PharmaSUG 2025 - Paper DS-028

Name That ADaM Dataset Structure

Nancy Brucken, IQVIA

ABSTRACT

This presentation will give a brief overview of the standard ADaM dataset classes, followed by an interactive segment where a commonly-used table shell will be displayed, and the audience will be encouraged to name the standard ADaM dataset class and variables that could be used to most easily generate that table. Alternative approaches will be discussed where appropriate, including the use of the ADaM Other class and when it is acceptable to produce a single table from multiple ADaM datasets. The goal is to get the audience thinking about designing ADaM datasets based on table requirements, instead of on the structure of the SDTM domains feeding into those datasets.

INTRODUCTION

The Analysis Dataset Model (ADaM) and ADaM Implementation Guide (ADaMIG) have defined 3 standard dataset structures, which support many common clinical trial data analysis methods:

- Subject-Level Analysis Dataset (ADSL) – a one record per subject dataset, regardless of study design, containing variables that are important for describing the subject and their experiences during the trial.
- Basic Data Structure (BDS) – a dataset containing one or more records per subject, per analysis parameter, and possibly per analysis timepoint.
- Occurrences Data Structure (OCCDS) – a dataset containing one or more records per subject and topic (event, intervention), typically used for counting such records.

The choice of which dataset structure to follow for a given dataset should be driven by the types of analyses that will be performed using that data. Note that, while standard dataset structures have also been defined for medical devices, those are outside the scope of this paper.

ADAM DATASET CLASSES AND SUBCLASSES

In addition to the 3 standard ADaM dataset classes, several subclasses have also been defined, in order to facilitate additional standardization and the development of more detailed conformance checks. Within the BDS class, the following subclasses have been defined:

- Time-to-Event (TTE) – a dataset designed to support survival analysis.
- Non-Compartmental Analysis (NCA) – a dataset designed to support non-compartmental analysis of pharmacokinetic (PK) data.
- Population PK (POPPK) – a dataset designed to support population PK analysis.

Within the OCCDS class, the following subclass has been defined:

- Adverse Events (ADAE) – a dataset consisting of adverse event (AE) records for each subject.

The ADaMIG has also defined a fourth analysis dataset structure, with a class of ADAM OTHER, to handle data which follow the fundamental principles of ADaM and basic ADaM concepts, but do not fit into one of the three standard dataset structures.

SUBJECT-LEVEL ANALYSIS DATASET (ADSL)

ADSL always consists of one record per subject, regardless of study design, and is the only required ADaM dataset. It contains important subject-level information, such as analysis population flags, treatment arms, demographic variables, baseline values, stratification variables, important dates, and subgroups.

BASIC DATA STRUCTURE (BDS)

BDS datasets consist of one record per subject, per parameter, and optionally, per analysis timepoint, and are considered to be “vertical” rather than “horizontal” datasets. That means they contain considerably few variables, but many records. BDS is the most commonly used ADaM dataset class, and supports a wide variety of analyses, including univariate summary statistics of values by timepoint, logistic regression, and repeated measures. The BDS also provides datapoint traceability via specific variables that point back to the source of the analysis value.

TIME-TO-EVENT SUBCLASS

The Time-to-Event subclass is a specific version of a BDS, structured as one record per parameter; analysis timepoints are not generally included. It is designed to support survival analysis, and contains an analysis value which represents either the time to the event or the time until a subject was censored, along with a censoring indicator and reason for censoring.

OTHER BDS SUBCLASSES

Other standard BDS subclasses include Non-Compartmental Analysis (NCA), which is used as input to software packages performing non-compartmental PK analyses on drug concentration data, and Population PK (POPPK), which is used as input to software packages performing population PK analyses.

OCCURRENCES DATA STRUCTURE (OCCDS)

OCCDS datasets are typically structured as one record per subject and term, where a “term” could be an event, a medication, or anything else that is being counted. They support analyses that involve counting occurrences of something related to that term, so there is no need for an analysis value on the record, and each record in an OCCDS dataset often corresponds directly to a single record in the source SDTM dataset.

