PharmaSUG China 2021 - Paper DM-18

Don't Panic: A Data Manager's Guide to SDTM

ABSTRACT

SDTM, as one of the required standards for data submission to FDA and PMDA, turns to be an inevitable topic to data managers, who usually works as the source data and perhaps also SDTM data deliverer. It is a big topic including programming, datasets, standards, and all different kinds of reports, full of complex technical terms or program language. This paper will however present the SDTM from a data manager's point of view, to give a general idea with detailed examples what are the key knowledges a DM should know about SDTM. We will focus on the parts of SDTM that are related and do help on DM's work, from CRF design to data delivery.

INTRODUCTION

SDTM, with full name as study data tabulation model, defines the standard structure for study data tabulations that are to be submitted as part of a product application to a regulatory authority like FDA and PMDA. Data managers works as the data deliverer, who are usually more responsible for the source data but less knowledge of SDTM data. This paper would like to provide some basic SDTM knowledge and to show how these SDTM knowledge will support DM's daily work with examples.

SOME BASIC SDTM TERMS AND CONCEPTS

Before we start the journey to see how SDTM knowledge helps on DM's work, it is important to know some basic SDTM terms and concepts. You can always find these terms and concepts in the CDISC: SDTM and SDTM Implementation Guide on the CDISC website.

Domains: Domain is defined as 'A collection of observations on a particular topic'. To say in an inappropriate way, it somehow likes a CRF form or the combine of several similar CRF forms. For example, AE domain data usually comes from the Adverse Event form and DS domain contains the data from forms like Eligibility, End of Treatment, End of Study.

Observations: Observations 'consist of discrete pieces of information collected during a study. Normally correspond to rows in a dataset. Each observation can be described by a series of named variables.' Or to say it in a simple way, one single row of record in the SDTM datasets.

Variables: Variable 'normally corresponds to a column in a dataset, can be classified according to its Role'. Like Study ID or Subject ID or AE term, will be consider as variables in a SDTM dataset.

Roles: Roles 'describes the type of information conveyed by the variable about each distinct observation and how it can be used'. They can be further classified into different types. There are five major roles in the SDTM dataset, identifier variables, topic variables, timing variables, qualifier variables and rule variables.

Knowing these key terms and concepts will help you to read the SDTM datasets and give you the first general idea of what SDTM dataset contains.

As a data manager, when we first look at the SDTM dataset, we will see many familiar datapoints. It is because some SDTM variables directly come from or are derived from CRF data or external data, especially when the CRF design follows the CDASH standard. You will find the one-one relationship between SDTM data and CRF data. Other origin of SDTM data can be classified as protocol, assigned and predecessor.

SDTM CONSIDERATION DURING CRF DESIGN

As data managers, CRF and database design is always a big part of our work at the beginning phase of a trial. And we will then ask the questions that what datapoints to be collected and how to collect them into our database. At this timepoint, SDTM rarely comes to our mind.

However, SDTM IG shares us a very interesting example in the FA domain even the topic of these example itself is not related to the CRF design.

The example shows us a CRF form of pre-specified Adverse Events, which can be mapped to FA and AE domain. However, the record in FA and AE domain cannot be linked due the there is no corresponding data collected on the CRF.

The example gives us a new consideration point when we design the CRF. We should not only consider the protocol, team review comments, system but also SDTM when we start the CRF and database.

SDTM VALIDATION RULES TO SUPPORT DATA CLEAN

When run the SDTM validation rules, you can see all kinds of findings. And I try to category them into below types which are more related to DM's work.

eCRF Data Related Findings: These findings include incorrect dates, duplication records, missing data, coding issue, and other findings. It also contains some irresolvable issues depending on the trial progress.

For example, a very common finding during the study conduct is that AE start date is after the latest disposition date. If a subject has entered the study but not reached to the End of Study yet, the latest DSSTDTC is usually the Eligibility/Randomization Date of this subject. It is very likely that the AE Start Date/Time is after the Eligibility/Randomization Date. (e.g., AE occurs after treatment.) So, this finding could be considered as acceptable during the trial conduct. But it must be resolved at the end of trial

Database Set-up Issue: There will be some findings caused by the database set-up issue like unexpected characters or missing TPTs.

SDTM Mapping/Program Related Findings: It is more related to programming part so I don't give examples here.

It is always helpful for a data manager to include SDTM validation rule finding review as part of the data clean activity. As, SDTM is the data to be submitted and it gives findings in source data as well.

CONCLUSION

This paper tries to provide data managers, especially to those who are the SDTM beginners, with some basic knowledges, and points to be considered as a data manager on SDTM. From my perspective, SDTM and SDTM IG are always the first and basic 'textbook' for those who would like to learn SDTM. Hope you all good luck on your journey of SDTM learning.

REFERENCES

Website CDISC Submission Data Standards Team 26 November 2013 *Study Data Tabulation Model* Version 1.4 https://www.cdisc.org/standards/foundational/sdtm/sdtm-v1-4

Website CDISC Submission Data Standards Team 26 November 2013 *Study Data Tabulation Model Implementation Guide: Human Clinical Trials* Version 3.2 https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-v3-2

Website Pinnacle 21Validation Rules https://www.pinnacle21.com/validation-rules/sdtm

RECOMMENDED READING

- *CDISC: Study Data Tabulation Model*
- *CDISC: Study Data Tabulation Model Implementation Guide: Human Clinical Trials*

CONTACT INFORMATION

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

Lanjin Jin
Boehringer Ingelheim
13918308487
Lanjin.jin@boehringer-ingelheim.com