

## ADaM Datasets with Multiple Participations per Subject

Elizabeth Dennis, EMB Statistical Solutions, LLC

Grace Fawcett, Syneos Health

Sandra Minjoe, ICON PLC

### ABSTRACT

The SDTM team has proposed a standardized technique for accommodating multiple participations per subject. In that proposal, the domain DM (Demographics) contains all subjects, with one primary record per subject, and the domain DC (Demographics for Multiple Participations) contains all subjects, with each participation per subject as a separate record.

There is often also a need to represent the multiple participations in analysis, so the multiple participations must be brought into ADaM. A technique similar to the SDTM solution may be used, where the standard ADaM dataset ADSL still contains all subjects, with one primary record per subject, but another dataset, which we call ADPL (Participation-Level Analysis Dataset) is used to contain all participations per subject, with one record per participation per subject.

This paper will summarize the SDTM DM vs. DC proposal and describe how a similar ADSL vs. ADPL solution could work in ADaM. We will provide some examples of what to include in ADSL vs. ADPL, showing how to use the data from these datasets in other datasets and for analysis. We will also give examples of issues that could become problematic across these datasets. Finally, we will give example text for an ADRG to make it clear what was done and where to find important information.

### INTRODUCTION

#### USUBJID VS SUBJID

The SDTMIG v3.2<sup>1</sup> provides CDISC Notes to explain the use of USUBJID and SUBJID:

- SUBJID: "Subject identifier, which must be unique within the study. Often the ID of the subject as recorded on a CRF."
- USUBJID: "Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product. This must be a unique number, and could be compound identifier formed by concatenating STUDYID-SITEID-SUBJID."

These notes handle the case when the same subject is enrolled in multiple studies. It does not cover cases where the same subject participates in a study more than once.

#### REASONS FOR MULTIPLE PARTICIPATIONS

There are several reasons why a person might participate in a single study more than once. A common scenario for multiple participations is when a screening failure subject then goes through the screening process again in another attempt. There are also studies where a subject may participate more than once with more than one body part, such as a subject who participates with the left eye, and then enrolls again with the right eye. In rare diseases, sometimes a subject is allowed to participate in a study once, completes the study, and then later participates again.

#### FDA INPUT

The Study Data Technical Conformance Guide<sup>2</sup> 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. For a study with multiple screenings or multiple enrollments, SUBJID should be included in all SDTM domains.

However, SDTM does not allow SUBJID to be a variable in datasets other than DM. Adding SUBJID to these domains will fire automated conformance checks.

## SDTM PROPOSAL

The CDISC Study Data Standards (SDS) team developed a proposal to accommodate multiple participations<sup>3</sup>. This proposal went out for public review in 2022 and the team is now addressing comments. It states that each enrollment should have a separate record in DC (Demographics for Multiple Participations), using a unique SUBJID for each record and a common USUBJID for each subject.

The domain DM (Demographics) will continue to contain only 1 record per subject. The SUBJID value used in DM is the value found in the “primary” DC record. The identification of which record is the “primary” record is left to the discretion of the sponsor. For cases where a subject goes through multiple screenings and eventually enrolls in the study, the primary SUBJID is probably the one where the subject was enrolled. For subjects who enroll in the study more than once, the primary SUBJID is a little tougher to choose, but the first enrollment might be a good choice.

Here is an example of a DC dataset for 2 subjects, who each have 2 participations. The first subject initially failed screening, but tried again and was enrolled. The second subject enrolled in the study on two separate occasions.

