The ADaM Solutions to Non-endpoints Analyses

Chengxin Li, Boehringer Ingelheim Pharmaceuticals Inc., Ridgefield, CT, USA

ABSTRACT

There always exist some analyses for non endpoints in the study. These analyses normally do not have inferential statistics, nor data/date imputations, nor timing windows, nor complicated derivations. With the example analysis on medical history based on SDTM MH domain, this paper describes different ADaM solutions to non-endpoints occurrence data: No ADaM Data Structure (NDS), ADaM Incidence Data Structure (IDS), or ADaM Basic Data Structure (BDS). Among the solutions, the NDS one is preferred in sense of simplicity and cost effective.

KEYWORDS

ADaM, Non-endpoints, Incidence Data

INTRODUCTION

The Clinical Data Interchange Standards Consortium (CDISC) [1] data models include data collection standard with Clinical Data Acquisition Standards Harmonization (CDASH) [1], tabulation data submission standard with Study Data Tabulation Model (SDTM) [2], and analysis dataset standard with Analysis Data Model (ADaM) [3]. To assist in the implementation, CDISC has been developing implementation guides such as Clinical Data Acquisition Standards Harmonization (CDASH) User Guide (CDASH UG) [4], Data Tabulation Model Implementation Guide (SDTM IG) [5], and Analysis Data Model (ADaM) Implementation Guide (ADaM IG) [6].

In current ADaM standard, there are two structures defined: the Subject-Level Analysis Dataset (ADSL) structure and the Basic Data Structure (BDS). In the meantime, the ADaM WGs have been working on "other" standards, e.g., recently published analysis data model for adverse event--the ADaM Data Structure for Adverse Event Analysis (ADAE) [7], analysis data model for time to event--the ADaM Basic Data Structure for Time-to-Event Analyses (ADTTE) [8]. Actually ADTTE can be treated as ADaM basic data structure (BDS) plus additional time-to-event variables.

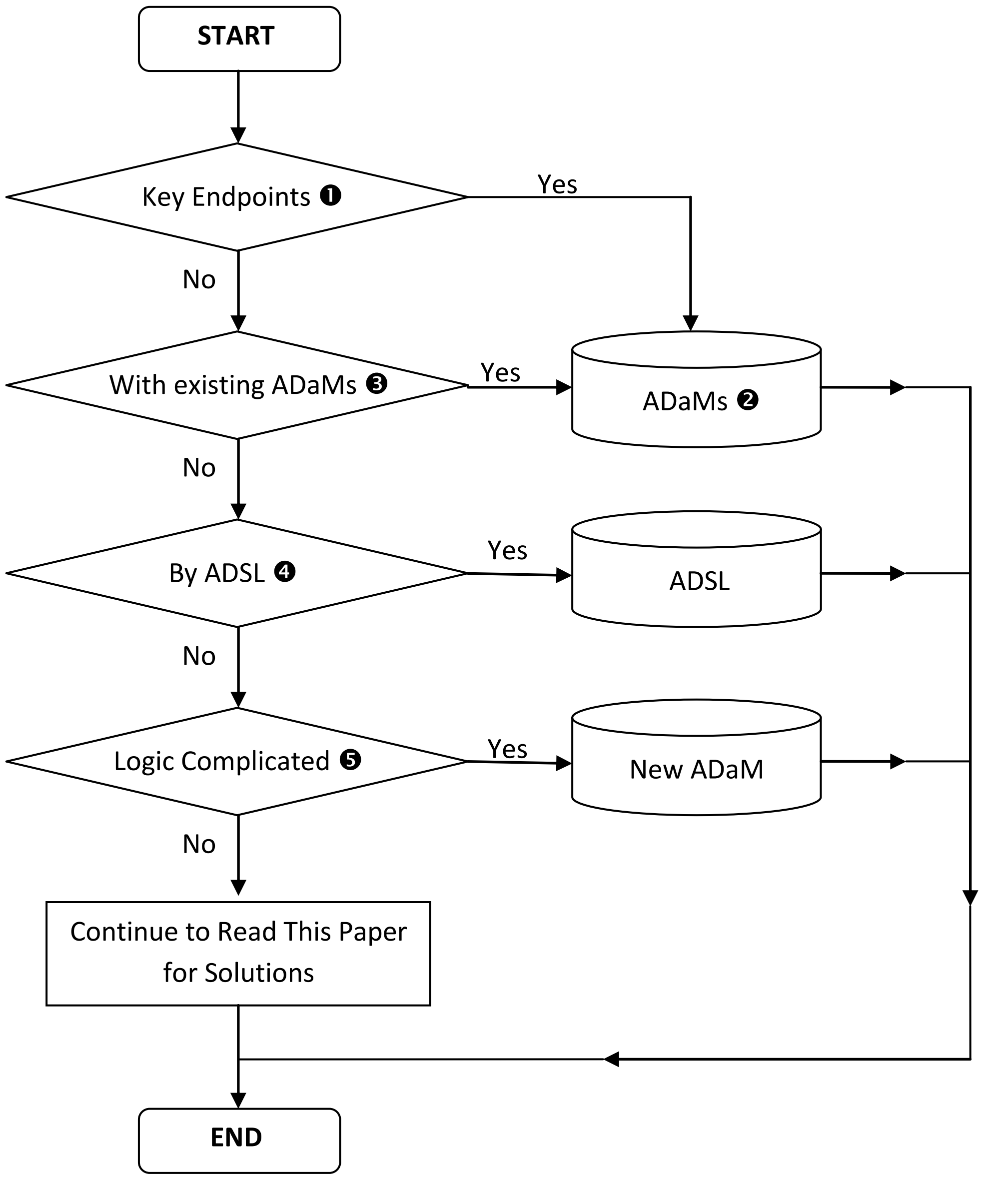
In clinical trials, the key endpoints, e.g., primary endpoints and secondary endpoints on efficacy analyses or safety analyses, are defined in the Statistical Analysis Plan (SAP). Those key endpoints should go with ADaM datasets such as ADaM BDS, ADTTE, or ADAE.

