

ADaM Implications from the “CDER Data Standards Common Issues” and SDTM Amendment 1 Documents

Sandra Minjoe, Octagon Research Solutions, Wayne, PA

ABSTRACT:

Over the past few years, the United States Food and Drug Administration (USFDA), specifically the Center for Drug Evaluation and Research (CDER), has been receiving more data from sponsors in a Clinical Data Interchange Standards Consortium (CDISC) or CDISC-like structure. Reviewers have had tools built and received training, but there are some technical issues with many submissions that are hindering their review process and thus their full adoption of CDISC. This prompted the issuance of a document entitled “CDER Common Data Standards Issues Document”.

Working closely with CDER to address many of these issues, the CDISC Submission Data Standards (SDS) team created an Amendment to the Study Data Tabulation Model (SDTM) version 1.2 and the SDTM Implementation Guide (IG) version 3.1.2. The CDISC Analysis Data Model (ADaM) team has not created a corresponding amendment, but there are many issues noted in the CDER document that have implications on ADaM. This paper examines issues that could affect ADaM, and describes how to handle them so that data and supporting documents submitted to FDA CDER are reviewer-friendly.

INTRODUCTION:

When first reviewing the CDER document, I was struck by how much ADaM material was included. I knew that the SDTM amendment had been created to address issues from the CDER document, and had thus expected that the CDER document would be focused specifically on SDTM. Instead, I found that many of CDER’s issues either directly or indirectly applied to ADaM.

As I absorbed the content of the document, I noticed that the ADaM implications fell into three basic categories: content, traceability, and integration. This paper is thus organized around those categories, and within each category I’ve described each issue, quoted the CDER document, and noted implications for ADaM.

ADAM CONTENT:

The first category of issues related to ADaM found in the CDER document is around the content of ADaM. The CDER document talks about the types of analysis data that should be part of our submissions. It describes how to handle imputations and mentions specific data that our submissions should include.

Imputed Data

Data imputation is mentioned on page 11 of the CDER document: “SDTM should not include any imputed data. If there is a need for data imputations, this should occur in an analysis dataset, and the relevant supporting documentation to explain the imputation methods must be provided.”

ADaM already provides for data imputations and the supporting documentation. Consider, for example, how a date is imputed in ADaM: not only do we produce the imputed numeric date variable, we must also provide the date flag to show what was imputed and metadata to describe the imputation rule followed. The CDER document is thus just emphasizing an already-existing ADaM concept, and noting that imputation is an ADaM task rather than an SDTM one.

Expected Analysis Data

There are several quotes on page 10 of the CDER document that talk about the types of analysis data they are expecting to receive. The first reads “All submissions containing standard data are expected to contain an ADSL file for each study.” (Underlining is from the CDER document and not added here). One of the reasons for including ADSL is because “SDTM datasets do not have core variables (such as demographic and population variables) repeated across the different domains. The need for such duplication of core variables across various domains can be fulfilled through their inclusion in the corresponding analysis datasets.”

ADSL is structured as one-record-per-subject, so it is easily merged or joined with any other data, including SDTM data. When reviewers receive both SDTM data and ADSL, they get all the advantages of using their tools with SDTM data, plus can tack on the ADSL core variables to the SDTM data for other review needs.

The ADaM document says that we should create ADaM data to support all of our analyses. The CDER document tells sponsors to "submit analysis datasets with their application to support key efficacy and safety analyses", implying that we may not need to submit some of the analysis data we create and use for generating analysis results.

If we're going to call a submission CDISC-compliant, it should then include SDTM, ADaM ADSL, and the ADaM data that supports our key analyses. Although we might create additional ADaM datasets for non-key analyses, these may not need to be included in a submission to CDER.

ADSL Content

The CDER document even includes some information about what they expect to see in our ADSL data. On page 10, it reads: "In addition to the variables specified for ADSL in the ADaM Implementation Guide, it is expected that the sponsor will include multiple additional variables representing various important baseline patient characteristics." This is not an oversight of the ADaM IG; baseline characteristics were not listed in the table in section 3.1 because they are study-specific. For example, the baseline characteristics needed for a cancer study are likely different than the baseline characteristics needed for a diabetes study.

In fact, section 1.3 of the ADaM IG makes a more general statement about the content of ADSL: "It contains variables such as subject-level population flags, planned and actual treatment variables for each period, demographic information, stratification and subgrouping variables, important dates, etc. ADSL contains required variables (as specified in this document) plus other subject-level variables that are important in describing a subject's experience in the trial."