ADVERSE EVENTS SUBCLASS

The Adverse Events (AE) subclass is a special subset of OCCDS records, limited to adverse events only, and is structured as one record per subject and adverse event. Since there is no requirement that an analysis dataset consisting of AE records has to be named ADAE, designating such a dataset as belonging to the AE subclass allows for the execution of additional conformance checks to ensure that required MedDRA dictionary coding fields, a serious AE indicator, and other required AE variables are present and populated in the dataset.

LET’S PLAY NAME THAT ADAM DATASET STRUCTURE!

Enough of the background information and preliminaries. It’s time to play “Name that ADaM Dataset Structure”! The rest of this paper will consist of a sample annotated table shell, with an explanation of why the specific ADaM dataset structure and associated variables were chosen to support it. Hands on your buzzers, and here we go… !

TABLE 1

Table 1

ADSL

SAFFL

TRT01P

AGE

SEX

ETHNIC

RACE

14.1.2.1 Subject Demographics and Baseline Characteristics
Safety Population

	BP3304 (N =xx)	Placebo (N =xx)	Overall (N=xx)
Age (years)			
N	xx	xx	xx
Mean (SD)	xx.x (xx.xx)	xx.x (xx.xx)	xx.x (xx.xx)
Median	xx.x	xx.x	xx.x
Min, Max	xx, xx	xx, xx	xx, xx
Gender [n (%)][a]			
Male	xx (xx.x)	xx (xx.x)	xx (xx.x)
Female	xx (xx.x)	xx (xx.x)	xx (xx.x)
Ethnicity [n (%)][a]			
Hispanic or Latino	xx (xx.x)	xx (xx.x)	xx (xx.x)
Not Hispanic or Latino	xx (xx.x)	xx (xx.x)	xx (xx.x)
Race [n (%)][a]			
White	xx (xx.x)	xx (xx.x)	xx (xx.x)
Black or African American	xx (xx.x)	xx (xx.x)	xx (xx.x)
Asian	xx (xx.x)	xx (xx.x)	xx (xx.x)
American Indian or Alaskan Native	xx (xx.x)	xx (xx.x)	xx (xx.x)
Native Hawaiian or Other Pacific Islander	xx (xx.x)	xx (xx.x)	xx (xx.x)
Other	xx (xx.x)	xx (xx.x)	xx (xx.x)

Reference: Listing 16.2.4.1
[a] Percentages are based on the number of subjects in the population.

Since this table summarizes basic subject demographic information, which is collected only once during the study, ADSL offers the best fit. Variables AGE, RACE and SEX are required to be present in ADSL by the ADaMIG, and you'll also need to include the safety population flag, SAFFL, subject ethnicity (ETHNIC), and at least one treatment variable in order to produce the table. For this example, we chose to summarize by planned treatment (TRT01P), but depending on the Statistical Analysis Plan (SAP), actual treatment (TRT01A) might be appropriate, as well.