In Trial Subjects section, Efficacy Evaluations section, or Safety Evaluations section, there also exist some non-endpoints analyses. ADSL can cover some non-endpoints summary tables at Trial Subjects section of which the data are subject attributes not varying over the visits during the course of study, e.g., the demographic analyses and subject set analyses. At Efficacy Evaluation or Safety Evaluation sections,

1 http://www.cdisc.org

some non-endpoints summary tables may also be covered by further deriving from ADaM BDS datasets along with key endpoint analyses, e.g., by deriving a new PARAM from the same SDTM data sources.

The ADaM decision tree describing above procedures can be illustrated in figure 1.



❶defined in SAP, e.g., primary and secondary endpoints of efficacy and safety

❷including ADaM BDS, ADAE, ADTTE, etc.;

❸possibly covered by required ADaM datasets in ❷, e.g, by further deriving a new PARAM with the same SDTM data sources;

❹ the analyses related to subject attributes;

❺ e.g., data/date imputation, time windowing, or complex derivation

Figure 1 the General ADaM Decision Tree

This paper will address the ADaM solutions for those non-endpoints analyses which can't be covered in the ADSL or available ADaM datasets (BDS or ADAE, etc.). This paper will take MH as an example for a summary table.

CDISC PROGRAMING MODEL

Before providing solutions, it needs to first introduce the CDISC programming model illustrated in figure 2 below.

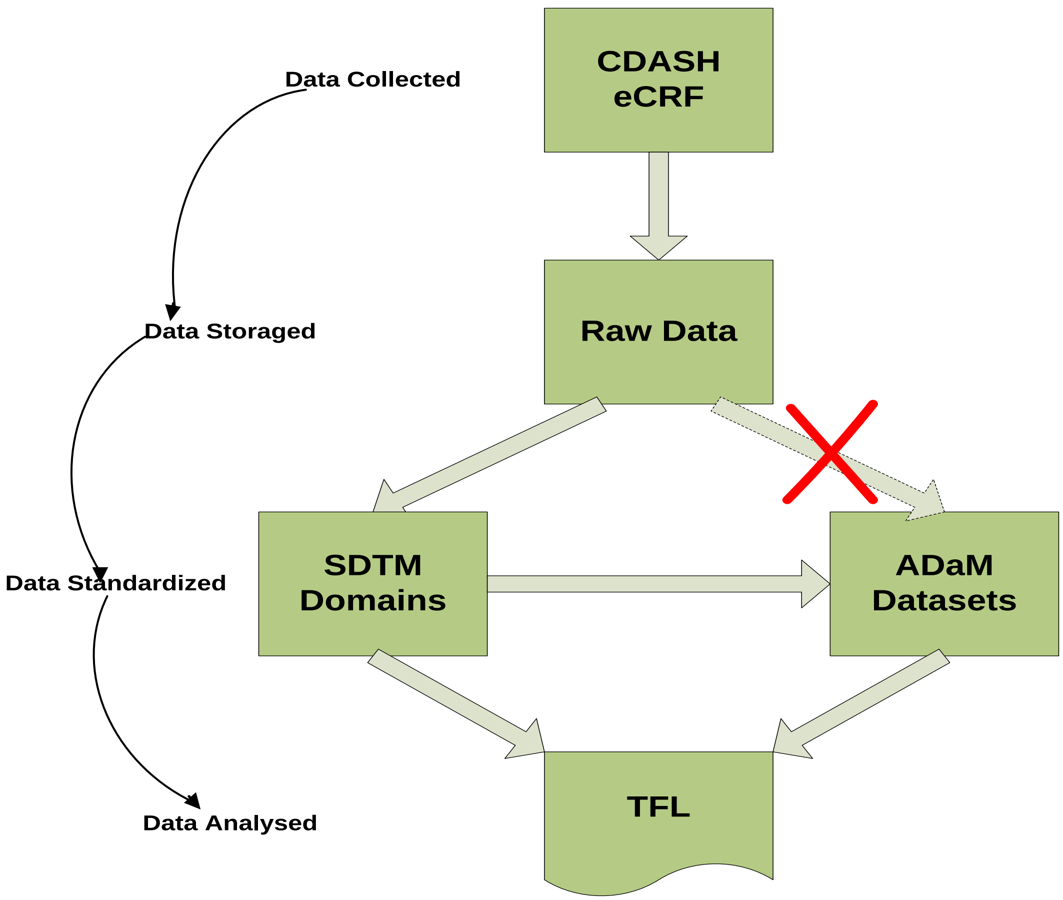


Figure 2 the CDISC Programming Model

The CDSIC data model requires that SDTM should fully reflect the collected data and the ADaM should only be derived from SDTM. The imputation for any missing data should be conducted only in ADaM rather than in SDTM. The key endpoint analyses, inferential analyses, or complicated analyses, in which cases the logic is not easy for reviewer to follow, should be designed with ADaM datasets. However, not every analysis needs to have a corresponding ADaM dataset. Some simple displays can be directly coded from SDTM domains such as data summary displays. To code these simple displays, neither derivation for timing widow, nor imputation for missing data, nor derivation with complicated algorithms is needed.

Within the CDISC programming model, the traceability should be facilitated wherever possible, from analysis result back to ADaM, to SDTM, and further back to collected raw data (or annotated Case Report Form).

The following sections will follow this programming model to design ADaM solutions to the specified non-endpoints analyses.

SIMULATED DATASET[2]

Suppose one clinical trial was designed as parallel, randomized, double blinded, and with two treatment groups. For simplicity, the actual treatment was supposed to be the same as planned treatment per subject. There was one summary table on patient medical history of prior and concomitant conditions (also called concomitant diagnoses) used as source data, i.e., the MH domain as the source data. The simulated partial data of MH domain with coding of MedDRA dictionary are depicted in figure 3, the simplified ADSL dataset simulated in figure 4, and display template for the summary table illustrated in figure 5.