We don't have to guess at what additional variables would be important to include in ADSL. To determine these, we can simply look at our analysis needs, and include variables that will allow us to produce the standard Demographics and Baseline Characteristics table(s) and the Disposition table(s). Thus for any study, ADSL will consist of the core set of variables specified in Section 3.1 of the ADaM IG, plus additional study-specific variables needed to produce tables for demography, baseline characteristics and disposition.

TRACEABILITY:

Traceability is the second category of issues from the CDER document that have implications to ADaM. In their document, CDER mentions their need to trace data within and across the CDISC models.

Traceability is also one of the fundamental principles of ADaM. Section 3.1.1 of the ADaM document starts off with this paragraph:

The concept of traceability is a cornerstone of the Analysis Data Model. This property enables the understanding of the data's lineage or the relationship between an element and its predecessor(s). Traceability facilitates transparency, which is an essential component in building confidence in a result or conclusion. Ultimately, traceability in ADaM permits the understanding of the relationship among the analysis results, the analysis datasets, and the SDTM domains.

Consistency Across Data

As surprising as it might seem to those of us who have been working with CDISC data for awhile, CDER seems to be receiving study data with inconsistent values in USUBJID. On page 11 of the CDER document they state "An individual subject should have the exact same unique identifier across all datasets, including SDTM and ADaM". They continue: "Improper implementation of the USUBJID variable is a common error that is seen with many applications, and often requires sponsors to re-submit their data." As mentioned earlier, reviewers often want to merge/join ADSL with SDTM data, but without a consistent USUBJID this isn't possible.

Different USUBJID content could be the result of ADaM data that is not created directly from SDTM, but instead from some other internal analysis data source. If internal processes don't allow us to create ADaM directly from SDTM, we need to be extra vigilant in checking for this issue, and watch for leading or trailing spaces, zeros, the placement of dashes, etc. Also note that ADaM has developed a list of automatable compliance checks, and check #53 will catch this.

Tracing ADaM back to SDTM

Page 6 of the CDER document describes their expectations of the relationship between SDTM and ADaM. Quite simply, "Analysis datasets should be derivable from the SDTM datasets".

All of our ADaM variable-level and parameter-level metadata should point back to SDTM, possibly via other ADaM data. If any interim datasets are used to create submitted ADaM data, these should also be submitted to show that traceability. The easiest way to meet this traceability expectation is to derive ADaM data directly from SDTM data, since that would allow us to use our specifications to write our metadata.

As mentioned earlier, some companies have internal processes that don't permit the creation of ADaM directly from SDTM, such as when SDTM data is created directly from the internal raw data source and ADaM data is created directly from the internal analysis data source. However, this statement in the CDER document means that even if we are not deriving ADaM data from SDTM, our documentation is expected to show how it would be done. This complicates matters, because the documentation provided with the ADaM data will then contain derivations that are different than what was actually performed to create the analysis datasets. Until SDTM and ADaM become part of their internal data flow, companies that don't follow a linear process to produce ADaM from SDTM will find it difficult to meet this expectation.

Derivation of Study Day

In SDTM, all study days, including --DY, --STDY, and --ENDY must be derived by subtracting the reference start date (RFSTDTDC) from the corresponding --DTC, --STDTC, or --ENDTC variable. The CDER document, on page 11, notes that "For most study designs, RFSTDTDC should be the start of treatment."

The study day in ADaM may need a different derivation than the SDTM counterpart did. Instead of deriving study day based on reference start date, our ADaM specification may have us deriving study day based on date of randomization, informed consent, or any other study date. This is perfectly acceptable for analysis data. Also, since SDTM doesn't allow for imputed dates, in ADaM we might be able to derive an analysis study day for records that have a missing study day counterpart in SDTM. Thus even if a study day variable exists in SDTM, we'll need to confirm that the derivation is the same before simply copying the content to the corresponding ADaM variable. If our analysis study day in ADaM differs from the SDTM study day, we'll also need to ensure that our data and metadata make it clear that there is a difference and denote which variable was used for analysis.

Permissible SDTM Variables that CDER Expects

The Core values of "Required", "Expected", and "Permissible" are SDTM terms and not CDER ones. Some of the variables listed in SDTM as permissible are, in fact, variables that CDER expects to see. As noted on page 10 of the CDER document, this includes baseline flags and variables --DY and --STDY.