TABLE 2

Table 2

ADSL

SAFFL

TRT01P

HEIGHTBL

WEIGHTBL

BMIBL

TRTSDT

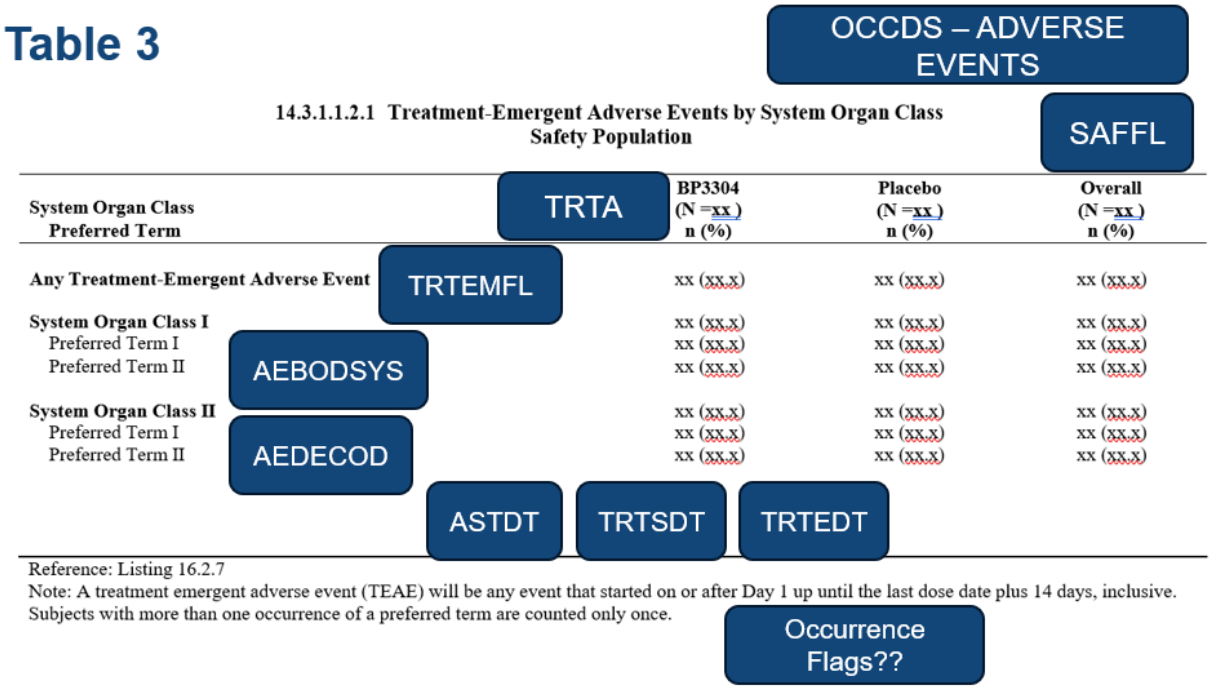
14.1.2.1 Subject Demographics and Baseline Characteristics
Safety Population

	BP3304 (N =xx)	Placebo (N =xx)	Overall (N=xx)
Height (cm)			
N	xx	xx	xx
Mean (SD)	xx.x (xx.xx)	xx.x (xx.xx)	xx.x (xx.xx)
Median	xx.x	xx.x	xx.x
Min, Max	xx, xx	xx, xx	xx, xx
Weight (kg)			
N	xx	xx	xx
Mean (SD)	xx.x (xx.xx)	xx.x (xx.xx)	xx.x (xx.xx)
Median	xx.x	xx.x	xx.x
Min, Max	xx, xx	xx, xx	xx, xx
Body Mass Index (kg/m^2)			
N	xx	xx	xx
Mean (SD)	xx.x (xx.xx)	xx.x (xx.xx)	xx.x (xx.xx)
Median	xx.x	xx.x	xx.x
Min, Max	xx, xx	xx, xx	xx, xx

Reference: Listing 16.2.4.1
Note: SD = standard deviation, Min = Minimum, Max = Maximum.

Since ADSL also contains important baseline characteristics for subjects in the study, it would be a good choice for producing this table, which displays univariate summary statistics for baseline height (HEIGHTBL), weight (WEIGHTBL) and body mass index (BMIBL). Note that there are no standard names for the baseline vital signs variables, but 'BL' is an ADaM variable naming fragment representing a baseline value, and should be used as a suffix in the variable name, according to the ADaMIG. Also, since the baseline value is often defined as the last non-missing value before the first dose of study medication, TRTSDT may be helpful when determining the baseline value. It's also a required ADaM variable. And again, at least one analysis population flag and treatment variable are required in ADSL.

