

A Taste of SDTM in Real Time

Changhong Shi, Merck & Co., Inc., Rahway, NJ
Beilei Xu, Merck & Co., Inc., Rahway, NJ

ABSTRACT

The Study Data Tabulation Model (SDTM) is a Clinical Data Interchange Standards Consortium (CDISC) standard. At Merck, pilot studies have been completed in which SDTM is implemented to the collected data up-front with the following proven advantages from SDTM: 1. standardization across studies; 2. facilitation of data integrity checking; and 3. tabulation data structures readily leveraged for ADaM (Analysis Dataset Model) data setup. This paper will describe these three advantages in detail.

Keywords: SDTM, CDISC, ADaM, ADSL

INTRODUCTION

SDTM defines a standard structure for study data tabulations that are to be submitted as part of a product application to a regulatory authority such as the FDA.

The SDTM model provides a framework to organize subject information collected during clinical studies. The collected data is reported into a series of domains depending on the topic of the information. The majority of the domains can be classified into three general observation classes: the intervention class, the events class, and the findings class. Each class has a series of standard variables with different roles in describing the observation including identifier variables, topic variables, timing variables, qualification variables, and rule variables.

Companies have different strategies concerning the implementation of SDTM into their internal process. In this paper, we will focus on the advantages of linear implementation of SDTM within the data collection process. This strategy is illustrated as;

Collected Data → SDTM Data sets → ADaM Data sets

When data is entered on a case report form (CRF) and is flowed to the database, it will be mapped to the SDTM domains where the SDTM data sets are created in the operational database. Note that the mapping of data collected on the CRF to the SDTM domains is not necessarily one-to-one. Usually, data in one domain corresponds to data that were collected on one CRF. However, depending on the topic, the data in one domain could be collected from one or more CRF's, and data collected on one CRF may be split into more than one domain.

In our experience, we have piloted studies implementing SDTM linearly in our product development process. This paper will describe some of the advantages we experienced in real time in three main areas: 1. standardization in programming for analysis and reporting; 2. facilitation in data management and integrity checking; 3. leverage of SDTM data structure for analysis data set.

1. STANDARDIZATION IN PROGRAMMING FOR ANALYSIS AND REPORTING

Under the SDTM model, different studies can have different designs, but the data structure would be the same. This is very helpful for programming consistency and efficiency across studies. Linear implementation of SDTM allows a logical flow of the program development life cycle. SDTM is the input for reporting and analysis; therefore it enables the downstream programming standardization in these areas:

- Standard safety reporting programs
- Setup programs for analysis data sets
- ISE/ISS

In Merck, we have developed a suite of standard macros which generate standard safety reports based on SDTM data. It can be used across studies with options which allow specific requirements to be passed to the macros. The chart below is an example of a standard macro called SPECIFIC0AE which generates three different specific AE tables based on the following options. The three standard AE tables generated are Drug-Related AEs, Serious AEs, and AEs causing discontinuation.

Deliverable File Name	ae0dr	ae0sae	ae0disc
Report Name	Subjects With Drug-Related Adverse Events	Subjects With Serious Adverse Events	Subjects With Adverse Events Causing Discontinuation
Subtitle	All Subjects as Treated	All Subjects as Treated	All Subjects as Treated
Report macro	specific0ae	specific0ae	specific0ae
CSR Section Number	12.2.4.1	12.2.4.1	12.2.4.1
Output format (RTF, PDF, SAS)	RTF	RTF	RTF
Therapy Grouping	TRTA	TRTA	TRTA
Additional Data Subsetting Parameters	AEREL="RELATED"	AESER="Y" and AEOUT ne "FATAL"	AEACN="DRUG WITHDRAWN"

Similarly, macros can be developed to create analysis data sets based on SDTM data. One quick hit is the ADSL (Subject Level Analysis Dataset). ADSL is the ADaM data set containing subject level information (e.g. demographics, population indicators, etc.). The source is primarily from the DM (Demographic) domain. A macro with options for protocol specific information can generate an ADSL data set for different studies. Other analysis data sets, e.g., analysis of change from baseline for an analysis endpoint, can also be created through a standardized macro which allows therapeutic area specifics.