ADaM datasets include analysis baseline flags and analysis study days, and in some cases we might be able to copy from SDTM rather than derive these variables in ADaM. However, because an ADaM baseline flag can have a different or more complicated derivation than the SDTM counterpart, and ADaM baseline flags and study days can make use of imputed dates, we must carefully review how the SDTM variable was derived before simply copying the content to the corresponding ADaM variable. If our analysis baseline flag in ADaM differs from the SDTM baseline flag, we'll also need to ensure that our data and metadata make it clear that there is a difference and denote which variable was used for analysis.

New SDTM Variables

The SDTM Amendment includes new variables in the events class and demography domain. These new variables were added specifically to help CDER reviewers. In fact, the CDER document references the SDTM Amendment and notes on page 4 that "Unless otherwise instructed by a division, these represent CDER-desired changes that sponsors should account for." With this endorsement, companies should be moving to adopt this content as soon as possible in order to help their reviewers at CDER.

This means that as we're preparing ADaM data, we'll now need to look for these new variables. Some of the new DM variables are similar to ADaM ADSL variables, and we might be able to copy from SDTM rather than derive a variable in ADaM. Examples include SDTM's ACTARM which could map to ADaM's TRTSEQA, and SDTM's RFXSTDTDC which could map to ADaM's TRTSTDT and TR01STDT. Keep in mind that an ADaM variable might need a different or more complicated derivation than the SDTM counterpart, or might use imputed dates, so we must carefully review how the SDTM variable was derived before simply copying the content to the corresponding ADaM variable. If any of our analysis variables in ADaM differ from similar variables in SDTM, we'll also need to ensure that our data and metadata make it clear that there is a difference and denote which variable was used for analysis.

SUPPQUAL

The supplemental qualifier (SUPPQUAL) data structure in SDTM was designed to hold any data that doesn't fit into variables found in the other structures. The SDTM documents make no mention of SUPPQUAL data as being any less important than other data, but because it doesn't reside in the main domain this data can be easily overlooked. On pages 7-8 of the CDER document, we're warned that "Discussion needs to occur if the sponsor intends to include important variables (that support key analyses) in the SUPPQUAL datasets."

When we develop our ADaM data, we must remember to look for source data not only in the SDTM domains but also in SUPPQUAL. Since ADaM can be derived from variables in both the main domain and SUPPQUAL, traceability to each location must be included. The values of SUPPQUAL variables QNAM, IDVAR (often set to --SEQ), and IDVARVAL will provide traceability to the specific SUPPQUAL record for the subject.

INTEGRATION:

Integration is the third category of issues noted in the CDER document with implications to ADaM. The document notes issues related to analysis that we need to take under consideration when we integrate data: the use of dictionaries and the potential for subjects to be included in multiple studies.

Although it is often an essential part of our submission, the topic of integration has not yet been well described in the SDTM and ADaM documents.

In March, 2012, the FDA participated in a Computational Sciences Symposium, and one of the work groups was devoted specifically to integration. That combined FDA and industry working group is currently considering options for the optimum place to integrate data, and noted that different options might be needed to handle different situations.

Dictionaries

With a set of studies that collected data over a long timeframe, we often find that a different dictionary or version was used in one study than was used in another. The CDER document explains on page 7 what to do with adverse event data when we integrate: "It is expected that the Adverse Event dataset for the Integrated Summary of Safety include MedDRA Preferred Terms from a single version of MedDRA." One reason is because "reviewers often want to analyze adverse events across trials, including the use of Standardised MedDRA Queries".

While adverse event data can make up a large portion of our integrated data, it is not the only type of integrated data that can be mapped to a dictionary. The CDER document continues: "It is expected that common dictionaries are used across trials and throughout the submission for each of the following: adverse events, concomitant medications, procedures, indications, study drug names, and medical history."

Data mapped into different dictionaries or versions can't easily be summarized, so these statements from CDER should come as no surprise. At the time of compiling integrated analysis datasets, we should confirm that a common dictionary has been used across all studies for any analyzed data.

Re-mapping data, especially from one dictionary to another, can be a lot of work. If not planned for ahead of time, this type of activity can cause a considerable delay in producing final deliverables for a submission.

Subjects in More than One Study

When combining data for an integrated analysis, sometimes we find subjects were enrolled in more than one study. As the CDER document points out on page 8, "In the DM domain, each subject should have only one single record per study." This might not, however, be the case with a multiple-study integrated DM domain. The CDER document continues: "Integrated summaries may contain more than one record per unique subject in the case that an individual subject was enrolled in more than one study".

The same issue applies to ADaM ADSL data. For a single study, we should have exactly one record per subject. In an integrated ADSL, when a subject has been enrolled in more than one study, it's possible to have more than one record per subject.