STUDYID	USUBJID	SUBJID	DCSEQ	RFXSTDTC	ARM	ARMNRS	COUNTRY
AB-123	AB-123-08	01	1			SCREEN FAILURE	USA
AB-123	AB-123-08	08	2	2023-01-23	Drug 1		USA
AB-123	AB-123-99	12	1	2020-01-24	Drug 1		USA
AB-123	AB-123-99	99	2	2023-03-18	Drug 2		USA

**Table 1 DC (Demographics for Multiple Participations)**

If the last DC record for the subject is chosen to be “primary”, then DM would contain the following:

STUDYID	USUBJID	SUBJID	RFXSTDTC	ARM	ARMNRS	COUNTRY
AB-123	AB-123-08	08	2023-01-23	Drug 1		USA
AB-123	AB-123-99	99	2023-03-18	Drug 2		USA

**Table 2 DM (Demographics)**

## ADAM APPLICATIONS

### ANALYSIS NEEDS

For screen failures, first determine if the screen failure records are even needed for analysis. The ADaM model document<sup>4</sup> states “Whether analysis datasets include data for subjects not analyzed (e.g., screen failures) is a sponsor decision and should be communicated with the reviewers or users of the data. If these data are included, they should be incorporated in the appropriate analysis datasets such as ADSL (as opposed to separate datasets for non-analyzed subjects) using appropriate flag variables to clearly differentiate these records. The metadata must specify that these data are included and how to distinguish them.”

### ADAM DATASETS ADSL AND ADPL

In cases where there is a need to analyze the multiple participations, how do we handle them in ADSL? ADSL is required to contain only one record per subject, so, like DM, it is not designed to handle multiple participations. A technique similar to the SDTM solution may be used, where the standard ADaM dataset

ADSL contains one primary record per subject, but another dataset, which we call ADPL (Participation-Level Analysis Dataset) is used to contain all participations per subject, with one record per participation per subject.

Which variables need to be defined at the participation-level is study dependent, but we will use Treatment, Age, and Baseline Weight in our examples. Here is an example of ADPL for these same 2 subjects shown in the SDTM examples above:

STUDYID	USUBJID	SUBJID	TRTSDT	TRT01A	ARMNRS	AGE	BLWT
AB-123	AB-123-08	01			SCREEN FAILURE	62	55
AB-123	AB-123-08	08	23JAN2023	Drug 1		62	57
AB-123	AB-123-99	12	24JAN2020	Drug 1		70	53
AB-123	AB-123-99	99	18MAR2023	Drug 2		73	46

**Table 3 ADPL (Participation-Level Analysis Dataset) with multiple participations per subject**

For this example, the last participation is the one used in ADSL. This is a study-specific choice.

STUDYID	USUBJID	SUBJID	TRTSDT	TRT01A	ARMNRS	AGE	BLWT
AB-123	AB-123-08	08	23JAN2023	Drug 1		62	57
AB-123	AB-123-99	99	18MAR2023	Drug 2		73	46

**Table 4 ADSL (Subject-Level Analysis Dataset)**

## OTHER ADAM DATASETS

Unlike SDTM datasets, SUBJID can always be included in ADaM datasets. Including both USUBJID and SUBJID as key variables in analysis datasets will allow the different participations for each subject to be easily identified.

If an ADaM dataset requires multiple participations to be included, then that dataset should use ADPL to bring in the participation-level data. In this way, items like baseline and treatment at time of assessment will be properly defined.

USUBJID	SUBJID	TRTA	VSSEQ	PARAMCD	VISIT	AVAL	ABLFL	BASE	CHG
AB-123-99	12	Drug 1	1	PULSE	Screening	60		62	
AB-123-99	12	Drug 1	2	PULSE	Cycle 1 Day 1	62	Y	62	
AB-123-99	12	Drug 1	3	PULSE	Cycle 2 Day 1	63		62	1
AB-123-99	12	Drug 1	4	PULSE	End of Treatment	62		62	0
AB-123-99	99	Drug 2	5	PULSE	Screening	64		64	
AB-123-99	99	Drug 2	6	PULSE	Cycle 1 Day 1	64	Y	64	
AB-123-99	99	Drug 2	7	PULSE	Cycle 2 Day 1	67		64	3
AB-123-99	99	Drug 2	8	PULSE	End of Treatment	68		64	4

**Table 5 ADVS (Vital Signs Analysis) with multiple participations per subject**

## PREPARING DATA FOR ANALYSIS

As with all analysis datasets, derivations must be made using definitions specified in the Statistical Analysis Plan (SAP). For studies with multiple participations, denominators for both safety assessments and efficacy assessments must be clearly defined. For example, treatment-emergent could be defined as any adverse event which occurred after the initial dose of study drug for a subject, including later participations.

