks, including the assessment of the current state of rules used in the top 5 SDTM validation tools being used by the industry. In April 2012 this CAB task force concluded their assessment. This presentation shares the Best Practices that were identified and the conclusions that were reached with respect to the current state of SDTM validation rules, as well as an update on the continuation of the efforts begun by this team and how they fit in with other industry activities for the development of validation rules.

INTRODUCTION

The CDISC Study Data Tabulation Model (SDTM) has been in use for several years, and the SDTM Model Document and Implementation Guide has been the primary source for determining the assumptions and rules that should be followed to ensure that the resulting datasets conform to the model. It has also been the mechanism the teams who have been defining SDTM have relied on to communicate their expectations and assumptions on the correct implementation of the model. Several tools exist that take the rules and assumptions from the SDTM and Implementation Guides and apply these programmatically to datasets and identify where the metadata deviates from the model. To date, there has not been an official list of rules defined by CDISC, so existing tools rely on someone manually reading through documents and translating them into concise rules that can then be programmed. This resulted in some variability between tools and can be subject to interpretation by the person reviewing the documents.

CDSIC Advisory Board members requested a way to ensure that their datasets were conformant with SDTM and that vendor tools claiming to validate SDTM datasets were reliable and consistent with the CDISC teams definition of the model. The CAB Validation Project was started and a team was assembled to look into the current state of SDTM validation tools and make recommendations to the CDISC Board of Directors on how best to meet the needs of the SDTM implementers for validation tools.

This paper summarizes the best practice recommendations that came from this team's assessments.

PROJECT TEAM

The CAB Validation Team had broad industry representation, including representation by the FDA and five vendors that offer SDTM validation tools. The team members brought their skills in experience with SDTM implementation as well as strong technical skills. The five validation tools that were included in the project were (alphabetical by tool name):

- CheckPoint from Octagon
- Data Model Compliance Checker from Business & Decision Life Sciences
- OpenCDISC Validator
- SAS Clinical Standards Toolkit
- WebSDM from PhaseForward

METHODOLOGY

The five vendors who participated in the project provided the list of the validation rules used in their tools. The team assessed the similarities and differences between each. The focus of this comparison was to identify differences in meaning of checks, categories of checks, determination of severity of failed checks and potentially missing checks based on the SDTM and Implementation Guide.

Each rule in all five tools were reviewed, along with the text of the message to align checks that may have been described differently but were in fact identifying the same issue. Severities were categorized as Errors, Warnings or Informative Notes in some tools, while others used High, Medium and Low.

FINDINGS

The majority of checks across all five tools were identifying the same types of issues, although often with differing descriptions or logic. Each tool also included extra checks that were unique. There were some rules that the team identified from the Implementation Guide that were not present in any of the validation tools.

CHECK TYPES

The checks could be grouped into four main types: SDTM conformance checks, data quality checks, system-specific checks and FDA implementation checks.

- SDTM Conformance Checks were defined as those that could be traced back to a specific reference in a CDISC document.
- Data Quality Checks were rules that are Good Clinical Data Management practices, irrespective of whether SDTM is used, such as identifying when a start date is after an end date.
- System-specific Checks were primarily seen in the WebSDM tool as this same tool may be used for data review, such as patient profiles, and features of these require additional validity checks unique to how the system processes the underlying SDTM datasets. System-specific checks were considered outside the scope of the project and were not included in other reviews
- FDA implementation checks were defined as those that could be traced back to the CDER Common Data Issues document ((www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm248635.htm) and were not already categorized as SDTM Conformance. An example of this was the FDA request for the inclusion of EPOCH, which is not a required variable per the SDTM IG v3.1.2 that was current at the time of the project

None of the tools identified the checks in this manner, which may lead to users being unaware that not all issues identified by any given tool were necessarily deviations from SDTM. To make it clear to users which category a check belongs to, a reference to the documentation supporting the definition of that check should be provided. The additional benefit of including a reference to the SDTM or FDA documentation is that users can easily find a fully detailed description of why an issue has been flagged and where to find out how best to address it.

SEVERITIES

All tools assigned a severity to each check, although the assignments were not consistent and the definitions for how severities were assigned was also quite varied. The team recommended simplifying the severities to either Errors or Warnings rather than Low, Medium or High. Some checks are very straightforward to identify and are obvious deviations from the rule referenced by the check, such as date fields not conforming to ISO8601 format. Other checks were more complex and while a program can identify a deviation, it may be necessary for further investigation to determine whether it can be addressed and corrected.

The determination of what must be addressed and what may be acceptable to leave as is was slightly different based on the type of the check as shown in the table below. The focus of the criteria was the implication if the issue was not resolved. Those checks that were identified as relating to data quality rather than conformance were not included in the severity definitions because these types of checks should be performed as part of the data cleaning process during data collection. It was decided that the focus of the validation tools should not duplicate data cleaning that should be done at the time of collection through a CRF 21 Part 11 compliant data management system. If an SDTM validation tool is used to identify data quality issues, this should be more for helping to confirm any data anomalies are explained in the Reviewers Guide. However, if the CDER Common Data Issues document references a specific type of data quality issue then this would fall under the rules for severity related to FDA Conformance.

