



PhUSE US Connect 2019

Paper SI08

Considerations for Using CDISC Standards in Observational Studies

Jon Neville, CDISC, Austin, TX, USA
Bess LeRoy, CDISC, Austin, TX, USA

ABSTRACT

Historically, CDISC standards have primarily been used for regulatory submissions of clinical trials data in support of approval to market medical products. However, recent expansion of CDISC standards through therapeutic area user guide (TAUG) development and an increase in CDISC visibility has led to the recognition of the value of data standards in other areas of medical research as well. The existing biomedical conceptual content of CDISC standards, described mostly in TAUGs, is study type-agnostic and aligns well with analogous concepts examined from limited comparisons of data collected in observational studies. Despite this alignment, there is still confusion about the suitability of CDISC for this application. By seeking broader input on the unmet needs of the research community and examining more use cases, CDISC aims to develop a considerations document to address issues in implementing standards in these types of studies.

INTRODUCTION

Observational studies differ from randomized controlled trials in significant ways regarding study goals, study design, subject populations, clinical settings, regulatory/study oversight requirements, and data collection / data management practices. Many of these differences present challenges that are seen as barriers to the adoption of CDISC standards in observational research. In this paper, we discuss the specific challenges they present, and discuss at a high level how to reduce barriers to broader adoption of CDISC standards in the areas of medical research.

As a part of CDISC's ongoing effort to address these challenges, we ultimately propose producing a considerations document designed to guide users within the observational research community on how best to implement CDISC standards in their studies. We aim to collaborate with the PhUSE *Data Standards for Non-Interventional Studies* Workgroup to produce considerations and solutions that meet the needs of a variety of users and use-cases as will be identified by stakeholder engagement.

In this paper, we focus on some of the most commonly identified challenges reported by stakeholders interested in this area: conformance to CDISC rules and gaps in biomedical concepts. We will briefly discuss analysis challenges that will ultimately also be addressed as part of this project. We will briefly discuss what a solution or recommendation to address some of these issues might look like, but these should not be viewed as CDISC's official positions on these issues. Our ultimate recommendations will be published in the considerations document that will come at a later date.

It is also important to note what is *not* in scope for this project. CDISC is not proposing at this point to produce a full implementation guide for observational studies. Also, we do not intend to produce solutions or guidance for addressing the numerous data quality issues inherent to legacy data conversion projects, another commonly-identified challenge in using CDISC standards for observational studies. Finally, at least for the first phase deliverables, we do not intend to focus on "real world data."

OBSERVATIONAL DATA

Unlike a randomized controlled trial, observational studies do not involve an intervention and no attempt is made on the part of the investigator to impact health outcomes. When collected in an academic or government research setting, observational data are often of high quality; these studies are protocol driven and subject to oversight by an Observational Study Monitoring Board. Like randomized controlled trials, observational studies vary in the study design employed and can be generally categorized as case-control, cohort, or cross-sectional studies. The intent of a randomized controlled trial is to determine the safety and/or efficacy of an intervention. In contrast, observational studies seek to relate potential risk factors to disease outcomes. Because of the lack of randomization, observational studies are more prone to bias and thus potential confounding factors must be collected in order to control for bias during analysis. Beyond research driven studies, observational data may also be generated from real world data sources including electronic health care records, claims and billing, patient registries, and mobile devices. These data have generally not been collected with the intent of supporting research and thus may be less complete and of lower



quality than data collected in a research setting. This paper's focus is limited to challenges of applying CDISC standards to observational data generated in research settings.

CHALLENGES AND CONSIDERATIONS FOR OBSERVATIONAL RESEARCH

In this section we discuss some of the issues identified to date by various stakeholders engaged in observational research, including the PhUSE *Data Standards for Non-Interventional Studies* working group team members.

Note that we do not consider the example issues described below to be a complete list of the issues users face when attempting to implement CDISC standards in observational research. Indeed, there are many more challenges not addressed in this work. A more comprehensive list will be developed by way of stakeholder engagement later. As mentioned in the introduction, legacy data quality issues and other issues related to legacy data conversion are not in scope for this project.

CONFORMANCE AND DATA VALIDATION

A commonly-identified challenge related specifically to using CDISC standards in observational studies is the inability to meet SDTM conformance rules and subsequently failing validation checks. Conformance rules can apply at the dataset level, the variable level and at the controlled terminology level. Validation checks cover these rules and additional rules applicable to regulatory submissions.

Table 1 below summarizes FDA validation rules around inclusion of datasets, and how these rules may present problems in observational research.

Table 1.: SDTM datasets required or expected by FDA rules, and the challenges these requirements would present in observational research.

