Preparing to Share CDISC SDTM Data: A Practical Evaluation of Potential Error Sources and Effective Review Processes

Steve Kirby, JD, MS [1] and Nancy Wang, PhD [2]
[1] ViroPharma Incorporated. [2] Celerion

OBJECTIVE

CDISC data standards can help get needed new drugs to patients by streamlining the FDA submission process and strengthening scientific collaboration in drug discovery. For that potential to be reached, CDISC standards must be accurately and efficiently applied. Our objective was to generate a comprehensive review process that addressed potential translation errors when programmatically mapping source information to CDISC SDTM supported by review documentation that would meet the needs of downstream data consumers.

METHODS

Processes associated with programmatically mapping source information to CDISC SDTM were evaluated to see what types of translation errors were possible. Those potential errors were categorized, and review methods (including documentation suitable for all downstream users) were specified by category.

RESULTS OVERVIEW

Review of programmatic mapping processes for translation errors indicated that the primary sources of error are:

- Misrepresenting The Source Data
- Not Conforming with the Applicable CDISC SDTM Standards
- Documentation that is Not Consistent with the SDTM Data

Given those primary error sources, data review needs to establish that the SDTM data are:

- Accurate: The SDTM Data are Complete and Consistent with the source Information
- Conformant: The SDTM Data Conform to the Applicable CDISC Standards
- Consistent: The define.xml and blankcrf.pdf are consistent with the SDTM data

Accurate: The SDTM Data are Complete and Consistent with the source Information

Review Methods

Primarily driven by manual review of information as collected and as mapped. That manual review can be programmatically supported in a variety of ways. More automation (and a simpler review process) is possible when source information is consistent with SDTM standards.

What to Check

Core Manual Check Categories Identified:

- 1 to 1 relations (including exact match)
- Not 1 to 1, more context is needed
- Date/time conversion
- Missing data references
- Is “Assigned” and ”Derived” content correct and consistent with any associated mapping specifications or documentation
- Completeness – Are all data relevant to submission included. “Permissive” variables are key for this category.

Tips

Printing all unique combinations of related variables as collected and as mapped (useful combinations tend to be within categories) usefully highlights potential mapping issues.

SDATE_ORIG	CMSTDAT
1 FEB 2011	2011-02-01
JAN 2012	2012-01
2010	
12 JAN 2012	2012-01-12

- Reviewers in this area do not have to be experts in SDTM if supported by content that makes the relation between SDTM data and original information easy to follow
- Harmonizing collection with CDISC controlled terminology minimizes substantive review effort
- Review can be effectively built into the mapping process and effectively linked with mapping specifications
- Validation of CDISC based report tables based on comparison with tables programmed using source data offers efficient verification of the most scientifically critical content

Process Guidelines and Review Documentation

Prove substantive validation was completed with signature document that references substantive checks completed

Conformant: The SDTM Data Conform to the Applicable CDISC Standards

Review Methods

Primary done programmatically. Good validation code easily available (some free, some for fee). Recent FDA statements indicate that they are using Open CDISC.

What to Check

Verify data using published conformance criteria met. Criteria not met should be minimized, and deviations supported. Key information needed to support use of tools:

- Exact version of SDTM used for mapping
- Version of Controlled Terminology used for mapping
- What versions are applied by check program used

Tips

Programmatic program verification tends to be complete and accurate - but be aware of limitations that need to be filled with manual review or manual documentation

- There can be a disjoint between the defined terms lists in the check program and in the specific controlled terms reference used for mapping as release schedules can vary
- Check programs naturally lag CDISC standard updates
- Automated conformance checks cast a wide net with warnings (especially related to controlled terminology).

The specific standards and verification code to be used can be effectively integrated into planning for mapping; consider whether a process is needed to support updating standards applied if planned study durations are very long.

