



PhUSE US Connect 2019

Paper SD08

CDISC Doesn't Tell You What to Collect, But Regulators Do!

Diane Wold, CDISC, Austin, Texas, US

ABSTRACT

One of CDISC's messages is that CDISC standards don't tell you what data to collect. For example, SDTM "required" variables are minimal, those without which a record would be useless. However, regulatory authorities can and do say what they expect sponsors to collect. That means a sponsor can't rely on CDISC standards to tell them all the requirements for a regulatory submission. This talk examines some regulatory documents which affect requirements for data collection and the preparation of a submission that uses the CDISC standards. Examples include the FDA Technical Conformance Guide requirement to include the EPOCH variable, the inclusion of LOINC in the FDA Data Standards Catalog, and the subject accountability data required by the FDA Bioresearch Monitoring Technical Conformance Guide.

INTRODUCTION

Those who are preparing datasets for a regulatory submission would love to have a "one stop shop" that tells them everything they need to know about preparing submission datasets and metadata. Unfortunately, CDISC can only provide information about correct use of their standards; regulatory agencies provide additional information about what they want. This paper examines information available from CDISC, from FDA, and other sources, such as PhUSE, and some of the gray areas between these sources.

DATA COLLECTED

Let's start by looking at what data is collected. CDASH designates some fields as "highly recommended", SDTM designates some variables as "required" or "expected" and ADaM designates some fields as "required" or "conditionally required." However, the decision about what data to collect in a clinical trial is ultimately driven by science and regulatory guidance, not the standards.

- CDASH core designations are recommendations based on the experience of team members, which may in turn be based on regulatory requirements or guidance.
- SDTM variables designated as "required" are those without which the record is unusable. For general observation class domains these variables are the identifiers STUDYID, USUBJID, DOMAIN, --SEQ and the topic variable for the class. One variable which is designated as "required" based on regulatory guidance is SEX, but this is a rare exception.
- SDTM variables designations of "expected" are based on the experience of team members. A study that did not collect this variable would result in a dataset in which all instances of an expected variable are null; this would be completely valid, as far as the standard is concerned.
- ADaM variables designated as "required" include the identifier variables similar to those for SDTM datasets, with PARAM and PARAMCD playing a role in ADaM datasets that use the basic dataset structure (BDS) similar to that of topic variables in SDTM datasets. Required ADSL variables also include ARM and treatment variables, as well as a few variables that reflect regulatory guidance (AGE, AGEU, SEX, RACE).

For any particular clinical trial, various regulatory guidance about what should be collected may apply. Regulatory guidance may be very general, such as the ICH guidance on adverse events, or specific to a certain trial design or a particular disease area, reflecting science in that area.

ADDITIONAL DATA ITEMS IN CDISC DATASETS

If a piece of information is not collected, can it appear in a CDISC dataset? The answer is not as straight forward as you might think.

- CDASH includes data items that are pre-specified, as well as those that are entered in a CRF or otherwise provided as individual patient data items. Pre-specified data items might include the study number, visit



numbers, and pre-specified conditions or medications that are queried. So CDASH data includes both pre-specified and collected data items.

- SDTM includes information data sourced from case report forms or other sources, but also includes other variables, such as derived timing variables like study day variables and EPOCH, and whole derived datasets such as Subject Elements and Subject Visits. It also includes trial design datasets, which are sourced from the protocol.
- ADaM includes data sourced from SDTM, may impute values missing in SDTM, and derives a variety of parameters.

All these exceptions notwithstanding, data which is not collected or otherwise specified in the trial protocol, statistical analysis plan, or other study plan document will not be present in a CDISC dataset. The data in CDISC datasets has to be traceable back to some source.

FDA ADVICE ON THE REPRESENTATION OF DATA IN CDISC DATASETS

For submissions to FDA, important sources of information about what the agency wants in CDISC datasets include the Technical Rejection Criteria, Data Standards Catalog, Technical Conformance Guide, and FDA business rules. All of these include requirements that are not part of the CDISC standards.

FDA TECHNICAL REJECTION CRITERIA

Technical Rejection Criteria include the lack of a Trial Summary datasets, a Demographics dataset, and an ADSL dataset. The requirement for ADSL reiterates the requirement in the ADaMIG, "ADSL and its related metadata are required in a CDISC-based submission of data from a clinical trial even if no other ADaM datasets are submitted." The SDTMIG does not explicitly require the submission a Demographics dataset, although the SDTM says, "Demographics special-purpose domain is included with human and animal studies." Neither the SDTM nor the SDTMIG requires the Trial Summary dataset; however this dataset includes the Study Start Date parameter, needed by FDA to determine whether CDISC datasets are required for the study. The Trial Summary requirement is thus rooted in the effective date for the regulatory requirement for data to be submitted in CDISC format.

FDA DATA STANDARDS CATALOG

The Data Standards Catalog includes submission terminology standards, including CDISC controlled terminology, MedDRA for adverse events, and WhoDrug for concomitant medications. Although the variable LBLOINC has been present in the SDTMIG since the first version (3.0), LBLOINC has a core designation of "permissible" and has seldom been populated in industry trials, so the addition of LOINC to the data standards catalog caused concern. The inclusion of LOINC in the data standards catalog is not the same as making LBLOINC a required SDTMIG variable, i.e., it does not mean that LBLOINC must be populated in every record. The document, "Recommendations for the Submission of LOINC® Codes in Regulatory Applications to the U.S. Food and Drug Administration" provides detail on what is expected.