In the same way, an ISE/ISS database can be built efficiently when based on the same raw SDTM data structure and the same standard macros across all studies.

2. FACILITATION OF DATA INTEGRITY CHECKS

In addition to the programming standardization for analysis and reporting, standard programming can be applied to data management and data integrity checks. Reports can be developed to report data issues during the active life of the study.

Another advantage of utilizing SDTM data in data management is the consolidation of data from the same topic. As stated in the introduction, data from the same topic is mapped to one domain though it is collected on multiple CRF's. The following example illustrates how we check subject status collected from different forms.

In one study, we have two phases - base and extension. The collector is designed to have two subject disposition forms; one for the base phase, and one for the extension phase. However, the data will not be represented in two domains. Instead, the data maps to one DS (Disposition) domain as two observations with one record representing the base status and one representing the extension status. Therefore, a subject will have one DS record for the base phase and one DS record for the extension phase if he/she entered into the extension.

A subject could have the following three possible statuses, with "/" separating the base and extension phase:

- COMPLETED/COMPLETED
- COMPLETED/DISCONTINUED
- DISCONTINUED

For the first scenario, the subject has one record with Disposition Event Subcategory, DSSCAT='COMPLETED' and its associated Start Date/time of the Disposition Event, DSSTDTC, for the base phase, and another record with DSSCAT='COMPLETED' and its associated DSSTDTC in the extension phase. For the third scenario, the subject discontinued in the base phase and there will be one record with DSSCAT='DISCONTINUED' and its associated DSSTDTC in the base phase.

- If a subject's status does not fall into the three scenarios above, it signals data issues. We can easily locate this issue by using the DSSCAT and DSSTDTC fields in the DS domain. Instead of looking into the two different disposition CRFs, we can check it in one place - the DS domain. A sample SAS[®] program below shows how to identify subjects with the problematic status of 'DISCONTINUED/COMPLETED':

```
*** Get All Randomized Subjects Status;
proc sort data = sdtm.ds(where=(subjid ne ' ' and dsscat ne ' '))
    out = ds(keep=usubjid subjid dscat dsscat dsstdtc);
    by usubjid dscat dsscat dsstdtc;
run;

*** Separate discontinued and completed status for both base and extension phases;
data discon comp;
    set ds;
    if dsscat eq 'DISCONTINUED' then output discon;
    else if dsscat eq 'COMPLETED' then output comp;
```

```

run;

*** If discontinuation date is before the completion date, it is a data issue;
data bad_discon_before_comp;
    merge discon(in=dd rename=(dsstdtc=dis_dt) drop=dscat dsscat) comp(in=cc);
    by usubjid;
    if dd and cc and (dis_dt < dsstdtc);
run;

```

The data set bad_discon_before_comp will contain any problematic records with discontinuation date before completion date.

3. TABULATION DATA STRUCTURES READILY LEVERAGED FOR ADAM DATA SETUP

The SDTM model adapts a long and skinny data structure which can be readily leveraged for ADaM data setup. For example, different vital sign measures are usually collected horizontally on the same form at a given visit. The VS (Vital Signs) domain transposes the horizontal data into a vertical structure by defining different Vital Signs Test Short Name/Vital Signs Test Name, VSTESTCD/VSTEST values to each vital sign measurement. Each subject will have multiple rows with different VSTESTCD/VSTEST values in the VS domain. This feature is consistent with the ADaM structure and minimal programming effort is needed to setup ADaM-compliant Vital Sign data.