	SDTM Conformance	**FDA Conformance**	**Data Quality**
Error	Must be addressed to conform to SDTM, clearly violates a rule as stated in the SDTM or SDTM IG	Must be addressed or the data cannot be reviewed	Not applicable - data quality issues should be addressed through Good Clinical DAta Management Practices and are not specifically SDTM conformance
Warning	May need to be investigated to determine if a conformance criteria have	May cause questions from the reviewer if not properly explained in the Reviewers	

	not been met	Guide or other documentation provided with the submission	

CATEGORIES OF CONFORMANCE RULES

Within the SDTM Conformance type of checks, there were four main categories of checks identified.

- Structural
- Controlled Terminology
- Content

Structural rules were those checks that compared domain-level and variable-level metadata to the SDTM. This includes checks such as are domain names, domain labels, variable names, labels and types consistent with SDTM.

Controlled Terminology rules check that variables are using the correct terminology and that no new terms have been added to non-extensible codelists.

Content rules were those that typically required the comparison of two or more variables, such as checking for duplicate Sequence Numbers within a Unique Subject Identifier, or that in the Related Records domain, if the Relationship Type is present, then the Identifier Variable Value should be null.

DEFINE.XML, XPT OR SAS DATASETS?

There was considerable discussion in the team regarding what files should be used as input to the checks. SDTM data can exist in multiple formats such as SAS datasets or SAS transport files. In addition, a define.xml file may also be available. The metadata in all three of these sources must be consistent, so regardless of which is used as the primary input to the validation tool, there should also be checks that confirm that the metadata in each of these sources is consistent.

CONCLUSION

There are many options available for programmatically checking your datasets for conformance to SDTM and there are also many organizations programming their own checking tools. The Best Practices identified in the CAB Validation Project are intended to make it easier for users to understand the rationale behind each check as well as the implications of not addressing a finding.

The suggested methods for accomplishing this are:

1. Provide a reference for each check to distinguish checks for SDTM conformance from other checks, provide link for users to find additional details if that check fails for their data and use the text from the reference as the “plain language” description of the check.
2. Simplify Severity to Errors or Warnings, making it clear to users what action is needed
3. Have consistent message text that clearly describes the issue in plain language rather than pseudo-code
4. Have a well defined process for maintaining checks that is coordinated with new releases of SDTM
5. Do not duplicate checks

The CAB Validation Project has been concluded since March 2012. The work started by this group is being continued in a variety of manners. The categorizations and severities of the SDTM conformance checks have been shared within CDISC for possible inclusion in future implementations of the CDISC SHARE project. Work on the FDA conformance checks and other activities have been shared with PhUSE FDA Working Groups.

ACKNOWLEDGMENTS

The author would like to acknowledge **Trisha Simpson** (UCB) for her key role in starting and leading the CAB Validation Project. In addition to Trisha, the hard-working team on the CAB Validation Project included:

Borowiak, Ken (PPD); **Castaldo,** Jeanne (Hurley Consulting); **Casteris**, Christina (Biogen Idec); **Cassells**, Sally (Next Step Clinical Systems); **Chen**, Hong (McDougall Scientific); **Chow,** Anthony (Allergan); **De Donder,** Nick (Business&Decision Life Sciences); **Freimark**, Nate (Clinical Research); **Fulda**, John (Premier Research); **Gibson**, Bill (SAS); **Gicqueau,** Ale (Clinovo); **Hernandez**, Joyce (Merck); **Hungria**, Marcelina (Dicore Group); **Hurley,** Margaret (Hurley Consulting); **Hurst**, Don (Premier Research); **Kalfas**, Thomas (Premier Research); **Koranda**, Adrienne (Premier Research); **Kanevsky,** Max (OpenCDISC); **Labout,** Shannon (CDISC); **Lightfoot,** Gene (SAS);

Low, Geoff; Malla, Amy (FDA); **Qian**, Wei (Takeda Global); **Smith,** Steve (SAS); **Tomioka,** Sam (Sunovion); **Truette,** Kim (KCTDM); **van Manen**, Robbert (Oracle); **van Reusel**, Peter (Business&Decision Life Sciences); **Verloet,** Stefaan (J&J); **Witczak,** Bobbie (FDA); **Wilkins,** Ryan (PPD).

CONTACT INFORMATION

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

Lauren Shinaberry
Business & Decision Life Sciences
Rue Saint-Lambert 141
Brussels B-1200
Work Phone: +32
Fax:
Email: Lauren.Shinaberry@BusinessDecision.com
Web: