

Hands-On ADaM ADAE Development

Sandra Minjoe, Accenture Life Sciences, Wayne, Pennsylvania

ABSTRACT

The Analysis Data Model (ADaM) Data Structure for Adverse Event Analysis¹ was released by the Clinical Data Interchange Standards Consortium (CDISC) ADaM team in May, 2012. This document is an appendix to the ADaM Implementation Guide (IG)¹, and describes the standard structure of the analysis dataset used for most of our typical adverse event reporting needs.

This hands-on training focuses on creating metadata for a typical adverse event (AE) dataset. Attendees will work with sample SDTM and ADaM data, finding information needed to create the results specified in a sample set of table mock-ups. Variable specifications, including coding algorithms, will be written. Some familiarity with SDTM data, AE reporting needs, SAS® data step programming, and Microsoft Excel® is expected.

Attendees will also learn how to apply the data structure for similar analyses other than adverse events.

INTRODUCTION: THE AE ANALYSIS NEED

Adverse event analysis results typically look something like this example table from the ADaM Data Structure for Adverse Event Analysis (ADAE) appendix¹:

Table 14.2.7.1		Page 1 of x	
Summary of Treatment Emergent Adverse Events by System Organ Class and Preferred Term Analysis Population: Safety			
SYSTEM ORGAN CLASS Preferred Term	Treatment A (N = xxx) n (%)	Treatment B (N = xxx) n (%)	
Number of subjects reporting at least one adverse event	x (x.x)	x (x.x)	
BLOOD AND LYMPHATIC SYSTEM DISORDERS			
At least one event	x (x.x)	x (x.x)	
Anaemia	x (x.x)	x (x.x)	
...	x (x.x)	x (x.x)	
CARDIAC DISORDERS			
At least one event	x (x.x)	x (x.x)	
Angina pectoris	x (x.x)	x (x.x)	
Coronary artery disease	x (x.x)	x (x.x)	
Ventricular tachycardia	x (x.x)	x (x.x)	
Myocardial infarction	x (x.x)	x (x.x)	
Ventricular fibrillation	x (x.x)	x (x.x)	
...	x (x.x)	x (x.x)	
<Other SOCs and PTs>			

N = Safety subjects, i.e., subjects who received at least one dose of study drug
n = Number of subjects reporting at least one treatment emergent adverse event
% = $n / N * 100$
Adverse events are presented by descending frequency within Treatment B
System organ classes and preferred terms are coded using MedDRA version x.x.

Figure 1: Example AE Analysis Table Output

To allow similar terms (e.g., headache, sinus headache, ache in head) to all be grouped together for reporting, each adverse event record must be mapped to a hierarchical dictionary. Typically the industry uses the Medical Dictionary for Regulatory Activities (MedDRA)² for AE mapping. MedDRA has the following levels of hierarchy: System Organ Class (SOC), High-Level Group Terms (HLGT), High-Level Terms (HLT), Preferred Terms (PT) and Lower-Level Terms (LLT). Dictionary mapping is typically done within the AE domain of the CDISC Standard Data Tabulation Model (SDTM), and this domain contains variables to hold the MedDRA mapping.

Usually two levels of the hierarchy are included in a summary, sometimes three. Within each level of hierarchy, subjects with any event are counted. Note that on standard tables the events themselves are not counted, just subjects. This means that a subject is counted once whether he had 17 headaches or just one. This type of counting is done at each level of the hierarchy reported.

ADaM has developed a standard analysis dataset as the step between the SDTM AE domain data and the analysis table. Sometimes analysis is very straightforward, and we might be able to use the AE domain along with some Subject Level Analysis Data (ADSL) tacked on. Often we need to create an analysis dataset, typically called ADAE, to handle any data imputations. The ADAE appendix describes some of these imputation needs¹.

THE AE ANALYSIS DATASET STRUCTURE

As described in the ADaM documents¹, analysis datasets are developed to address analysis needs. At this time there are three ADaM classes or data structures that are standard: ADSL (one record per subject); BDS (one record per subject, per analysis per parameter, and if needed per timepoint); and ADAE. The ADAE structure is not the same as or even based on the ADSL and BDS structures.

Adverse event data doesn't fit well into the BDS structure for a three main reasons, as described in the ADAE appendix¹:

- There is no need for AVAL or AVALC. Occurrences are counted in analysis, and there are typically one or more records for each occurrence.
- A dictionary is used for coding the occurrence, and it includes a well-structured hierarchy of categories and terminology. Mapping this hierarchy to BDS variables PARAM and generic *CAT variables would lose the structure and meaning of the dictionary.
- Dictionary content is typically not modified for analysis purposes. In other words, there is no need for analysis versions of the dictionary hierarchy.

Instead of forcing adverse event data into a BDS structure, the ADAE structure was developed.

Actually, the ADAE structure is based closely on the AE domain from SDTM. Typically ADAE and AE have the same number of rows, though some exceptions are explained the ADAE appendix¹. Any columns from AE and the transposed SUPPAE can be included in ADAE. In addition, variables from ADSL, including population flags, demographics, and baseline characteristics, can be added. Finally, we often need to derive some variables for analysis purposes, such as a treatment-emergent flag.

The list of the most common ADAE variables is included in section 4 of the ADAE appendix¹. These are not the only allowed variables in the dataset, and other variables can certainly be added to support analysis and reviewer needs. Additionally, many of the variables listed in the ADAE appendix are permissible, so an actual adverse event analysis dataset can have many fewer columns than are included in the document.

APPLICATION: CREATING AN AE ANALYSIS DATASET

The process of a creating an AE analysis dataset usually goes something like this:

1. Determine the analysis needs. Table mockups are generally a good source of information for this.
2. Determine all the variables that can be copied from input datasets, such as AE, SUPPAE, and ADSL. Any variables needed for analysis that are not copied from input datasets must be derived.
3. Bring in the SDTM AE dataset, and keep the variables needed for analysis. In some cases rows are subset or extra rows are created, but this is not typical.
4. If needed, transpose the SUPPAE dataset, using QNAM for the variable name and QLABEL for the variable label, and then merge any needed data by the appropriate keys.
5. Merge on ADSL variables needed for analysis.
6. Derive any additional analysis variables needed for analysis.

In the Hands-On Training exercise (for which this paper was written), we step through the process using the following example.

ANALYSIS NEEDS

This is an example table layout that we need to produce:

Figure 2: Example AE Analysis Table Mockup

In addition to this table in Figure 2, several other AE tables need to be produced. Below is a table of contents showing the list of all the AE tables that are needed for this example study.

Table 1: List of AE Analysis Tables to Produce

Table #	Description
Table 15.3.1.1	Frequency of Subjects with Adverse Events by Treatment, Primary System Organ Class and Preferred Term Treated Population
Table 15.3.1.2	Frequency of Subjects with Serious Adverse Events by Treatment, Primary System Organ Class and Preferred Term Treated Population
Table 15.3.1.3	Frequency of Subjects with Severe or Life-Threatening Adverse Events by Treatment, Primary System Organ Class and Preferred Term Treated Population
Table 15.3.1.4	Frequency of Subjects with Investigator-Defined Drug-Related Adverse Events by Treatment, Primary System Organ Class and Preferred Term Treated Population
Table 15.3.1.5	Frequency of Subjects with Adverse Events Leading to Discontinuation of Study Drug by Treatment, Primary System Organ Class and Preferred Term Treated Population
Table 15.3.2.1	Frequency of Subjects with Adverse Events by Treatment, Primary System Organ Class and Preferred Term Per-Protocol Population
Table 15.3.2.2	Frequency of Subjects with Serious Adverse Events by Treatment, Primary System Organ Class and Preferred Term Per-Protocol Population
Table 15.3.2.3	Frequency of Subjects with Severe or Life-Threatening Adverse Events by Treatment, Primary System Organ Class and Preferred Term Per-Protocol Population
Table 15.3.2.4	Frequency of Subjects with Investigator-Defined Drug-Related Adverse Events by Treatment, Primary System Organ Class and Preferred Term Per-Protocol Population
Table 15.3.2.5	Frequency of Subjects with Adverse Events Leading to Discontinuation of Study Drug by Treatment, Primary System Organ Class and Preferred Term Per-Protocol Population

Each of these tables will be structured similarly to the table mocked in Figure 2, but with different content as described in the title. For example, notice that all tables numbered 15.3.2.x have the same titles as 15.3.1.x, but for the Per-Protocol population rather than the Treated population.

INPUT DATA

Because full SAS datasets can't be easily included in a conference paper such as this, below is some metadata for each of the three input datasets used in the example: AE, SUPPAE, and ADSL.

Table 2: Selected Metadata for Input Dataset AE

NAME	LABEL
STUDYID	Study Identifier
DOMAIN	Domain Abbreviation
USUBJID	Unique Subject Identifier
AESEQ	Sequence Number
AETERM	Reported Term for the Adverse Event
AEDECOD	Dictionary-Derived Term
AEBODSYS	Body System or Organ Class
AESER	Serious Event
AEACN	Action Taken with Study Treatment
AEREL	Causality
AEOUT	Outcome of Adverse Event
AESDTH	Results in Death
AESLIFE	Is Life Threatening
AECONTRT	Concomitant or Additional Trtmnt Given
AETOXGR	Standard Toxicity Grade
VISITNUM	Visit Number
VISIT	Visit Name
EPOCH	Epoch
ETCD	Element Code
ELEMENT	Description of Element
AESTDTC	Start Date/Time of Adverse Event
AEENDTC	End Date/Time of Adverse Event
AESTDY	Study Day of Start of Adverse Event
AEENDY	Study Day of End of Adverse Event

Table 3: Selected Metadata for Input Dataset SUPPAE

QNAM	QLABEL
ACTNSMED	Action taken - Study Medication
ACTNTMZ	Action taken – TMZ
AELLT	Low Level Term
AELLTCD	Low Level Term Code
AEPTCD	Preferred Term Code
RELDX	Causal Relationship to Disease
RELSMED	Causal Relationship - Study Medication
RELINF	Relative to Infusion

QNAM	QLABEL
RELTMZ	Causal Relationship to TMZ

Table 4: Selected Metadata for Input Dataset ADSL

NAME	LABEL
STUDYID	Study Identifier
USUBJID	Unique Subject Identifier
SUBJID	Subject Identifier for the Study
SITEID	Study Site Identifier
ARMCD	Planned Arm Code
ARM	Description of Planned Arm
AGE	Age
AGEU	Age Units
AGEGR1	Pooled Age Group 1
AGEGR1N	Pooled Age Group 1 (N)
SEX	Sex
RACE	Race
COUNTRY	Country
TRT01P	Planned Treatment for Period 01
TRTSDTM	Datetime of First Exposure to Treatment
TRTSDT	Date of First Exposure to Treatment
TRTEDTM	Datetime of Last Exposure to Treatment
TRTEDT	Date of Last Exposure to Treatment
ENRFL	Enrolled Population Flag
SAFFL	Safety Population Flag
ITTFL	Intent-To-Treat Population Flag
DEATHDT	Date of Death

COMBINING INPUT DATASETS

As described earlier in this paper, the typical process is to:

- Pull in the AE dataset, keeping all variables needed for analysis
- Transpose and then merge on the SUPPAE data needed for analysis
- Merge on the ADSL data needed for analysis

Not all variables from each input dataset need to be included. Some companies choose to keep only the bare minimum that are needed for analysis and traceability; other companies choose to keep all variables from each input dataset. Most companies are somewhere in between.

The point of the paper is not to walk through SAS code to do these steps. There are many papers and other documentation available to describe how to subset, transpose, and merge.

DERIVATIONS NEEDED

Continuing our example, the Statistical Analysis Plan (SAP) specifies that only treatment-emergent events are to be included, so this variable must be derived. There is a treatment-emergent flag on the SDTM AE dataset, however:

- Imputation is not allowed in SDTM, so any records with partial dates probably don't have a treatment-emergent flag in SDTM.
- Study days are created based on the study reference start date assigned in SDTM. If the SDTM study day is what is used to create the SDTM treatment-emergent flag, it is probably not be what is needed for our analysis.

For these reasons, it is common to derive the analysis treatment-emergent flag in ADaM rather than use the SDTM variable.

Also described in the SAP are a handful of imputations that must be done when data is missing:

- If *Serious* is unknown, impute to Serious
- If *Drug-Related* is unknown, impute to Related
- If *Leading to Discontinuation of Study Drug* is missing, impute to Yes

Note that for each of these imputations we must derive a new variable, and leave the variable copied from SDTM unchanged. This is based on the harmonization principle, described in the ADaM document version 2.1 as the requirement to “maintain the values and attributes of SDTM variables if copied into analysis datasets without renaming”¹. The new variable name should be based on the SDTM variable, replacing the “AE” prefix with an “A” for analysis, as described in the ADAE appendix¹. This means we’ll have a variable AESER with exactly the same content it had in SDTM, plus a new derived variable ASER where any values missing from AESER are imputed to “Serious”. This same type of imputation derivation must be done for the analysis versions of variables for drug-relatedness and leading-to-discontinuation.

Again, the point of the paper is not to walk through SAS code to do these steps. There are many papers and other documentation available to describe how to derive variables.

PRODUCING RESULTS

The adverse event structure described in the ADAE appendix allows us to easily derive the numbers on our adverse event analysis tables. Each full table will probably not be produced with one statistical procedure, but each count can be done via one procedure. This allows for traceability back from each number on the table to variables and rows in the dataset.

