PharmaSUG 2025 - Paper DS-63

The Winding Road to ADSL

Elizabeth Dennis, EMB Statistical Solutions LLC;
Grace Fawcett, Syneos Health

ABSTRACT

The typical study will use SDTM data as the sole origin for the ADaM datasets. ADSL is derived using SDTM.DM, EX, and other domains.

But what's the best process to use when the roadmap to ADSL contains twists and turns? There are times when the typical process is not robust enough to produce a well-documented ADSL that meets all the analysis needs.

This presentation will go through three scenarios where details of the study and the analysis require a more complicated process to produce ADSL:

1. The derivations for a flag variable (usually a population flag variable) are so complicated that an intermediate dataset is helpful.
2. The derivations for a flag variable (usually a population flag variable) require clinical adjudication.
3. There are multiple participations per subject.

INTRODUCTION

Statistical programmers are very familiar with ADSL. It's required for all submissions[1], and its purpose is to provide the variables that describe the attributes of a subject. The structure must be one record per subject. It brings together data from many different SDTM sources to summarize the subject's study experience, and sets study population flags. It also provides a convenient single source to pull the subject information for all other analysis datasets. In a study's set of ADaM datasets, ADSL is usually derived first.

The statistical analysis plan (SAP) provides the details of the definition of the study populations. Many of the populations are simple to define. The safety population, for example, generally flags subjects who received any study drug.

There may be studies, though, where the variables necessary for ADSL have complex derivations. They could require many sources of data for the definition. All of these items need to be documented and traceable. In such cases, deriving ADSL as the first ADaM dataset is not practical.

This presentation will discuss are three specific scenarios where creating intermediate ADaM datasets prior to ADSL is an efficient technique. While going through this process, it's important to keep in mind two of ADaM's fundamental principles: facilitate clear and unambiguous communication of the content and source of the datasets, and provide traceability between the analysis data and its source data. By using intermediate datasets, the process has many small, clear, well-defined derivations, instead of a single large derivation that's hard to define and document.

CASE STUDY 1: AN EXTREMELY COMPLICATED POPULATION FLAG

Correctly identifying the analysis population is important for every study. For the rare situation where the derivation is very complex, using intermediate datasets can aid in the derivation, traceability, and documentation.

Case study 1 involves an efficacy analysis population that had 19 different criteria. A subject could violate more than 1 criterion. For analysis it was necessary to document which criteria each subject violated. The primary efficacy endpoint was an assessment done after two years on the study. Examples of subject exclusions include specific adverse events during the time on study, specific concomitant medication use, medical history exclusions, and others.

For this example, the team created a dataset called ADPRE1, which contained all the usual ADSL variables, with the exception of the efficacy population flag. Next, a Basic Data Structure (BDS) dataset (ADPRE2) with 19 parameters per subject was created. Each subject was flagged Yes/No for each parameter, and other identifying variables were populated to aid traceability.

SUBJID	TRT01P	PARAMCD	PARAM	AVALC	SRCDOM	SRCSEQ
502-01	DRUG A	CRIT1	Exclusionary AE occurred	Y	AE	12
502-01	DRUG A	CRIT2	Received prohibited concomitant medication	N		
502-01	DRUG A	CRIT3	Received prohibited concomitant procedure	Y	PR	5
502-01	DRUG A	Additional criteria through CRIT19				

Table 1 BDS Variables in ADPRE2

ADSL was then created by combining ADPRE1 with ADPRE2. This process resulted in many simple, well documented derivations instead of one large untraceable derivation.

Dataset Name	Description	Class	Source Data
ADPRE1	Pre-ADSL Analysis	ADAM OTHER	SDTM.DM EX DS
ADPRE2	Efficacy Population Analysis	BASIC DATA STRUCTURE	ADAM.ADPRE1 SDTM.AE CM PR
ADSL	Subject-Level Analysis Dataset	SUBJECT LEVEL ANALYSIS DATASET	ADAM.ADPRE1 ADPRE2

Table 2 Datasets leading to ADSL

CASE STUDY 2: POPULATION FLAG REQUIRES CLINICAL ADJUDICATION

This study indication was infection disease. In these studies, it's not unusual for a subject to qualify for participation based on clinical signs, but final determination is based on infection growth that's not determined until months later. Qualifying for the efficacy population can depend on many different factors, like prior medications received and bacterial growth. To determine a subject's infection status, a clinician needs to look at a variety of data.

Much like Case Study 1, the ADaM team started with a dataset called ADPRE1, which contained all the usual ADSL variables, with the exception of those associated with the population determination. Next the team created a BDS dataset called ADPRE2, with one parameter for each information piece needed by the clinician. This information was sent to the clinical team in a spreadsheet, with additional columns to be completed by that team. After review, the clinical team made a blinded determination of each subject's suitability for the efficacy population and entered information to document their findings.

That spreadsheet was read back into ADPRE2 to create additional parameters associated with the final population determination.

The table below shows a few of the parameters the clinical team required, in addition to the parameters that the clinical team populated.

SOURCE	PARAMCD	PARAM	AVALC	ADT
Derived by ADaM team from SDTM	CRIT1	Relevant prior medications	Doxycycline	10Jan2024
	CRIT2	Quantitative Culture Colony Count	36 CFU/ml	02Feb2024
	CRIT3	MIC - GENTAMICIN	1 ug/mL	02Feb2024
Received from clinical team	EFFPOP	Include in population?	N	02Apr2024
	EXCREAS1	Reason 1 excluded from population	No baseline MAC quantitative sputum culture result	02Apr2024
	EXCREAS2	Reason 2 excluded from population	NA	02Apr2024

Table 3 BDS Parameters in ADPRE2

The final version of ADPRE2 in the submission contained both sets of parameters: those programmed by the ADaM team and sent to the clinical team, and those determined by the clinical team and sent back to the ADaM team. ADSL was then easy to produce, combining ADPRE1 and ADPRE2. Value level metadata in the define.xml clearly identified the source of all data.