Conformance Rule	Flag Type	Challenge Presented
Demographics (DM) dataset must be included in every submission.	Error	Inclusion of the dataset should not present a problem. However, some required/expected variables will not be available (See table 2 below)
Adverse Events (AE) dataset should be included in every submission.	Warning	Depending on study type, these data may not be available
Lab Test Results (LB) dataset should be included in every submission.	Warning	Depending on study type, these data may not be available
Vital Signs (VS) dataset should be included in every submission.	Warning	Depending on study type, these data may not be available
Exposure (EX) dataset should be included in every submission.	Warning	Observational studies are not interventional studies. As such, exposure data will not be relevant.
Disposition (DS) dataset should be included in every submission.	Warning	Subjects won't likely meet formal milestones, nor will they have formal study completion/withdrawal dates.
Subject Elements (SE) dataset should be included in every submission.	Warning	Trial arms and elements are not relevant to observational research. Therefore, neither are subjects' progression through these.
Trial Arms (TA) dataset should be included in every submission.	Warning	Observational studies do not have rigid study designs with planned arms.
Trial Elements (TE) dataset should be included in every submission.	Warning	Without trials arms there are no elements to describe.
Trial Summary (TS) dataset must be included in every submission.	Error	Observational studies are not trials, but investigators could possibly create study parameters to describe here. It would require new controlled terminology and could be burdensome if it were considered a "required" dataset in observational research.

In addition to these dataset-level conformance rules, Table 2 describes several variable-level expectations and the challenges those may present. Given the nature of the data sources in observational studies, potentially any required/expected--or conditionally warranted/useful) SDTM variable could be missing. Note that this list focuses on variables whose requirement would most likely present a challenge broadly across observational research; potentially many more expected variables could present challenges on a case by case basis. Additionally, when it is expected that an entire dataset would most likely be wholly missing (as described in Table 1), required and expected variables from that dataset are not all shown here. It is not likely that investigators would be able to produce even partial datasets in these cases and would therefore not encounter validation errors from individual missing variables within.

Table 2.: Examples of SDTM variables required or expected by FDA rules, and the challenges these requirements would present in observational research.

Variable(s)	Domain	Core	Challenge Presented
RFSTDTC / RFENDTC	DM	Expected	Study reference periods will not always be relevant. Defining these dates can be challenging. Sometimes dates will be missing altogether.
RFXSTDTC / RFXENDTC	DM	Expected	Observational research does not include regimented exposure to a protocol-defined drug. Phase IV studies / Post-marketing surveillance could possibly provide these



SITEID	DM	Required	Observational research includes observations from across healthcare and clinical settings. These will likely vary and not be available in the data anyway
ARM / ARMCD ACTARM / ACTARMCD	DM	Required	There are no arms to describe in observational research.
VISITNUM	Multiple	Sometimes Required	The concept of “visit” may not be relevant in observational research
EPOCH	Multiple	Sometimes Required	Use cases for observational research have not been explored. Existing controlled terminology is specific to clinical trials

GAPS IN EXISTING BIOMEDICAL CONCEPTS

It would seem reasonable that the biomedical conceptual content of observational studies would not present significant challenges to implementing CDISC standards. After all, biomedical concepts are agnostic to the type of study in which they are used, so the only gaps present when compared to the biomedical concepts already represented in CDISC standards should theoretically be small. This is not to say that gaps in biomedical concepts are completely unexpected; indeed, this is another obstacle frequently cited by stakeholders interested in using CDISC standards in their observational studies.

In CDISC’s only first-hand experience mapping observational study data (a cancer epidemiology study), though most concepts collected in the study aligned well with existing concepts from published CDISC guides, approximately 25% of the items collected on the CRFs were conceptually new. Observational studies are often interested in assessing risk and risk-mitigating factors, including environmental exposures, lifestyle habits, socioeconomic factors, and other considerations often not relevant to clinical trials. CDISC has approached implementation strategies for some of these concepts, mostly through the development of infectious disease therapeutic area standards (e.g., Tuberculosis, Ebola, and HIV) and the draft Environmental and Social Factors (ER) domain. However, these solutions are incomplete, and some remain unpublished.

A more thorough survey of observational studies data is required to better determine the nature of the gaps in conceptual content and which strategies, including new concept development, would best address them. Any new development work would follow the procedures outlined in CDISC Operating Procedures COP-001.¹

ANALYSIS CONSIDERATIONS

As mentioned above, the concept of visit in the context of observational studies is not nearly as rigid as a visit in a clinical trial setting. Defining visits and windowing around those visits in ADaM therefore also presents a challenge. This may be particularly true in retrospective studies where partial and missing dates may be common and not resolvable.