Interesting to note here is that this issue in the CDER document was specifically pointed out for DM, but there was no mention of it also applying to ADSL. We should be wary of reading too much into this. Although it could imply that CDER expects data to be integrated at the SDTM level, it may have simply been an oversight by CDER not to mention ADSL.

FDA AND CDISC:

FDA and CDISC have been working closely together for years. FDA reviewers sit on CDISC teams and contribute to the CDISC standards.

The CDER document and the SDTM Amendment were developed jointly and have been aligned where relevant. Members of the CDISC leadership team reviewed and contributed to the development of the CDER document. The SDTM Amendment will become part of SDTM version 1.3 and SDTMIG version 3.1.3, expected to be released by mid-2012.

The ADaM document and ADaMIG did not undergo an update to coincide with the release of the CDER document. Nothing stated in the CDER document contradicts anything in the current ADaM documents.

CBER and CDRH have not created a common data standards issues document like CDER, simply because they haven't yet received enough CDISC data to have common issues with it. However, representatives from the CBER organization were part of the development of the CDER document and have said at public meetings that they agree in principle to the statements made there.

FDA, PhUSE (Pharmaceutical Users Software Exchange), and CDISC are now working together to address a set of topics put forth by FDA. The annual Computational Sciences Symposium held in March, 2012 kicked off a set of working groups. Each working group will post information on the PhUSE wiki as it is developed, including recommendations to industry.

CONCLUSION:

The CDER document was developed to help us prepare our data for submission. However, it should be noted that on page 1, it reads: "The document is not intended to replace the need for sponsors to communicate with review divisions regarding data standards implementation approaches or issues, but instead, it is designed to compliment and facilitate the interaction between sponsors and divisions." The science must always come first, and there may be cases where it makes more sense to do something other than what is specified in the document. We should use this document as a tool to help us talk with our reviewers.

CDER has taken the time to describe many of the different issues with data that they have received. The document has been aligned with CDISC, making implementation fairly straightforward for us. It's clear that this is what the CDER reviewers want, so sponsors are encouraged to make these changes to their CDISC data as soon as possible. Sponsors submitting to CBER might also want to consider implementing these changes, though it's always good to have a discussion with the reviewers just to be sure it's what they need.

As more sponsors submit CDISC data to CDER, they will likely find new common issues, and we should expect to see new versions of this document over time. Also, when CBER and CDRH receive enough CDISC data to determine what their common issues are, we might then see either separate documents from them or a joint document across all of FDA (CDER, CBER, and CDRH). Future common issues might result in both the CDISC standards evolving over time to becoming more specific and the creation of additional documents by working groups such as those kicked off at the recent Computational Sciences Symposium.

REFERENCES:

"Analysis Data Model (ADaM)". <http://www.cdisc.org/adam>. The current version is downloadable from web page and available to CDISC members and non-members (access date 12Jan2011).

"Analysis Data Model (ADaM) Implementation Guide". <http://www.cdisc.org/adam>. The current version is downloadable from web page and available to CDISC members and non-members (access date 12Jan2011).

"CDER Common Data Standards Issues Document".

<http://wcmis.fda.gov/FDAgov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM248635>. The current version is downloadable from this FDA web page. A standard web search on the document title will quickly find the document (access date 21Dec2011).

"Study Data Tabulation Model." <http://www.cdisc.org/sdtm>. The current version is downloadable from web page and available to CDISC members and non-members (access date 12Jan2011).

"Study Data Tabulation Model Implementation Guide." <http://www.cdisc.org/sdtm>. The current version is downloadable from web page and available to CDISC members and non-members (access date 12Jan2011).

OTHER INFORMATION:

http://www.phusewiki.org/wiki/index.php?title=FDA_Working_Groups is the PhUSE Wiki for the FDA Working Groups that kicked off at the Computational Sciences Symposium in March, 2012. Members of the working groups post their own material as it is developed.

<http://www.phuse.eu/css> contains information about the FDA/PhUSE Annual Computation Sciences Symposium. The website includes links to the presentations and posters plus ways to get involved.

ACKNOWLEDGEMENTS:

Special thanks to those who helped review this and the predecessor material, including

- Mario Widel of Roche Molecular Diagnostics
- Steve Wilson of US FDA
- Fred Wood of Octagon Research Solutions

CONTACT INFORMATION:

Your comments and questions are valued and appreciated. The author can be reached at:

Sandra Minjoe
Octagon Research Solutions
585 East Swedesford Rd
Wayne, PA 19087 U.S.A.
sminjoe@octagonresearch.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.