We may need occurrence flags in the dataset to denote a single event per subject to count at each reported level of hierarchy, however many companies have standard reporting tools that do this for them.

EXTENSION: USING THE AE ANALYSIS DATASET STRUCTURE FOR OTHER DATA

Concomitant medications and medical history data that is mapped to a hierarchy and summarized by that hierarchy, similar to what was described above for adverse events, does not currently fit into any ADaM structures. The ADAE appendix reads “Adverse events are just one example of data that can use the structure described within this document. An ADaM sub-team is working to expand this to other data where there is no need for an analysis variable or parameter as would be seen in a BDS structure because records are simply counted for analysis. Example data for these types of analyses are concomitant medications and medical history.”¹

As of this writing, there is an ADaM sub-team developing a document describing this structure. As suggested, this document will include applications not only for adverse events, but also for concomitant medications and medical history. This document is in late draft status, and is soon expected to be published for public comment. After comments are reviewed and any changes made, the final document, which should contain a single structure for all of these hierarchical occurrence needs, will be posted to the CDISC website.

Until the time that this data structure is final, we can apply some of the rules seen in the ADAE appendix to help us with our concomitant medications and medical history needs. For example, we can create the concomitant medication analysis dataset by following these steps, which you’ll note are remarkably similar to those stated earlier in the document for developing an adverse event analysis dataset:

1. Determine the analysis needs. Table mockups are generally a good source of information for this.
2. Determine all the variables that can be copied from input datasets, such as CM, SUPPCM, and ADSL. Any variables needed for analysis that are not copied from input datasets must be derived.
3. Bring in the SDTM CM dataset, and keep the variables needed for analysis. In some cases rows are subset or extra rows are created, but this is not typical.
4. If needed, transpose the SUPPCM dataset, using QNAM for the variable name and QLABEL for the variable label, and then merge any needed data by the appropriate keys.
5. Merge on ADSL variables needed for analysis.
6. Derive any additional analysis variables needed for analysis.

We can also use the naming convention rules in the ADAE appendix for new variables derived in our analysis dataset, imputed from SDTM variables. In the case of concomitant medications, we’d replace the 2-letter domain prefix “CM” with “A” for analysis.

One thing to keep in mind is that when developing our metadata for concomitant medication and medical history analysis data, we can’t use a class/structure of “ADAE”. Instead, until the new structure is defined, we will have to use the class structure of “OTHER”.

For further information on how to apply the ADAE appendix to concomitant medications and medical history data,

consider reading the paper titled “Using the ADaM ADAE Structure for non-AE Data” written for the 2013 SAS Global Forum³.

CONCLUSION

It is a pretty straightforward process to assemble an adverse event analysis dataset from SDTM and ADSL data to produce the common adverse events analysis tables. The resulting AE analysis dataset generally has the same number of records as the SDTM AE domain, and is usually assembled by simply combining the SDTM AE domain + any SUPPAE variables after transposition + any ADSL variables + any derived variables. In fact, many people refer to it as an “SDTM-plus” structure.

This structure is different from BDS in that it doesn’t have analysis values or parameters. However, variable naming conventions and rules for adding rows are similar to BDS, so it can be helpful to reference the ADaMIG in addition to the ADAE appendix document when building the adverse event analysis dataset.

Although not yet developed, we can apply the concepts in the ADAE appendix to produce similarly structured datasets for common concomitant medication and medical history analyses. As long as the analysis need is to count subjects within different levels of a standard hierarchy, and that hierarchy isn’t modified from how it is captured in SDTM, this structure works nicely.

In all cases, the analysis dataset is derived to support the needed analyses. If your analysis needs are different from those described in this document, you may need to create data in a different structure.

REFERENCES

¹All CDISC documents are available from: <http://www.cdisc.org>

²MedDRA Terminology is available at http://www.meddrasso.com/subscriber_download.asp

³Minjoe, S. and Widel, M. Paper 177-2013 “Using the ADaM ADAE Structure for non-AE Data” for the 2013 SAS Global Forum will be located at <http://support.sas.com/events/sasglobalforum/2013/index.html>. At the time of this writing the paper had not yet been uploaded to the conference website, so a direct link can’t be provided.

ACKNOWLEDGMENTS

The author would like to thank the CDISC organization, specifically the ADaM team and its ADAE and Occurrence sub-teams, for developing the standards described in this document.

CONTACT INFORMATION

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

Sandra Minjoe
Accenture Life Sciences
585 East Swedesford Road
Wayne, PA 19087
Sandra.Minjoe@accenture.com

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.

Other brand and product names are trademarks of their respective companies.