The ADaM use case for observational studies data is a primary focus of the PhUSE *Data Standards for Non-Interventional Studies* working group and was the subject of a presentation at the PhUSE EU Connect 2018 meeting.²

NEXT STEPS: DEVELOPMENT OF A CONSIDERATIONS DOCUMENT

To better address the challenges encountered when using CDISC standards in observational research, CDISC proposes development of a considerations document to guide users on best-practice recommendations. Though no formal time line has yet been established, CDISC has begun to develop a project plan. A brief overview of this draft plan follows.

SCOPE

Project scope will be defined as part of a normal scoping phase as described in COP-001. Scope definition will take into consideration which CDISC standards to cover for v1.0 (end-to-end, SDTM-only, etc.) as well as which observational study types and data sources will be covered. As discussed above, it is our current intention to focus on academic research and exclude electronic health records and electronic medical records (EMR/EHR) data in the initial scope.

DELIVERABLES

The principle deliverable of this effort will be the considerations document itself. This concise document will introduce readers to the issue, and CDISC’s proposed path forward. This will likely include some details on stakeholder feedback and limited use case evaluation. This document will be sent for broader community review as a CDISC “position paper” and as the beginnings of a proposed solution. This document will provide direction and recommendations for implementing the CDISC standards in observational studies. Implications for conformance rules will be discussed within this document as well.

Additionally, a 30-60 min presentation will be developed to introduce the user community to the outcome of the



scoped work. This presentation will be presented via three web conferences: one each at a time convenient for users based in the U.S., Europe and Japan.

STAKEHOLDER ENGAGEMENT

In order to produce a document that is relevant to the user community's needs, CDISC will devise a stakeholder engagement plan to solicit feedback and input from CDISC members and collaborators from government agencies and patient research organizations. The engagement plan will include surveys and one-on-one calls as necessary or requested. Collaborating institutions may include, among others, FDA, NIH, PMDA, the Bill and Melinda Gates Foundation, World-Wide Antimalaria Resistance Network, Oxford University, Japan Agency for Medical Research and Development (AMED), Frontier Science Foundation, and any other stakeholder interested in participating.

POSSIBLE OUTCOMES

Though this work has not formally begun, we can envision how some details might unfold over the course of development as they pertain to the issues outlined above. For example, it is possible that CDISC, in collaboration with stakeholders, could develop a set of conformance rules relevant to observational research. Whether or not this involves any new rules or just a reduction in the current rules is yet to be seen, but we do know that some existing rules pertaining to the presence of certain datasets and variables would not be part of the conformance rules for observational studies. Once these rules are developed, CDISC could then discuss with technology vendors like Pinnacle 21 the possibility of developing validation rules or possibly allowing the toggling on and off of certain validation checks on a study by study basis.³

In the event that new biomedical concept development is required, such development would go through the process described in COP-001 as stated earlier. The implementation strategies developed to address these concepts would be published on a provisional basis, perhaps as an appendix to the considerations document until such time that they are rolled into the foundational standards to which they pertain, as applicable. It is not envisioned that a user guide of any sort will be developed to accommodate the material developed in this work at this time. CDISC may revisit that decision depending on the outcome of the project.

CONCLUSION

CDISC standards have gained much visibility in the research community in recent years. Standards implementation guidance in this area continues to be requested and remains a largely-unmet need. Despite the fact that the majority of existing CDISC standards and associated rules are compatible with research data, there are some gaps that need to be addressed. Various workarounds to some of the issues described herein have been developed and reported by stakeholders and collaborators, but these strategies are inconsistent across studies and investigators, and none have been formally endorsed by CDISC. A true, robust solution entails the development of a unified approach that considers broad use cases and stakeholder needs, as well as existing CDISC rules. CDISC will be soliciting input and feedback from stakeholders on the issues confronted while attempting to implement CDISC standards in observational research and will use the information gathered to guide the development of a considerations document as a formal statement of CDISC's position and recommended approach to handling observational data.

REFERENCES

1. CDISC Operating Procedure COP-001: Standards Development, Revision 3.0. 15 July 2017 <https://www.cdisc.org/sites/default/files/resource/CDISC-COP-001-Standards-Development-2017.pdf>. Accessed January 2019
2. Bahatska, Y and Ivanushkin, V. *First Aid Kit for an Observational Study*. 2019 PhUSE EU Connect. Paper PP24
3. Pinnacle 21. SDTM Validation Rules. <https://www.pinnacle21.com/validation-rules/sdtm>

ACKNOWLEDGMENTS

The authors wish to acknowledge Janet Reich (Amgen) and Yuliia Bahatska (Syneos Health) for their input on this project.

CONTACT INFORMATION

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

Jon Neville
CDISC
jneville@cdisc.org
www.cdisc.org



Brand and product names are trademarks of their respective companies.