

**Update: CDISC Interchange Abstract**  
**Harmonizing CDISC Models-SDTM &ADaM**  
**Edward Helton(SAS®), Stephen Ruberg(Eli Lilly), Cathy**  
**Barrows(GlaxoSmithKline), Musa Nsereko(Cephalon) and Greg Anglin(Eli**  
**Lilly)**

**ABSTRACT:**

Submission of data to FDA has been a requirement for many years and will continue to be so in the future. CDISC has focused considerable efforts on standardizing these efforts in order to help FDA understand industry data, to minimize programming and rework of the data at FDA for review, and to maximize the integration of data across studies and pharmacological class for broader scientific/medical evaluation.

CDISC has worked closely with FDA on these matters and the Submission Data Tabulation Model (SDTM) and the Analysis Data Model (ADaM) have matured to a point that they are commonly referenced in industry forums. They have garnered the attention of many in the industry as the industry seeks to implement them in ways that will streamline submission and facilitate review of the data. CDISC recognizes that the unity or interoperability of data standards is a requirement for both the submission and the review and approval process.

This presentation will discuss the use of these ADaM and SDTM models in the near-term as they emerge as useful data submission standards. It will then focus on the desired future state for CDISC models and a reasonable transition to that future state. There will be an update on the CDISC ADaM/SDTM Pilot Project with the US FDA to unify the implementation and application of these models and to explore the iterative process to standardize the e-submissions required by the FDA.

**INTRODUCTION/PLAN:**

The goal of the pilot project is to produce a working mock submission with the available CDISC standards. It will illustrate how the various CDISC components can be used to result in a submission of electronic data that are in a format that is acceptable to the FDA and meets the needs of both medical and statistical reviewers. To do this, the project team will use the tools currently available (with very minor modifications if any) to produce the pilot submission package.

The “full” submission package will include SDTM datasets, analysis datasets, all relevant metadata, analysis results, and an abbreviated report. It will address only the more common elements of a submission, including safety data, the primary

efficacy outcome, and one secondary efficacy outcome. A representative set of analyses will be included. A summary report of the pilot submission project will be produced that will include issues encountered, strengths and weaknesses of the models, as well as the FDA reviewers' report. The results of the pilot project will be made available to the public by posting a sample submission package and the summary report to a website.

The Pilot Project includes the evaluation of the submitted datasets (SDTM and AdAM), metadata and documentation from the perspective of the FDA statistical and medical reviewers. Further, an assessment of the package by FDA reviewers as being "reviewable" and "meets expectations" is required for the success of the project.

Legacy data to be used in the pilot submission are real clinical trial data, provided by Eli Lilly. The data have been de-identified, so there is no possible way to trace back to the original patient (maintaining confidentiality and privacy of patients was of paramount importance).

The objectives for the data and metadata is to both use and demonstrate where are the overlaps and differences between SDTM and Analysis models. Consideration will be given as to what improvements might be considered to optimize the SDTM and ADaM models. One emphasis is to demonstrate how derived variables in analysis datasets will be linked to SDTM (or the source data) and how derived variables included in SDTM will be referenced back to the derivation in ADaM. This will provide traceability of the data, ideally down to the row level.

## **CONCLUSION/RESULTS/FUTURE PLANS:**

The Pilot Project has begun by conducting a first case study using legacy data (real clinical trial data, warts and all) with CDISC SDTM domains and ADaM datasets and associated metadata. The submission of this case study package to FDA for mock review is being used to identify issues to be resolved in SDTM, ADaM and other CDISC models. Resolutions of any issues will be addressed in the subsequent iterations of the pilot. The status and progress to date will be presented at PharmaSUG 2006 and the results of the first iteration are to be presented at the CDISC Interchange in September 2006. A guidance with all results will be posted on the CDISC website.

Efforts and considerations for the future iterations (to include noted issues) of the pilot are underway as well and are briefly reviewed here. The CDISC Pilot Project research effort has a strong motivation that CDISC have the long-run objective of a "single data stream" submission, thereby having integrated submission of collected

data, derived elements, analysis data structures, and design and conduct information such as protocols, analysis plans, analysis planned vs. actual, etc. Obviously reaching this state will take much time and effort, but we think we can identify near-term steps and reach the long term goals with reiterative steps.

These objectives also include the research of issues involved in achieving a metadata-driven CDISC submissions data model with focus on issues identified during the pilot effort that provide suggestions to the CDISC working groups about potentially helpful changes. Thus we want to identify gaps in metadata standards relative to the “single stream” objective (e.g. resolve issues with “source” of derived variables on SDTM and ADaM datasets). We would like to show that it is technically possible to have one data stream and leverage the infrastructure we have while experimenting with updated ODM / Define features, metadata-driven dataset construction and prescriptive rather than descriptive metadata. Future iterations of the ADaM/SDTM Pilot Project might:

- Fully and completely work the example we have (rather than a subset)
- Use studies from other therapeutic areas
- Go beyond a single study ...
  - Data from multiple studies e.g. for which multiple coding dictionaries / code lists have been used?
  - Submitting different sets of sponsor data (e.g., NDA, safety update)?
- Address how to send updates (e.g. additional derived variables) effectively?

**CONTACTS:** [ed.helton@sas.com](mailto:ed.helton@sas.com) and [www.cdisc.org](http://www.cdisc.org)