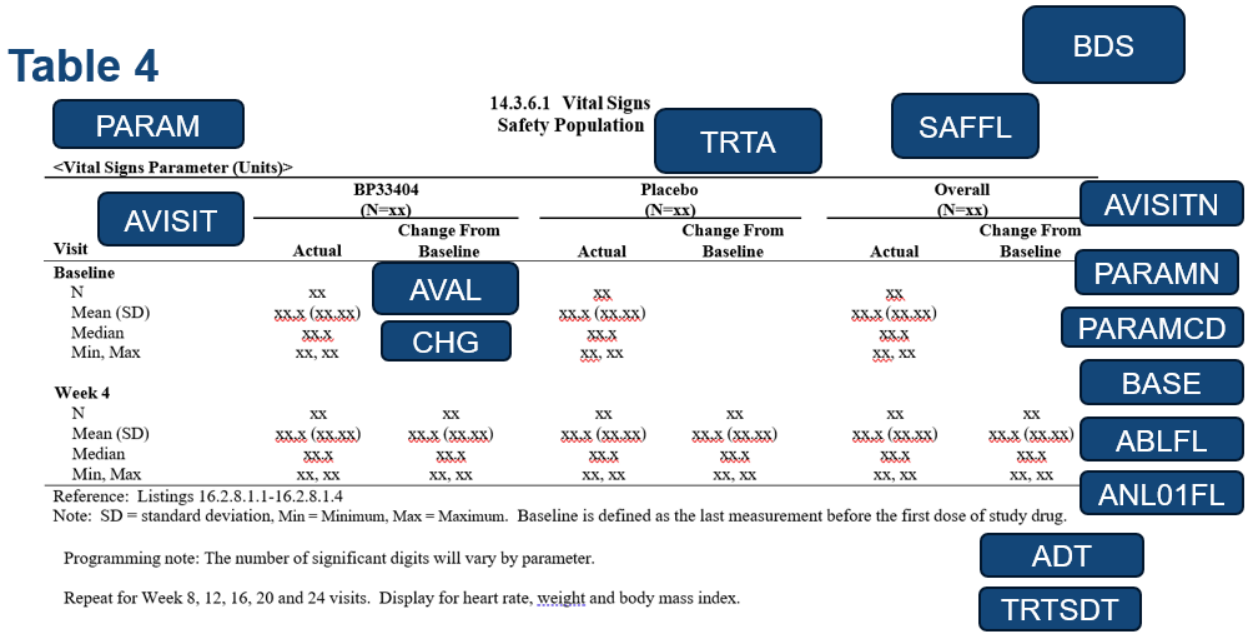
TABLE 3



This table displays counts of subjects in the safety population (SAFFL) who reported AEs by System Organ Class (AEBODSYS) and Preferred Term (AEDECOD). Since we're simply counting subjects with AE records, there is no analysis value defined, and the OCCDS structure, with a designated subclass of ADVERSE EVENTS, should be used. Note that since only treatment-emergent AEs (TEAEs) should be included, the required variable TRTEMFL is needed. Additionally, since the footnote defines treatment-emergent as any event that started on or after the first dose of study med up until the last dose date plus 14 days, both the treatment start date (TRTSDT) and treatment end date (TRTEDT) must be included in this dataset, along with the adverse event start date (ASTDT), for traceability when deriving TRTEMFL.