STUDYID	DOMAIN	USUBJID	MHSEQ	MHTERM	MHDECOD	MHCAT	MHBODSYS	MHDTC
XXXX	MH	XXXX_0001	1	ASTHMA	Asthma	GENERAL MEDICAL HISTORY	Respiratory, thoracic and mediastinal disorders	2004-09-17
XXXX	MH	XXXX_0001	2	CEPHALEA	Headache	GENERAL MEDICAL HISTORY	Nervous system disorders	2004-09-17
XXXX	MH	XXXX_0001	3	CORONARY HEART DISEASE	Coronary artery disease	CARDIAC MEDICAL HISTORY	Cardiac disorders	2004-09-17
XXXX	MH	XXXX_0001	4	ATRIAL FIBRILLATION	Atrial fibrillation	CARDIAC MEDICAL HISTORY	Cardiac disorders	2004-09-17
……	……	……	……	……	……	……	……	……

Figure 3 the Simulated MH Domain

2 The simulated datasets are for serving this paper only.

STUDYID	USUBJID	SUBJID	SITEID	AGE	AGEU	SEX	RACE	FASFL	SAFFL	RANDFL
XXXX	XXXX_0001	1	XYZ001	48	YEARS	M	WHITE	Y	Y	Y
XXXX	XXXX_0002	2	XYZ006	58	YEARS	F	WHITE	Y	Y	Y
……	……	……	……	……	……	……	……	……	……	……

(Cont.)

ARM	TRT01P	RANDDT	TRTSDT	TRTEDT
Drug A	Drug A	2004-09-21	2004-09-21	2005-06-06
Drug B	Drug B	2004-09-23	2004-09-23	2005-06-08
……	……	……	……	……

Figure 4 the Simulated ADSL Dataset

	Drug A N (%)	Drug B N (%)
Number of patients	xxx (100.0)	xxx (100.0)
Patients with Concomitant Diagnoses	xxx (xx.x)	xxx (xx.x)
MHCAT	xxx (xx.x)	xxx (xx.x)
MHBODSYS	xxx (xx.x)	xxx (xx.x)
MHDECOD	xxx (xx.x)	xxx (xx.x)
……	……	……

Figure 5 the Simulated Display Template: Concomitant diagnoses – safety set

To work out this summary table depicted in figure 5, there may have several optional solutions. The following section will be providing different ADaM solutions for coding this table.

ADaM SOLUTION

OPTION 1 – No ADaM Data Structure (NDS)

As specified in the CDISC programming model, the analysis can be directly based on SDTM domain under the specific circumstances such as neither data imputation nor complicated derivation. For MH summary table, there are no interests in any data imputation, and the coding logic is simple enough for reviewer to follow. So the key parts of summary table can be coded directly from SDTM MH domain; the information for calculating denominator can be obtained from ADSL with SAFFL=’Y’; and the treatment groups can be dragged either from ADSL (or from DM domain). The solution is summarized in figure 6.

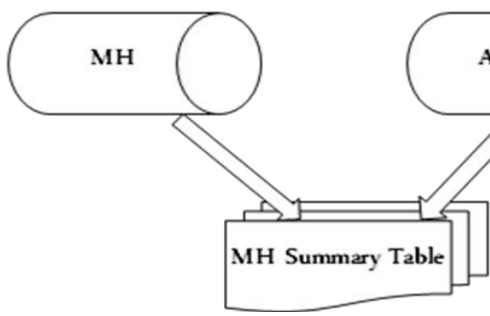


Figure 6 the Option of No ADaM Data Structure (NDS)

In more details, to work out this summary table (output displayed in figure 7), take the following steps:

- Count the denominators (Drug A, Drug B) from ADSL where ADSL.SAFFL='Y'
- Merge ADSL.ARM with SDTM MH domain as a temporary dataset
- Count categorized patient number from the above temporary dataset

	Drug A N (%)	Drug B N (%)
Number of patients	100 (100.0)	120 (100.0)
Patients with Concomitant Diagnoses	98 (98.0)	106 (88.3)
General Medical History	80 (80.0)	90 (75.0)
Respiratory, thoracic and mediastinal disorders	50 (50.0)	65 (54.2)
Asthma	50 (50.0)	65 (54.2)
Nervous system disorders	35 (35.0)	46 (38.3)
……	……	……

Figure 7 the Output of Concomitant Diagnoses

As this solution is very simple, from cost effective perspective, this solution should be highly preferred if the conditions of NDS are met. To completely ensure review easy, the logic illustrated in figure 6 may be further specified in Review's Guide.