Treatment emergent flags and baseline flags from SDTM in these situations should be carefully scrutinized to determine if they meet the analysis need, as defined in the SAP. For example, the ADaM study day variable ADY may need to be derived differently SDTM VSDY.

USUBJID	SUBJID	VSSEQ	TRTSDT	VISIT	ADT	VSDY	ADY
AB-123-99	12	1	24JAN2020	Screening	01JAN2020	-23	-23
AB-123-99	12	2	24JAN2020	Cycle 1 Day 1	24JAN2020	1	1
AB-123-99	12	3	24JAN2020	Cycle 2 Day 1	21FEB2020	29	29
AB-123-99	12	4	24JAN2020	End of Treatment	20MAR2020	57	57
AB-123-99	99	5	18MAR2023	Screening	01MAR2023	1133	-17
AB-123-99	99	6	18MAR2023	Cycle 1 Day 1	18MAR2023	1150	1
AB-123-99	99	7	18MAR2023	Cycle 2 Day 1	18APR2023	1181	32
AB-123-99	99	8	18MAR2023	End of Treatment	28APR2023	1191	42

**Table 6 ADVS with Study Day derived using each SUBJID's participation**

Here SDTM study day VSDY is derived using the date of first treatment from DM.RFXSTDTC (see Table 2). However, in ADaM study day ADY needs to be derived using the date of first treatment from ADPL.TRTSTDT (see Table 4), so that each SUBJID can be analyzed separately. Notice that the value of VSDY is the same as ADY for the first participation, but is different for the second enrollment.

## DOCUMENTATION

The participations and derivations must be clearly documented in the define.xml and the Analysis Data Reviewer's Guide (ADRG)<sup>5</sup>.

In the define.xml documentation for the ADVS dataset, be sure to reference dataset ADPL. Cases where ADPL is used instead of the typical ADSL are shown in red in the example below. (*Note that not all rows or columns for the define.xml are shown.*)

Variable	Label / Description	Type	Origin / Source / Method / Comment
USUBJID	Unique Subject Identifier	text	Predecessor: VS.USUBJID
SUBJID	Subject Identifier for the Study	text	Predecessor: VS.SUBJID
TRTA	Actual Treatment	text	Predecessor: <b>ADPL</b> .TRT01A
VSSEQ	Sequence Number	integer	Predecessor: VS.VSSEQ

Variable	Label / Description	Type	Origin / Source / Method / Comment
PARAMCD	Parameter Code	text	Predecessor: VS.VSTESTCD
VISIT	Visit	text	Predecessor: VS.VISIT
TRTSDT	Treatment Start Date	Integer	Predecessor: <b>ADPL</b> .TRTSDT
ADT	Analysis Date	Integer	Derived: numeric version of VS.VSSTDTC
ADY	Analysis Date	Integer	Derived: ADT – TRTSDT + (ADT >= .TRTSDT)
ABLFL	Baseline Record Flag	text	Derived: Set ABLFL='Y' on the last record where ADT <= TRTSDT.
AVAL	Analysis Value	float	Predecessor: VS.VSSTRESN
BASE	Baseline Value	float	Derived: AVAL when ABLFL=Y
CHG	Change from Baseline	float	Derived: CHG = AVAL – BASE

**Table 7 define.xml ADVS documentation referencing dataset ADPL**

The ADRG section titled “Subject Issues that Require Special Analysis Rules” is a good place to call attention to the multiple participations. Example text might look something like this:

### 3.3 Subject Issues that Require Special Analysis Rules

Subjects were allowed to enroll in this study more than once. In all cases, these subjects maintain the same unique subject identifier (USUBJID), but for each participation they have a different subject identifier (SUBJID).

Here is the list of subjects who had multiple participations:

USUBJID	SUBJID
AB-123-08	01
	08
AB-123-99	12
	99
...	