CASE STUDY 3: MULTIPLE PARTICIPATIONS PER SUBJECT

There are many studies where a subject can only participate once. There are also studies where multiple participations are allowed. The most common occurrence is when a screen failure subject is allowed to re-screen. There are other cases with rare disease study populations where a subject can participate, complete a study, and later re-enroll and participate again.

The FDA Study Data Technical Conformance Guide[2] says that a subject who is screened and/or enrolled more than once in a study should use a different SUBJID for each unique screening. It also states that a subject should be identified using the same USUBJID across the application. That causes challenges for ADSL, where the rules remain the same: it must be one record per subject. To capture the data needed to summarize the individual participations, ADPL[3] (Participation-Level Analysis Dataset) can be used.

This process generally produces ADPL first, to meet the needs of any of the study's analysis where multiple participations are included.

STUDYID	USUBJID	SUBJID	TRTSDT	TRT01A	AGE	BLECOG
01-ABC	01-ABC-01	01	02Feb2023	Drug 1	51	1
01-ABC	01-ABC-01	05	06Aug2024	Drug 1	52	1
01-ABC	01-ABC-38	38	12Mar2023	Drug 1	47	1
01-ABC	01-ABC-38	73	27Nov2024	Drug 1	48	0

Table 4 ADPL with multiple participations per subject

Consideration must be taken when producing ADSL. It must be one record per subject, but how it's reduced depends on the analysis needs. Are there analyses that will be done per subject instead of per participation? If so, is the first treatment used or the last treatment, or a combination of both? These details should be documented in the Statistical Analysis Plan.

In Case Study 3, the SAP detailed that all safety analysis would treat multiple participations as a single subject based on the initial treatment date. ADSL was modeled to facilitate that analysis.

STUDYID	USUBJID	SUBJID	TRTSDT	TRT01A	AGE	BLECOG
01-ABC	01-ABC-01	01	02Feb2023	Drug 1	51	1
01-ABC	01-ABC-38	38	12Mar2023	Drug 1	47	1

Table 5 ADSL with one record per subject

DOCUMENTATION

The datasets and derivations must be clearly documented in the define.xml and the Analysis Data Reviewer's Guide (ADRG)[4].

DEFINE.XML

For Case Study 2, the define.xml uses Value Level Metadata to show the origin for each parameter. The relevant excerpt of the define.xml is below.

Variable	Where Condition	Label / Description	Type	Origin / Source / Method / Comment
AVALC		Analysis Value (C)	text	
	PARAMCD = ‘CRIT1’		text	Derived If subject has a record where SDTM.CMDECOD = ‘Doxycycline’ then AVALC = CMDECOD
	PARAMCD = ‘CRIT2’		text	Derived SDTM.MB.MBORRES where MBTESTCD = ’COLCOUNT’
	PARAMCD = ‘CRIT3’		text	Derived SDTM.MS.MSORRES where MSTESTCD = ‘MIC’ and MSAGENT = ‘GENTAMICIN’
	PARAMCD = ‘EFFPOP’		text	eDT Using external file [CLINADJ], set to value in column ‘Population Determination’

Table 6 Define.xml documentation of Case Study 2

ANALYSIS DATA REVIEWER’S GUIDE

The ADRG offers a few locations to more fully describe the process used to create these datasets. The sections[4] ‘Source Data Used for Analysis Dataset Creation’, ‘Intermediate Datasets’, and ‘Analysis Datasets’ all are appropriate sections to detail the derivations.

CONFORMANCE TO ADAM STANDARD

Conformance checks done using Pinnacle 21 do not flag any issues with the approaches used in these case studies. Conformance to the ADaM model, though, goes further than simply passing Pinnacle 21 checks. The strategy used in Case Study 2 goes against the ADaM standard of using only SDTM domains and other ADaM datasets as the source. However, these other important principles were met:

1. Variables that have been derived or imputed in ADaM datasets should not be copied back into the SDTM source data.
2. The ADaM datasets meet the analysis need.
3. The dataset and variable sources are clearly defined.

For these reasons, the ADaM team used these strategies to create datasets that are simple to understand and easy to trace back to the data origin.

CONCLUSION

While it is common and usually appropriate to produce ADSL before any other ADaM dataset, there are studies when that process doesn't work. Creating intermediate ADaM datasets can help simplify the analysis and produce better documentation and traceability.

REFERENCES

1 CDISC: ADaM 2.1 ADaM v2.1 | CDISC

2 FDA: "Study Data Technical Conformance Guide v5.9 https://www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources

3 PharmaSUG: Dennis, E. et al. "ADaM Datasets with Multiple Participations per Subject" ADaM Datasets with Multiple Participations per Subject

4 PHUSE. "ADRG v1.2" https://advance.phuse.global/display/WEL/Analysis+Data+Reviewer%27s+Guide+%28ADRG%29+Package.

CONTACT INFORMATION

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

Elizabeth Dennis
EMB Statistical Solutions LLC
edennis@embstats.com

Grace Fawcett
Syneos Health
grace.fawcett@syneoshealth.com