OPTION 2 -- ADaM Incidence Data Structure (IDS)

Currently the CDISC ADaM sub-team is working on the general structure supporting analysis of incidence data such as concomitant medications, medical history, etc [3], in which cases the AVAL/AVAC or PARAM/PARAMCD would not be needed. The published ADAE is the first such example supporting analysis of incidence of adverse events.

From ADAE, for ADaM incidence data structure, there are no PARAM/PARAMCD nor AVAL/AVALC variables. The dataset is based on the corresponding SDTM domain plus some derived variables. The denominator used for the calculation of the percentages is obtained from ADSL. However, it's not clear yet how the final structure for the general incidence data looks like and when the standard will be published.

Mostly the analyses of adverse events are specified as safety endpoints in SAP, thus utilizing ADaM datasets (e.g., ADAE, ADTTE) is a good choice. However, the analyses of medical history are not defined as key endpoints, and neither date imputation nor indicators are of interests. Modeling from ADAE, the ADMH following IDS is illustrated in figure 8 (only necessary variables selected to ADMH).

STUDYID	USUBJID	TRTA	MHSEQ	MHTERM	MHDECOD	MHCAT	MHBODSYS	MHDTC
XXXX	XXXX_0001	Drug A	1	ASTHMA	Asthma	GENERAL MEDICAL HISTORY	Respiratory, thoracic and mediastinal disorders	2004-09-17
XXXX	XXXX_0001	Drug A	2	CEPHALEA	Headache	GENERAL MEDICAL HISTORY	Nervous system disorders	2004-09-17
XXXX	XXXX_0001	Drug A	3	CORONARY HEART DISEASE	Coronary artery disease	CARDIAC MEDICAL HISTORY	Cardiac disorders	2004-09-17
XXXX	XXXX_0001	Drug A	4	ATRIAL FIBRILLATION	Atrial fibrillation	CARDIAC MEDICAL HISTORY	Cardiac disorders	2004-09-17
……	……		……	……	……	……	……	……

Figure 8 the ADMH Dataset With IDS

The TRTA was directly copied from DM.ACTARM or can also be derived from ADSL.

---

[3] http://www.cdisc.org/adam-future

The solution is summarized in figure 9.

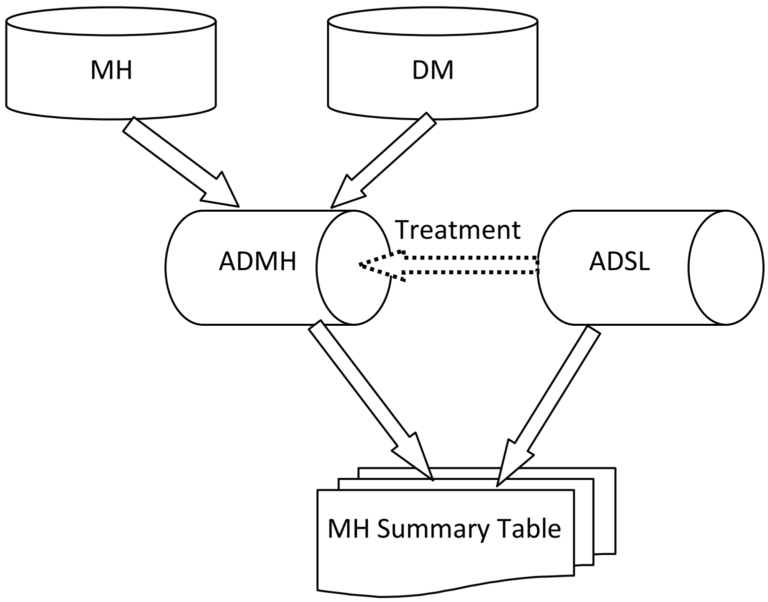


Figure 9 the Option of ADaM Incidence Data Structure (IDS)

To work out the summary table (figure 5), similar to option 1, take the following steps:

- Calculate the denominators (Drug A, Drug B) from ADSL where ADSL.SAFFL='Y'
- Count categorized patient number from ADMH

The output is the same as in figure 7.

Comparing ADMH with MH, the structures for SDTM MH and ADMH are very close. The coding processes for the summary table are similar as well. The IDS does not show any advantages over NDS for the analysis.

OPTION 3 -- ADaM Basic Data Structure (BDS)

Although the ADaM IG explicitly states that *the BDS was not designed to support analysis of incidence of adverse events or other occurrence data (p5, ADaMIG v1.0)*, it can still have a way to map those simple incidence data to ADaM BDS. The challenges lie in setting required AVAL/AVAL and PARAM/PARAMCD, and keeping the dictionary hierarchy in BDS if dictionary coding is used in the corresponding SDTM domain.

Compliant with ADaM BDS, the ADMH is illustrated in figure 10. It should be noted that MHDECOD was mapped to AVALC, and to keep the dictionary hierarchy as it is, MHCAT mapped to AVALCAT1, MHBODSYS mapped to AVALCAT2. ADT was derived from MH.MHDTC. PARAM was set as "Occurrence of Concomitant Diagnoses", PARAMCD as "MHOCC". Optionally, instead of mapping with AVALCATy, the variable MHCAT/MHBODSYS can be directly copied to ADMH. Different with IDS, a BDS should incorporate all needed variables into the dataset to have fully analysis ready. The Param-Level Flag SAFPFL is listed here to serve the purpose indicating the treated patient set are merged from ADSL by ADSL left join SDTM MH and with ADSL.SAFFL='Y'. The record of XXXX_1001 shows the case of patient 1001 in safety set but not in MH, which leads to redundancy in ADMH although the redundancy is at minimum.

STUDYID	USUBJID	TRTA	AVALC	AVALCAT1	AVALCAT2	ADT
XXXX	XXXX_0001	Drug A	Asthma	GENERAL MEDICAL HISTORY	Respiratory, thoracic and mediastinal disorders	2004-09-17
XXXX	XXXX_0001	Drug A	Headache	GENERAL MEDICAL HISTORY	Nervous system disorders	2004-09-17
XXXX	XXXX_0001	Drug A	Coronary artery disease	CARDIAC MEDICAL HISTORY	Cardiac disorders	2004-09-17
XXXX	XXXX_0001	Drug A	Atrial fibrillation	CARDIAC MEDICAL HISTORY	Cardiac disorders	2004-09-17
XXXX	XXXX_1001	Drug B				
……	……	……	……	……	……	……

(Cont.)

PARAM	PARAMCD	SAFPFL	SRCVAR	SRCSEQ
Occurrence of Concomitant Diagnoses	MHOCC	Y	MHDECOD	1
Occurrence of Concomitant Diagnoses	MHOCC	Y	MHDECOD	2
Occurrence of Concomitant Diagnoses	MHOCC	Y	MHDECOD	3
Occurrence of Concomitant Diagnoses	MHOCC	Y	MHDECOD	4
Occurrence of Concomitant Diagnoses	MHOCC	Y		
……	……	……	……	……

Figure 10 the ADMH Dataset With BDS

The solution is summarized in figure 11.

