Paper SI03

# USUBJID Challenges: Representing Multiple Participations for Analysis & Reporting

Emily Bernstein-Hanlon, Pfizer Inc., USA
Sirisha Davuluri, Pfizer Inc., USA

## ABSTRACT

Multiple enrollments in a study can be challenging to address, particularly with increasing complexity of trial designs, evolving regulatory requirements and without clear data standards. Scenarios range from study designs where re-enrollment by the same participant is allowed to participants enrolling multiple times when a protocol does not allow it. Per the FDA Study Data Technical Conformance Guide, SUBJID should be unique for each screening or enrollment in a study, but USUBJID should remain the same across all studies involving the product. Appropriate and timely data collection is necessary to be able to link all the enrollments together. Challenges can arise when initial data entry is incomplete, and updates are made between multiple submissions. Several approaches can be used to handle re-enrollment, each with its own advantages and disadvantages. This paper will discuss these scenarios and possible approaches, with a focus on ADaM and the corresponding analysis and reporting.

## INTRODUCTION

Multiple participations challenge our existing structures for data collection, SDTM[1] mapping, ADaM[2] design, and analysis and reporting. Multiple participations can occur either as part of planned re-enrollment of the same participant to gain efficiencies in complex study designs or as unplanned protocol deviations where a participant enrolls multiple times in a study without site's or sponsor's prior knowledge. The approach for SDTM mapping can be similar between these two cases, while approaches for ADaM design reporting depend on the needs of a particular analysis and require careful consideration.

The FDA Study Data Technical Conformance Guide (TCG)[3] states that if a single participant screens or enrolls in a study multiple times, a unique SUBJID should be assigned for each participation. SUBJID should be included in all related domains besides the DM (Demographics) domain, even though this can cause Pinnacle 21 (P21) validation errors. The DM domain should contain the primary enrollment/screening, and additional enrollments or screenings should be included in a custom domain structured like the DM domain. The Clinical Data Interchange Standards Consortium (CDISC) Multiple Subject Instances Team proposed that this domain be called the DC domain, Demographics for Multiple Participations, which is expected to be released in SDTM Implementation Guidelines (IG) version 4.0 at the earliest[4].

CDISC SDTM standards require that USUBJID uniquely identify a participant "across all studies for all applications or submissions involving the product"[1]. In cases where a participant enrolls or screens multiple times in a study, the requirement is that USUBJID will remain the same throughout the study, while SUBJID identifies each unique participation. The TCG recommends including an explanation in the data reviewer's guide for any such cases along with a table showing all SUBJID for each USUBJID in the study.

The TCG and proposed DC domain provide relatively straightforward guidance for SDTM mapping, but the approach for ADaM has not been formalized and depends on the needs of the particular analysis. In this paper we will explore several multiple participation scenarios, provide potential approaches, and discuss the advantages and disadvantages of each.

## DATA COLLECTION

USUBJID must uniquely identify a participant across all studies for all applications or submissions involving the investigational product. In the case of multiple participations, SUBJID uniquely identifies each participation, while USUBJID remains the same. Data collection of prior study participation is critical to be able to assign the correct USUBJID. In case there is only a single prior participation expected, the prior STUDYID, SITEID and SUBJID should be collected. If there are multiple prior participations, it becomes more complicated. The first participation details would be required to assign USUBJID, and a sponsor may want to collect additional participations between the first and the current one depending on data entry and reconciliation needs. Collecting the most recent participation details besides the first could assist with automatic flow of data like medical history and prior medications from the most recent participation to the current participation.

Difficulties arise when sites do not understand how to complete the questions correctly, deviating from the CRF completion guidelines, or enter the data late, not understanding the criticality of these fields, resulting in an incorrect USUBJID being assigned. Under a protocol design which allows re-enrollment this would be anticipated, but in cases of unplanned multiple enrollments when sites are not aware, it can be challenging to track down when two participations are the same person. There are various detection methods used by sponsors based on demographic data or using artificial intelligence, but still these cases can be difficult to fully detect. If prior participation information is not initially entered, or entered incorrectly, and if there are interim data cuts for submissions, the situation can arise where USUBJID changes between data cuts causing significant

challenges for those submissions. Thorough site training, monitoring and source data verification can help mitigate some of these risks to get the data entered correctly.