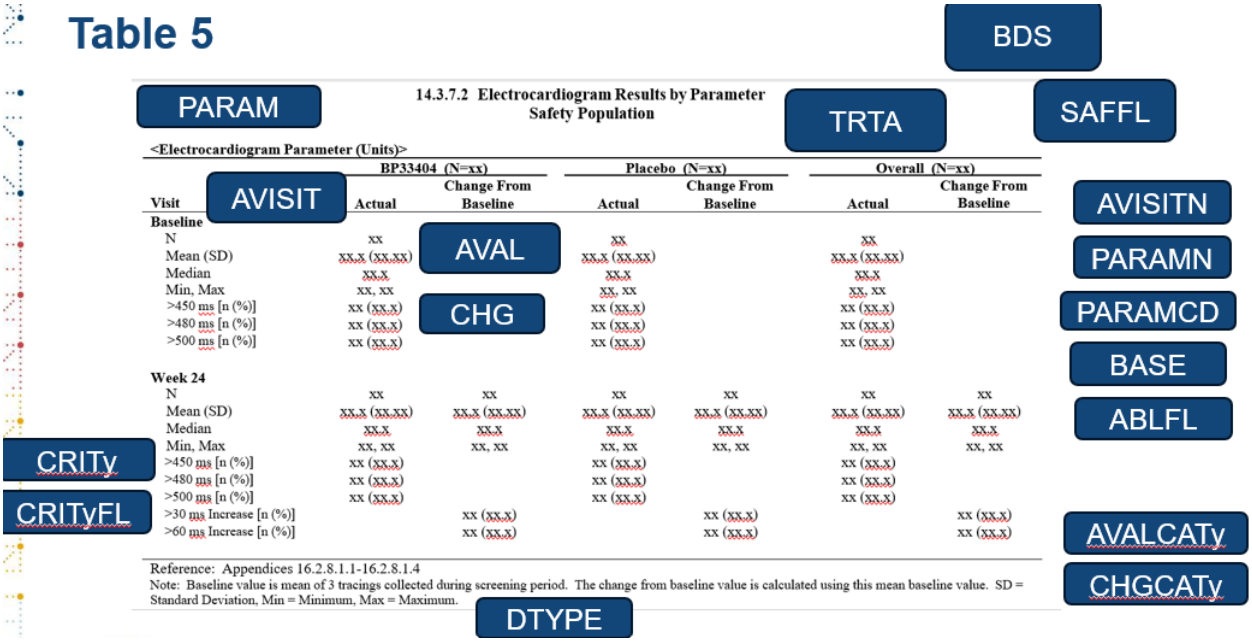
One question that often comes up when designing datasets to support AE tables is whether to include occurrence flags (AOCCzzFL and related variables). The ADaM team was never able to reach consensus on the use of occurrence flags, so they remain as permissible variables within the OCCDS structure. They can be extremely helpful, especially in cases where the selection criteria are complex, such as in the identification of records representing the maximum severity of an AE. However, since AE tables generally summarize events at the subject, SOC and PT levels, and each level requires a separate occurrence flag, the number of occurrence flags in the analysis dataset can multiply rapidly, to the point where they reach an unmanageable number.

TABLE 4



This table displays univariate summary statistics for vital signs and change from baseline over time. Since there is a clearly defined analysis value (AVAL), the vital sign measurement, the BDS is the best standard structure for supporting the table. Change from baseline is also summarized, so CHG should be included, and in order to calculate change from baseline, the baseline value must first be identified (ABLFL), and then added to every record for that parameter (BASE). In addition, the footnote indicates that baseline is defined as the last measurement before the first dose of study drug, so the vital signs measurement date (ADT) and treatment start date (TRTSDT) must also be included. Finally, since the vital signs measurements are summarized by parameter and analysis visit, PARAM, PARAMCD (for identification), PARAMN (for sorting), AVISIT (for identification) and AVISITN (for sorting) are also needed. This example also includes ANL01FL, which is a permissible variable used here for identifying specific records to be summarized if a subject has multiple values recorded for a parameter at a single visit.

TABLE 5



14.3.7.2 Electrocardiogram Results by Parameter
Safety Population

<Electrocardiogram Parameter (Units)>

Visit	BP33404 (N=xx) Actual	BP33404 (N=xx) Change From Baseline	Placebo (N=xx) Actual	Placebo (N=xx) Change From Baseline	Overall (N=xx) Actual	Overall (N=xx) Change From Baseline
Baseline						
N	xx		xx		xx	
Mean (SD)	xx.x (xx.xx)		xx.x (xx.xx)		xx.x (xx.xx)	
Median	xx.x		xx.x		xx.x	
Min, Max	xx, xx		xx, xx		xx, xx	
>450 ms [n (%)]	xx (xx.x)		xx (xx.x)		xx (xx.x)	
>480 ms [n (%)]	xx (xx.x)		xx (xx.x)		xx (xx.x)	
>500 ms [n (%)]	xx (xx.x)		xx (xx.x)		xx (xx.x)	
Week 24						
N	xx	xx	xx	xx	xx	xx
Mean (SD)	xx.x (xx.xx)	xx.x (xx.xx)	xx.x (xx.xx)	xx.x (xx.xx)	xx.x (xx.xx)	xx.x (xx.xx)
Median	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
Min, Max	xx, xx	xx, xx	xx, xx	xx, xx	xx, xx	xx, xx
>450 ms [n (%)]	xx (xx.x)		xx (xx.x)		xx (xx.x)	
>480 ms [n (%)]	xx (xx.x)		xx (xx.x)		xx (xx.x)	
>500 ms [n (%)]	xx (xx.x)		xx (xx.x)		xx (xx.x)	
>30 ms Increase [n (%)]		xx (xx.x)		xx (xx.x)		xx (xx.x)
>60 ms Increase [n (%)]		xx (xx.x)		xx (xx.x)		xx (xx.x)