To manage the core data needed at the subject-level and the participation-level, the following two ADaM datasets were used:

- ADSL (Subject-Level Analysis Dataset): contains core information at the unique subject-level (one record per USUBJID).
- ADPL (Participation-Level Analysis Dataset): contains similar core information as ADSL, but at the subject participation-level (one record per SUBJID).

Other ADaM datasets use either ADSL or ADPL to bring in the necessary core variables, depending on the analysis need. Information about which datasets use ADSL versus ADPL for core content is found in the [Analysis Datasets](#) section below. Additionally, variable derivations for these other ADaM datasets show from which dataset any core content was obtained.

**Table 8 ADRG section: Subject Issues that Require Special Analysis Rules**

The “Analysis Datasets” section of the ADRG is where dataset ADPL itself should be described in more detail, and where other datasets that contain multiple participations should be noted. Example text might look something like this:

<b>5.2 Analysis Datasets</b>			
<b>Dataset</b>	<b>Class</b>	<b>...</b>	<b>Structure</b>
ADSL Subject-Level Analysis Dataset	ADSL		One record per subject
ADPL Participation-Level Analysis Dataset	ADAM OTHER		One record per subject per participation
ADVS Vital Signs Analysis Dataset	BDS		One record per subject per participation per parameter per visit
...			

**5.2.1 ADSL – Subject-Level Analysis Dataset**

This dataset contains subject-level content, including core variables. Because it is structured as one record per subject, it does not contain individual participations when a subject has more than one participation. For subjects that participated in the study more than once, the content from the last participation is used in ADSL. See section [Subject Issues that Require Special Analysis Rules](#) for the list of subjects who participated more than once.

**5.2.2 ADPL – Participation-Level Analysis Dataset**

This dataset contains participation-level content, including core variables. Because it is structured as one record per subject per participation, it contains individual participations when a subject has more than one participation. See section [Subject Issues that Require Special Analysis Rules](#) for the list of subjects who participated more than once.

**5.2.3 ADVS – Vital Signs Analysis Dataset**

This dataset contains content from multiple participations across the same subject. It uses dataset ADPL rather than ADSL for all core variables.

...

**Table 9 ADRG section: Analysis Datasets**

**CONFORMANCE**

Advantages of this approach are that it follows the ADaM principles and doesn't break any ADaM rules. The ADaM datasets designed as shown in the examples will pass conformance checks.

Alternate approaches, such as where ADSL is structured as one record per subject participation or not including ADSL in favor of just ADPL, will break ADaM rules and fail conformance checks. These approaches are not recommended.

**CONCLUSION**

A study with multiple participations per subject requires extra attention in SDTM. If the multiple participations are necessary in analysis, ADSL alone is not sufficient. ADSL must be structured as one record per subject (USUBJID). The new proposed dataset ADPL is similar to ADSL, except that it is structured as one record per SUBJID. When deriving other analysis datasets, ADSL can still be used if subject-level content is needed, but ADPL provides this data with one record per participation. This method follows the ADaM fundamental principles and conformance check rules.

When using ADPL in addition to ADSL, be sure to document which dataset was used for any core variables, both in the define.xml and in the ADRG.

## REFERENCES

- 1 CDISC. “SDTMIG v3.2” <https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-v3-2>.
- 2 FDA: “Study Data Technical Conformance Guide v5.0” <https://www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources>.
- 3 CDISC. “Demographics Data (Draft new section)” <https://wiki.cdisc.org/pages/viewpage.action?pageId=157584287>.
- 4 CDISC. “ADaM v2.1” <https://www.cdisc.org/standards/foundational/adam/adam-v2-1>.
- 5 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  
[edennis@embstats.com](mailto:edennis@embstats.com)

Grace Fawcett  
Syneos Health  
[grace.fawcett@syneoshealth.com](mailto:grace.fawcett@syneoshealth.com)

Sandra Minjoe  
ICON PLC  
[sandra.minjoe@iconplc.com](mailto:sandra.minjoe@iconplc.com)