FDA TECHNICAL CONFORMANCE GUIDE (TCG)

The scope of the FDA Technical Conformance Guide extends beyond CDISC datasets. It includes, for example, discussion of the Study Data Reviewers Guides developed through PhUSE. It does include a number of "should" statements about CDISC datasets. Those statements sometimes reiterate information in a CDISC document, often add requirements not in the CDISC standards, and occasionally conflict with CDISC documents.

An example of a TCG statement that reiterates SDTMIG information is the sentence, "In the DM domain, each subject should have only one single record per study." This is a restatement of the structure of the DM dataset as "One record per subject."

An example of a requirement in the TCG that is in addition to the SDTMIG is the statement, "The variable EPOCH should be included for clinical subject-level observation (e.g., adverse events, laboratory, concomitant medications, exposure, and vital signs)." EPOCH is a timing variable, so has always been allowed in general observation class datasets, but the SDTMIG has not designated it as "expected."

An example of a statement in the TCG that was inconsistent with the SDTMIG is the statement, "Screen failures, when provided, should be included as a record in DM with the ARM, ARMCD, ACTARM, and ACTARMCD field left blank." SDTMIG versions 3.2 and earlier directed that arm variables for screen failures be populated with special values.

FDA BUSINESS RULES

Occasionally there is a more serious discrepancy between regulatory advice and a CDISC standard. One example is the FDA Business Rule FDAB005, "Age or age range should be provided for all subjects, except for Screen Failures." This presents a dilemma for sponsors who collect year of birth, but do not collect age, since the TCG also says, "SDTM and SEND datasets should not contain imputed data."



FDA BIMO DATA

The FDA Bioresearch Monitoring Technical Conformance Guide data needed by FDA for the process of selecting sites to audit. Some of the data items described are ones which could be represented in CDISC datasets. However, these datasets described are not datasets that follow either SDTM or ADaM. Sponsors may use CDISC datasets in their processes for creating datasets that conform to this guidance, but CDISC does not currently provide advice on how to do this.

RECONCILING FDA AND CDISC REQUIREMENTS

When a new version of FDA documents such as the TCG appears, members of the CDISC community note differences between the FDA document and the implementation guides, and often ask CDISC to update the standard to conform to FDA requirements. CDISC then has to decide whether to make changes to the standard, and what kind of changes to make.

In response to the FDA TCG advice about EPOCH, EPOCH was added to domain tables in SDTMIG v3.3, to avoid the impression that EPOCH is allowed only in domains that list it in the domain specification. (All timing variables are allowed in all general observation class domains, but a surprising number of implementers don't understand that.) The core designation for EPOCH remained "permissible" rather than "expected." The difference between these two core designations affects validation checks, but the difference is ultimately based only on the opinion of the SDS team, who developed and maintain the SDTMIG. Expected variables are those variables the team thought would usually be collected. Even commonly used derived variables such as study day variables are only "permissible," although the TCG also says that these should be included. ("Whenever --DTC, --STDTC or --ENDTC, which have the role of timing variables, are included in a general observation class domain, the matching Study Day variables (--DY, --STDY, or --ENDY, respectively) should be included.") EPOCH and the study day variables are clear cases where relying on the standard alone will cause you to miss regulatory requirements.

The TCG statement about the population of arm variables for screen failures led to a bigger change in SDTMIG v3.3. The SDS team decided that including special values in the arm variables created overloading of the variables, since it combined arm information with information about why the subject was not assigned to or did not follow an arm. SDTMIG 3.3 introduced two new variables, ARMNRS (Reason Arm and/or Actual Arm is Null) and ACTARMUD (Description of Unplanned Actual Arm) to relieve this overloading.

The FDA business rule about including age is not easy to reconcile. A sponsor may have based their decision to collect year of birth rather than age on their interpretation of privacy regulation, and CDISC cannot dictate that sponsors change to collecting age. A sponsor who collected only year of birth would have to decide whether to include an imputed age or leave age null, and either would run contrary to some piece of FDA advice. This decision would best be made in consultation with the relevant reviewing division at the agency.

CONCLUSION

Those who prepare CDISC datasets for regulatory submission must consult regulatory documents in addition to the CDISC standards. Those regulatory documents include guidance on what data should be collected and advice about the representation of data in CDISC datasets.

REFERENCES

CDISC standards, available from the CDISC website <https://www.cdisc.org/standards>

FDA Study Data Resources, available at

<https://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm>

"Recommendations for the Submission of LOINC® Codes in Regulatory Applications to the U.S. Food and Drug Administration," available at

<https://www.fda.gov/downloads/forindustry/datastandards/studydatastandards/ucm586363.pdf>

"Bioresearch Monitoring Technical Conformance Guide," available at

<https://www.fda.gov/downloads/drugs/developmentapprovalprocess/formssubmissionrequirements/ucm332468.pdf>

CONTACT INFORMATION

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

Diane Wold
CDISC
401 West 15th Sreet
Suite 800
Austin, TX 78701
Office Phone: +1 512 363 5826
Mobile: +1 919 306 2414
Email: diane.wold@cdisc.org
Web: cdisc.org



Brand and product names are trademarks of their respective companies.