Paper 385-2011

WHAT CDISC MEANS TO SAS PROGRAMMERS

Kevin Lee, Cytel, Inc., Chesterbrook, PA

ABSTRACT

The paper is intended for SAS programmers in the pharmaceutical industry who are interested in SAS programming in a CDISC environment. This paper discusses the basic structure of CDISC (Clinical Data Interchange Standards Consortium) and the current roles of SAS programmers in a CDISC environment. It also discusses the future of SAS programmers in a CDISC environment.

INSTRUCTION OF CDISC

The goal of CDISC is to catalyze the information flow through pre-clinical and clinical research process, from study protocol and various data collection to analysis and reporting through regulatory submission and electronic data archive. CDISC mainly consists of the following:

- Study Data Tabulation Model(SDTM)
- Analysis Data Model(ADaM)
- Operational Data Model(ODM)
- Laboratory Data Model(LAB)
- Case Report Tabulation Data Define Specification(CRTDDS) – Define.xml
- Protocol Representation(PR)
- Trial Design Model(TDM)
- Clinical Data Acquisition Standards Harmonization(CDASH)
- Terminology

CURRENT INVOLVEMENT OF SAS PROGRAMMERS IN CLINICAL TRIAL

The role of Clinical Trial SAS Programmers varies from simple SAS programming to generating the analysis data and the table, listing and graphs. The following are examples of various roles for clinical trial SAS programmers:

- Edit Check Programming
- eCRF annotation
- Creating the derived data sets
- Creating the tables, listings and graphs.
- Importing the external data and converting to SAS data sets.

WAHT IS NEW WITH CDISC

The implementation of CDISC standards requires more than programming from SAS programmers. The role of SAS programmers in CDISC depends on the job description in each company, but SAS programmers can be exposed to the following areas in CDISC environment other than SAS programming skill.

- SDTM – Electronic Data Capture(EDC) and Electronic Case Report Form(eCRF) annotation
- ADaM – Statistics
- ODM – XML programming knowledge
- LAB – Conversion of eternal data to SAS data set format
- Define.xml – Metadata of SDTM and ADaM and XML programming knowledge
- PR – Protocol
- TDM – Statistical Analysis Plan(SAP) and SDTM
- CDASH – EDC system
- Terminology

As indicated above, each standard requires different skills and backgrounds from SAS programmers. Usually SAS programmers get involved in developing SDTM and ADaM.

SDTM can be developed from raw data from EDC by SAS. So SAS programmers need to understand the following:

- Electronic Data Capture(EDC) system (ex. Oracle Clinical, Inform, ClinTrial, etc.)
  - Database Structure
  - Domains or tables
  - Data Type – numeric, character, timing, etc.
- Conversion of data to SAS format
- SDTM Concepts

ADaM is usually derived from SDTM. The basic principles of ADaM are the analysis readiness, the traceability and the link to metadata. In order to provide traceability and metadata, SAS programmers need to understand SDTM. To create analysis-ready ADaM, SAS programmers need to understand the statistical method for some of efficacy

analysis. SAS programmers also need to understand the exact SAS statistical procedures for analysis such as proc means, proc freq, proc ttest, proc glm, proc npar1way, proc reg, proc mixed, proc lifetest, proc phreg, proc logistics, and so on.

For example, when the analysis is COX model, SAS programmers need to know what procedure and statements are needed in COX model for the correct ADaM data set. Below are the sample COX model SAS codes.

```
**** P value for treatment and covariates;
ods trace on;
ods output parameterEstimates=_estimate;
proc phreg data=all;
    where param = 'OVERALL SURVIVAL';
    model aval*cnsr(1) = trt01pn covar1 covar2;
run;
ods output close;
ods trace off;
```

The above codes indicate that ADTTE(Time to Event ADaM Data) needs to have the covariates, COVAR1 and COVAR2. And the parameter, OVERALL SURVIVIAL, should be unique in each subject to satisfy the correct assumption. By knowing the exact SAS statements for all the analyses, SAS programmers are able to create analysis-ready ADaM data sets.

To provide the metadata of analysis, SAS programmers also need to know how to create XML output.

DISADVANTAGE

There will be more software package to analyze CDISC data beside SAS. Because CDISC is standardized form, it is much easier to import the data into the software for the automatic analysis. For example, JMP® Clinical recognizes CDISC data and automates the analytics and reporting of the safety data even though the users do not have CDISC knowledge. Therefore, some of SAS programming could be replaced because of the emergence of other CDISC specific software.

Because of the emergence of ODM, there will be a great demand in XML data format. XML format data will be used in data transfer more.

CDISC is a part of electronic clinical trial rather than paper, so the pace of CDISC clinical trial will go faster, so SAS programmers do not have as much time as before.

CDISC is the standardized process, so SAS programmers need to work in more structured and standardized setting than before.

ADVANTAGE

There will be more data in the health care industry to analyze because all the data including the clinical trial data become standardized. It will be much easier to obtain and analyze the data and SAS will provide the best solution. SAS won't be the only software, but simply one of the best out there specially in the data manipulation and analysis.

Due to the current involvement in the clinical trial, the role of SAS programmers will expand. Because SAS programmers are the last group in CDISC path, we will be in the ideal position to review the final output as well as all the CDISC clinical trial process. The SAS programmers' involvement in the CDISC clinical trial is likely to expand.

There will be more opportunities to SAS programmers who have CDISC experiences. Rather than starting from scratch, some sponsors will outsource CDISC implementation and manage that process. Therefore, SAS programmers who have extensive CDISC experiences will have more opportunities as consultants as well as project managers.

THINGS TO CONSIDER

The pharmaceutical industry heavily depends on SAS programming for data manipulation and statistical analysis. But, the FDA likes to be neutral in terms of vendor selection. XML is an open data model and vendor-neutral so even FDA discusses the possibility to accept SDTM and ADaM by XML format not by SAS XPORT format in the future. It is not sure if the transition from XPORT to XML will ever happen because of the nature of clinical data, but it is sure that XML format data will be used more. It will be a great advantage if SAS programmers know about XML data format.

Since all of the clinical data is standardized, it will be easier to integrate the clinical data. So there will be more opportunities in data mining. The company and agencies will try to analyze the clinical data across different clinical trials and therapeutic areas.

CDISC is a part of drive that all the data should be in the standardized form, so CDISC will follow the direction that its data can be merged into other standardized data such as HL7. So SAS programmers who have the ability to understand the different data structures and formats will be able to take a full advantage of the standardization movement in health care data. For example, unlike the normal tabular structure in the clinical trial such as SDTM and ADaM, ODM XML data structure is hierarchical.

In addition, SAS programmers will spend more time in developing the structured codes such as CDISC specific macros rather than the custom SAS programming.

SAS programmers should have a clear understanding on CDISC for FDA submission to avoid the delay or rejection by the FDA because the submission does not meet FDA expectation.

CONCLUSION

CDISC will present the disadvantages and advantages to SAS programmers. If SAS programmers want to succeed in the CDICS environment, they need to understand the purpose of CDISC implementation and each critical path and furthermore, learn the necessary skills. In addition, the current CDISC is not a final product. It will keep evolving. It is not sure how CDISC will evolve, but if SAS programmers are open to changes and are able to adapt to the new changes and skills, SAS programmers will be able to take a full advantage of it as CDISC progresses.

CONTACT INFORMATION

Your comments and questions are valued and welcomed. Please contact the author at

Kevin Lee

Cytel, Inc.
640 lee Road, Suite 201
Chesterbrook, PA 19087
(610) 994-9840
Email:Kevin.lee@cytel.com

.