Process Guidelines and Review Documentation

CDISC Standards Applied		
SDTM Version	1.2	
SDTM IG Version	3.1.2	
Controlled Terms Version	22JUL2011	
Conformance Code Used		
Program	Open CDISC Validator	
Program Version	1.2.1	
Program Inputs	SDTM domains and define.xml	
Conformance Findings		
Type	Message	Notes
Warning	Value for MHBODSYS not found in SOC controlled terminology codelist	MedDRA body systems as submitted are in mixed case as provided by the dictionary. The term list used for this check was in upper case.

Ensure that all validation issues are explained (and related to the specific code and standards used) in a format that can be included in submission package [typically located in Reviewer’s Guide].

Consistent: The define.xml and blankcrf.pdf are consistent with the SDTM data

Review Methods

Both programmatic checks (variable characteristics/code lists in define consistent with the data) and manual review (Is the define.xml origin consistent with the blankcrf.pdf and with the data) are needed.

What to Check

blankcrf.pdf

- Are annotations consistent with SDTM mapping
- Are annotations consistent with define.xml origin [match with “CRF” page references, consistent otherwise]

define.xml

- Is data documentation consistent with data
  - Domain, Variable and Value attributes
  - Code list values
- Is origin accurate and consistent with blankcrf.pdf
  - Origin references are consistent with SDTM mapping
  - CRF page references match blankcrf.pdf annotations
  - Accurate derivation comments support each derived variable

Tips

- Be aware of limitations in programmatic review of consistency between define.xml and SDTM data as any gaps (such as domain labels) need to be otherwise supported
- Building define.xml creation/review into the mapping process or creating define.xml based on data nicely supports consistency
- Supporting define.xml origin assignment and blankcrf.pdf annotation with local rules facilitates effective, efficient application and review

The define.xml origin should be CRF if and only if :

- There is a direct (exact or functionally equivalent), 1 to 1 match between data as collected and SDTM data.
- There is a direct (exact or functionally equivalent), 1 to 1 match between information pre-printed on the CRF and SDTM data.

Please see above for definition of “functionally equivalent”

Process Guidelines and Review Documentation

Prove consistency with

- Signature document that references manual review of source references (blankcrf.pdf annotations, define.xml origin references) completed [“Origin Audit”]
- Documented programmatic verification that define.xml references (outside of header information) are otherwise consistent with the data.

Note: Additional verification of blankcrf.pdf and define.xml are also needed to ensure that all applicable formatting standards are met. [See, e.g. MSG-SDTM].

CONCLUSIONS

Generating and applying a complete and accurate set of review steps for each category of possible translation error when mapping source information to SDTM makes it so conforming, accurate data and documentation can be consistently delivered.

Those review steps must ensure the data accurately reflects the source information, conforms with all applicable CDISC standards, and has associated documentation (define.xml and blankcrf.pdf) that is consistent with the data. Where CDISC guidance supports more than one mapping or documentation choice, local rules are needed to clarify the choice to be made.

Preparing to share CDISC SDTM data is more than just preparing accurate, conformant SDTM domains and documentation; it is being able to show a wide audience why you are certain that the SDTM data provided are accurate, conformant and consistent. Proving that a comprehensive review was completed in a way that will satisfy a broad audience of downstream users makes it so those review steps do not need to be needlessly repeated.

Thought should be given to having review documentation available in a form that will serve the needs of multiple audiences. While each specific detail of the review process is important, many end users are most interested in and best served by a concise overview of the process. One effective approach is to have the final signature document give an overview of substantive areas checked with further detail incorporated by reference.

AUTHOR CONTACT

The authors welcome any comments and questions. Please feel free to contact:
Steve Kirby, Manager Data Standards Implementation, ViroPharma: steven.kirby@viropharma.com
Nancy Wang, Associate Director, Biostatistics, Celerion: nancy.wang@celerion.com

DISCLOSURE

Authors of this presentation have the following to disclose concerning possible financial or personal relationships with commercial entities that may have a direct or indirect interest in the subject matter of this presentation:
Steve Kirby: Nothing to disclose
Nancy Wang: Nothing to disclose