The example CRF below displays one mock subject's Vital Signs: blood pressure, pulse rate, respiratory rate, temperature, and body weight collected at Visit 2:

Visit 2 : VITAL SIGNS - Form Version: 09-Jun-2008 01:20	
Site: Patient: Patient No:	
Vital Signs	
1. Enter assessment date if different from visit date	<input type="text" value="7"/> / <input type="text" value="Feb"/> / <input type="text" value="2009"/>
2. Position for Blood Pressure and Pulse Rate	<input checked="" type="radio"/> sitting <input type="radio"/> semi-recumbent <input type="radio"/> standing <input type="radio"/> supine <input type="radio"/> prone
3. Blood Pressure (Systolic/Diastolic)	<input type="text" value="100"/> / <input type="text" value="80"/> mmHg
4. Pulse Rate	<input type="text" value="70"/> beats/min
5. Respiratory Rate	<input type="text" value="20"/> breaths/min
6. Temperature (XXX.X)	<input type="text" value="37"/> °C <input type="radio"/> °F
7. Temperature Method	<input checked="" type="radio"/> Oral <input type="radio"/> Tympanic <input type="radio"/> Rectal <input type="radio"/> Axillary
8. Weight	<input type="text" value="65"/> lb <input checked="" type="radio"/> kg

The data collected is stored horizontally as a single row in the front end collection application. However, in the VS domain, this single row record was transposed to multiple rows with different VSTESTCD/VSTEST values:

VSSEQ	VSTESTCD	VSTEST	VSPOS	VSORRES	VSORRESU	VSLOC	VISIT	VSDTC
1	SYSBP	Systolic Blood Pressure	SITTING	100	mmHg		Visit 2	2009-02-07
2	DIABP	Diastolic Blood Pressure	SITTING	80	mmHg		Visit 2	2009-02-07
3	PULSE	Pulse Rate	SITTING	70	beats/min		Visit 2	2009-02-07
4	RESP	Respiratory Rate		20	breaths/min		Visit 2	2009-02-07
5	TEMP	Temperature		37	C	ORAL	Visit 2	2009-02-07
6	WEIGHT	Weight		65	Kg		Visit 2	2009-02-07

In the ADaM data set, ADVS, these records were easily adapted from the VS domain above with traceable variables like SRCDOM, SRCVAR and SRCSEQ which trace back the records to the SDTM VS domain:

PARAM	PARAMCD	AVISIT	ADT	AVAL	BASE	CHG	SRCDOM	SRCVAR	SRCSEQ
Systolic Blood Pressure (mmHg)	SYSBP	Week 1	2/7/2009	100	90	10	VS	VSORRES	1
Diastolic Blood Pressure (mmHg)	DIABP	Week 1	2/7/2009	80	75	5	VS	VSORRES	2
Pulse Rate (beats/min)	PULSE	Week 1	2/7/2009	70	69	1	VS	VSORRES	3
Respiratory Rate (breaths/min)	RESP	Week 1	2/7/2009	20	22	-2	VS	VSORRES	4
Temperature (C)	TEMP	Week 1	2/7/2009	37	36.8	0.2	VS	VSORRES	5
Weight (kg)	WEIGHT	Week 1	2/7/2009	65	65	0	VS	VSORRES	6

CONCLUSION

The main purpose of the Study Data Tabulation Model is for the preparation of the study data tabulation for submission. The standard ensures consistency across studies within the industry and greatly facilitates data review for agency reviewers. This paper discussed some benefits to the industry of implementing SDTM up front in the product development process. Our experience with SDTM data in real time studies suggests that it not only increases the efficiency of data submission and data review, but also provides benefits in data management, data integrity checking, as well as data analysis and reporting.

REFERENCES

CDISC Study Data Tabulation Model Implementation Guide: Human Clinical Trials

Version 3.1.1

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.

ACKNOWLEDGEMENTS

The author would like to thank the management team for their review of this paper.

CONTACT INFORMATION

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

Changhong Shi

Merck Co. & Inc.

RY34-A320

P.O. Box 2000

Rahway, NJ 07065

(732)-594-1383

Changhong_shi@merck.com

Beilei Xu

Merck Co. & Inc.

RY34-A320

P.O. Box 2000

Rahway, NJ 07065

(732)-594-9980

beilei_xu@merck.com