Reference: Appendices 16.2.8.1.1-16.2.8.1.4
Note: Baseline value is mean of 3 tracings collected during screening period. The change from baseline value is calculated using this mean baseline value. SD = Standard Deviation, Min = Minimum, Max = Maximum.

This table is very similar to the previous example, in that it also displays univariate summary statistics by parameter and visit, so the BDS is an appropriate choice. The footnote indicates that the baseline value is the average of 3 tracings collected during the screening period; since that represents a composite baseline definition, derivation type (DTYPE) is included to identify the derived baseline records that will need to be generated. Counts of subjects with ECG parameter values and increase from baseline above various thresholds are also displayed, and those categories are not mutually exclusive, so CRITy/CRITyFL variables were chosen to indicate how the records should be counted. However, since the categories depend solely on analysis value and change from baseline, respectively, AVALCATy and CHGCATy could also be used. In that case, the category definitions would be mutually exclusive, and the record counts obtained by subsetting on combinations of category values.

TABLE 6

Table 6

OCCDS

14.3.7.3 Electrocardiogram Abnormalities
Safety Population

TRTA

SAFFL

ECG Abnormality	BP3304 (N =xx) n (%)	Placebo (N =xx) n (%)	Overall (N =xx) n (%)
Any Abnormality	xx (xx.x)	xx (xx.x)	xx (xx.x)
Abnormality I	xx (xx.x)	xx (xx.x)	xx (xx.x)
Abnormality II	xx (xx.x)	xx (xx.x)	xx (xx.x)
Abnormality III	xx (xx.x)	xx (xx.x)	xx (xx.x)

ATERM

Reference: Listing 16.2.9.1
Note: Percentages are based on the number of subjects in each population. Subjects identified as having abnormalities more than once are only counted once for that abnormality. Only abnormalities reported after the start of study medication are displayed.

ADT

TRTSDT

This table summarizes counts of subjects with various recorded ECG abnormalities, and represents a good example of a case where an ADaM OCCDS dataset is created from an SDTM Findings domain. Since we are simply counting abnormalities, there is no analysis value defined on the record, and analysis term (ATERM) is used to hold the abnormality description. The footnote indicates that only abnormalities reported after the start of study medication are displayed, so analysis date (ADT) and treatment start date (TRTSDT) are needed in the dataset in order to identify such records. They could then be identified with an occurrence flag, or an analysis flag (ANLzzFL) could be added to the dataset.

Note that this table could also be generated from a BDS, with the abnormality stored in AVALC, and ANLzzFL used to identify the first occurrence of each abnormality for a subject. However, since the table simply counts subjects with abnormalities over the entire study, OCCDS seemed a more appropriate structure.

CONCLUSION

Choosing which ADaM dataset structure to use for generating a specific table, listing, or figure should be driven by the type of analysis that needs to be performed, and not by the structure of the source SDTM or ADaM datasets. Be sure to include any ADSL variables required for summarization, such as subject-level population flags and study treatment start and end dates. Keeping additional variables providing traceability, such as those used in derivations, is also recommended.

REFERENCES

CDISC ADaM documents are available at https://www.cdisc.org/standards/foundational/adam. This paper referenced the following documents at this site:

- Analysis Data Model (ADaM) v2.1
- Analysis Data Model Implementation Guide (ADaMIG) v1.3
- ADaM Structure for Occurrence Data (OCCDS) Implementation Guide v1.1
- ADaM Basic Data Structure for Time-to-Event Analyses v1.0

ACKNOWLEDGMENTS

Many thanks to those who reviewed the slides and paper and offered suggestions.

CONTACT INFORMATION

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

Nancy Brucken
IQVIA
nancy.brucken@iqvia.com