The scenarios described below cover a case of planned re-enrollment, where all prior participation details were entered correctly so the two SUBJID values can be linked to a single USUBJID. Unplanned multiple enrollments, once identified and entered, could follow the same examples discussed below. Whether planned or unplanned, similar challenges arise on how to map and analyze the data.

Two analysis scenarios will now be discussed for a case with a subject who re-enrolls in the same study after a four-month gap, one where a single participation is summarized and the second where both participations are summarized. The analysis needs will determine how data is mapped in SDTM and ADaM to be ready for analysis and reporting.

## SCENARIO ONE – ONE PARTICIPATION SUMMARIZED

The first scenario is where a single enrollment is summarized when a specific participation is selected for analysis. This occurs in a planned re-enrollment or unplanned multiple enrollments where only one participation is analyzed instead of all participations per study analysis needs. For an interim CSR with a subset of analysis in scope, either a subset can be applied to SDTM and ADaM including only the participation of interest to avoid complications with multiple participations, or cumulative domains can be created, however this introduces additional challenges. We will discuss both approaches to Scenario One below.

### SDTM Approaches

In this example, a participant enrolled first to the 10mg study intervention, then returned and enrolled again in the 30mg study intervention eight months later. One approach when only a single participation needs to be summarized is to apply a subset to the data when creating SDTM to select the second participation data. For a multi-enrolled study with single participation summarized there will be no change to mapping for SDTM domains as the focus will be on the required participation (second participation in this approach), with all dates referencing to the corresponding participation. One limitation of this approach is it would not be possible to include data from all participations in listings.

Below is an example of the DM domain if an individual participation is summarized, repeated for each participation. When the second SUBJID is selected, RFSTDTC, RFENDTC and RFICDTC come from the second participation; however, USUBJID is mapped based on the individual's first participation and ARMCD will be set to C for ARM = 'STUDY DRUG 30 mg'. If the first participation is considered instead of the second, then the corresponding participation dates and ARM/ARMCD values would be used.

*dm.xpt for first participation*

| USUBJID | SUBJID | RFSTDTC | RFENDTC | RFICDTC | BRTHDTC | ARMCD | ARM |
|---|---|---|---|---|---|---|---|
| AAA 001 1001 | 1001 | 2023-01-01 | 2023-03-30 | 2022-12-03 | 1998-01-01 | A | STUDY DRUG 10 mg |

*dm.xpt for second participation*

| USUBJID | SUBJID | RFSTDTC | RFENDTC | RFICDTC | BRTHDTC | ARMCD | ARM |
|---|---|---|---|---|---|---|---|
| AAA 001 1001 | 1003 | 2023-09-01 | 2023-11-30 | 2023-08-05 | 1998-01-01 | C | STUDY DRUG 30 mg |

Another approach to this scenario is to retain all participations by creating cumulative SDTM domains, using ADaM datasets to flag the records to use for analysis. The DM domain needs to be redefined to summarize all participations in a single record. This can be achieved by setting RFSTDTC from the first participation, RFENDTC from the last participation and adding alternative 'SUBJIDy' variables like the below table. Since these are nonstandard variables (NSV), it may be preferable to place them in the supplemental domain to avoid conformance issues during P21 validation. Mapping additional variables, like prior participation RFICDTC, RFSTDTC and ARM could become cumbersome. This is where a DC domain would be useful to contain all participation data for traceability.

*dm.xpt*

| USUBJID | SUBJID | SUBJID2 | RFSTDTC | RFENDTC | RFICDTC | BRTHDTC | ARMCD | ARM |
|---|---|---|---|---|---|---|---|---|
| AAA 001 1001 | 1003 | 1001 | 2023-01-01 | 2023-11-30 | 2022-12-03 | 1998-01-01 | C | STUDY DRUG 30 mg |

