SDTM vs CDASH: Why You Need Both!

Abstract

Some think CDISC's CDASH data capture standard is unnecessary. They say it's very similar to SDTM, and the few differences create confusion and extra work. CDASH *is* similar to SDTM, but they solve different problems. Used together they positively impact data capture, quality, usability, repurposing, and traceability.

We explore differences between CDASH and SDTM and why *both* standards are critical.

The Same or Different?

CDASH and SDTM are in fact very similar.

- 67% of CDASH v2.0 maps directly to SDTMIG variables, and *CDASH v2.0 includes mapping*
- 86% of CDASH maps directly with standard mappings included (e.g., dates)
- 14% are different for a reason

SDTM is optimized for tabulation, analysis dataset creation, & data submission.
CDASH is optimized for data capture, investigator site activities, & data quality.

Different requirements, different approaches, but with the same end in mind.

SDTM	Why the Difference?	CDASH
Show me the data, not lack of data	SDTM assumes that if there is no record then nothing happened. This works but only if it was checked in data capture, which requires a question and record (e.g., Were there any AEs?)	Absence of evidence is not evidence of absence: must check that missing data is missing
Machine-readable data: - ISO 8601 Dates/Times: 1 variable, YYYY-MM-DDThh:mm:ss - Duration: P1M3D	SDTM machine-readable formats for variables such as dates are good for data reusability but are not user-friendly for data capture. Sites recording data in unfamiliar formats increases risk of errors	Human-readable: - Dates/Times: 2 or more variables, DD-MMM-YYYY, HH:MM:SS - Duration: 1 month, 3 days
Variables must be in order by domain; non-standard variables are stored in different datasets (e.g., FA, SUPP--)	Domain-driven organization is critical for standard tools, but data must make sense to the site. This can mean to split domains across CRFs and CRFs across domains, and not split custom and standard variables	Data structure harmonized with SDTM but variables can be arranged to make data capture easier.
Collected relationships between data are represented in RELREC, a separate dataset	RELREC is based on collected data, but data is not captured like that. Entering line numbers in the related datasets is simpler, requiring no derivations (e.g., adding AE line # to related con med)	Links among records are explicit (e.g., this AE related to that CM), or implicit (e.g., AE severity changes going into FA) in data collection
Findings data must be in a normalized or vertical structure; answers are already in variables	Normalized structures can store new tests without changing dataset structures, but most EDC systems can't do this; also, different tests in a domain may need different controlled terms (e.g., different answers for different questions in a survey)	Findings data may be horizontal, letting each test have a different code list; SDTM CT is used for variable names & CRF prompts
Metadata centers on tabulations, e.g., variable labels and roles	SDTM labels identify tabulation data. CDASH has question texts and prompts designed to elicit clear responses on CRFs. CRF instructions convey SDTM and CDASH assumptions in a data capture context	Metadata includes capture needs, e.g., question text/prompt, CRF completion instructions

Conclusions

To use SDTM instead of CDASH for data capture, take out derived variables, records and datasets; add in data quality indicator variables; put all custom and FA variables into parent datasets; reformat variables that are not user-friendly; reword variable labels to questions; and restructure vertical data to horizontal.

This effectively produces CDASH. Except each organization will do it differently, resulting in reduced data quality and traceability

Whether Regulatory Affairs assembling a submission, FDA reviewers seeking safety signals, or Big Data miners searching for as-yet unknown reasons, future users must be confident that the data represents the "truth."

Using CDASH facilitates consistent, well-defined data across studies. Without that confidence, at best the data will produce vague associations; at worst, it may kill us.

Authors & Acknowledgements

Kit Howard, Director Education, CDISC, khoward@cdisc.org
Shannon Labout, VP Education, CDISC, slabout@cdisc.org

With thanks to the CDASH and SDS teams

cdisc