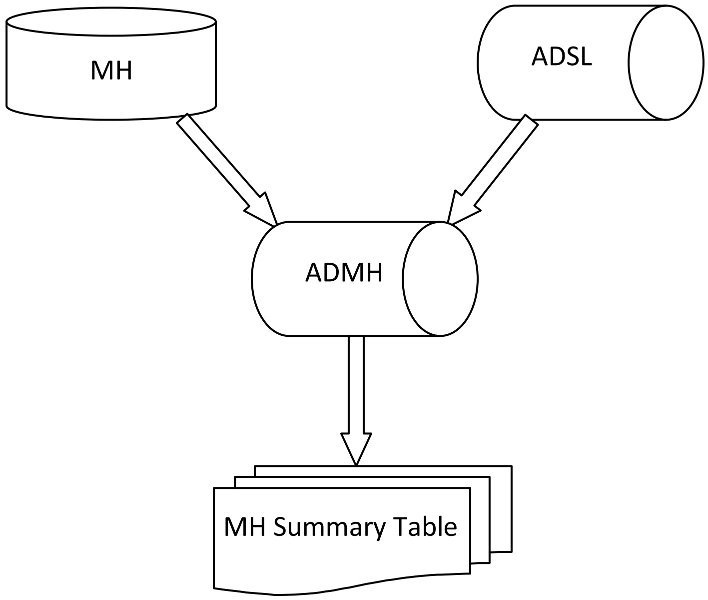


Figure 11 the Option of ADaM Basic Data Structure (BDS)

To work out the summary table (figure 5), take the following steps:

- Count distinct USUBJID as the denominators (Drug A, Drug B) from ADMH
- Count categorized patient number from ADMH dataset

The output is the same as in figure 7.

This option has more challenges in generating ADMH, but with fewer efforts in producing the table as the ADMH is equipped with fully analysis ready.

SUMMARY

The above ADaM solutions are summarized in figure 12. The NDS solution should be highly preferred to non-endpoint analysis if neither data imputations nor complicated derivations are needed.

Solution	Pros	Cons	Actions
NDS	No extra ADaM dataset, thus no define.xml; low cost; no redundancy	Requiring logic simple enough for reviewer to follow, e.g., just with one proc FREQ from the SDTM domain; table coding with SDTM + ADSL	Highly preferred under specific circumstance; put the coding logic in Review's Guide
IDS	no redundancy; dictionary hierarchy kept if dictionary coding used	Table coding with ADaM + ADSL; ADaM similar to SDTM; Restricted to incidence/occurrence data; define.xml needed for the dataset	Temporarily modeling from ADAE, otherwise wait until the standard of general structure for analysis of incidence data released
BDS	Fully analysis ready (all in one dataset)	Difficulty in producing ADaM dataset; redundancy; define.xml needed for the dataset	Avoided

Figure 12 the Summary of ADaM Solutions to Non-endpoint Analysis

CONCLUSION

There always exist some analyses for non endpoints in the study. These analyses normally do not need inferential statistics, nor data/date imputations, nor timing windows, nor complicated derivations. For these analyses especially for those occurrence data analyses, there may have several ADaM compliant solutions: No ADaM Data Structure (NDS), ADaM Incidence Data Structure (IDS), or ADaM Basic Data Structure (BDS). Among the solutions, the NDS should be preferred in sense of simplicity and cost effective. The example on MH analysis has fully demonstrated this.

ACKNOWLEDGEMENT

The author would like to thank Nancy Bauer for the invaluable inputs to this paper.

REFERENCE

[1] CDISC CDASH Team, the Clinical Data Acquisition Standards Harmonization (CDASH) v1.1, 18 January 2011.

[2] CDISC Submission Data Standards Team, the Study Data Tabulation Model v1.1, April 28, 2005.

[3] CDISC Analysis Data Model Team, the Analysis Data Model (ADaM) Version 2.1, December 17, 2009.

[4] CDISC CDASH Project Team, Clinical Data Acquisition Standards Harmonization (CDASH) User Guide, V1-1.1, 12 April 2012.

[5] CDISC Submission Data Standards Team, the Study Data Tabulation Model Implementation Guide: Human Clinical Trials V3.1.3, 6/25, 2012.

[6] CDISC Analysis Data Model (ADaM) Team, the Analysis Data Model (ADaM) Implementation Guide Version 1.0, December 17, 2009.

[7] CDISC Analysis Data Model (ADaM) Team, the Analysis Data Model (ADaM) Data Structure for Adverse Event Analysis Version 1.0, May 10, 2012.

[8] CDISC Analysis Data Model (ADaM) Team, the ADaM Basic Data Structure for Time-to-Event Analyses Version 1.0, May 8, 2012.

CONTACT INFORMATION

Author Name: Chengxin Li
Company: Boehringer Ingelheim
Address: 900 Ridgebury Road
City State Zip: Ridgefield, CT 06877
Email: chengxin.li@boehringer-ingelheim.com