In Events and Findings domains, careful consideration must be made to the study day (--DY) variables and EPOCH. The anchor date for all --DY, --STDY, and --ENDY variables is defined as DM.RFSTDTC in SDTM IG v3.2 and v3.3[5,6]. In the AE example below, data from both participations was retained for USUBJID = "AAA 001 1001" with two different SUBJID values populated showing the multiple participations. Even though the focus is on the second participation for this USUBJID, still DM must contain the first RFSTDTC value of "2023-01-01" from the first participation. Otherwise, there would be conformance issues of observations prior to study participation start/informed consent when a subset is not applied. This means the study day variables and EPOCH will be derived based on first participation even though the second participation is the one of

interest. Similarly, the treatment emergence flag in SDTM (SUPPAE.AETRTEM) would be defined based on first participation, whereas in ADaM it would be based on the second participation. The SUBJID variable also needs to be included in the AE domain to present the different subject numbers for each participation to distinguish which events are associated with each participation.

*ae.xpt*

| USUBJID | SUBJID | AEDECOD | AESTDTC | AEENDTC | AESTDY | AENDY | EPOCH | SUPPAE. AETRTEM |
|---|---|---|---|---|---|---|---|---|
| AAA 001 1001 | 1001 | Anaemia | 2023-01-03 | 2023-01-05 | 3 | 5 | TREATMENT | Y |
| AAA 001 1001 | 1001 | Irritable bowel syndrome | 2023-01-10 | 2023-01-15 | 10 | 15 | TREATMENT | Y |
| AAA 001 1001 | 1003 | Headache | 2023-08-10 | 2023-08-12 | 222 | 224 | FOLLOW-UP | Y |
| AAA 001 1001 | 1003 | Cough | 2023-10-10 | 2023-10-20 | 283 | 293 | FOLLOW-UP | Y |

## ADaM Approaches

While standards are still evolving for ADaM datasets, there are several approaches for creating analysis ready domains. There is more flexibility with ADaM structure compared to SDTM datasets. In the first approach to Scenario One, SDTM was subset to the second participation so the below ADSL will follow a general parallel study design structure with one record per participant.

*adsl.xpt*

| USUBJID | SUBJID | TRT01P | TRTSDT | TRTEDT | EOSDT | RFICDTC | BRTHDTC |
|---|---|---|---|---|---|---|---|
| AAA 001 1001 | 1003 | STUDY DRUG 30 mg | 01SEP2023 | 01OCT2023 | 20NOV2023 | 2023-08-05 | 1998-01-01 |

In the second approach to Scenario One where cumulative ADaM domains are created to include multiple participations, ADSL might not see many challenges besides needing to add variables for the multiple subject identifiers along with different treatment periods. An NSV multiple enrollment flag (MULENRFL = 'Y') can be assigned to identify which participants have multiple enrollments/participations. If instead a subset of data is submitted, this flag is not needed since there is only a single participation included in the data.

*adsl.xpt*

| Row | USUBJID | SUBJID | SUBJID2 | TRTSDT | TRTEDT | TRT01P | TR01SDT |
|---|---|---|---|---|---|---|---|
| 1 | AAA 001 1001 | 1003 | 1001 | 01SEP2023 | 01OCT2023 | STUDY DRUG 10mg | 01JAN2023 |

| Row | TR01EDT | TRT02P | TR02SDT | TR02EDT | MULENRFL |
|---|---|---|---|---|---|
| 1 | 01FEB2023 | STUDY DRUG 30mg | 01SEP2023 | 01OCT2023 | Y |

The rest of the BDS (Basic Data Structure) and Occurrence ADaM datasets can retain the required participant level data per analysis needs. Below are the examples for ADAE and ADVS to show how cumulative datasets for multiple participations with a focus on single participation analysis will look like. The SAFRFL (Safety Analysis Record-Level Flag) variable enables inclusion/exclusion of data per reporting needs. Since the focus of this scenario is on the second participation, adding SAFRFL='Y' to the subset condition will include the required participation on the corresponding table, listing, or figure (TLF). APERIOD corresponds with the xx in the ADSL.TRTxxP assigned to the record-level TRTP based on the observation/start date of the record. When each participation is a single period, APERIOD can be used to denote the participation and could also be used in the subset condition. The analysis day variables are flexible to use whichever anchor date is defined in SAP. In this example, the second participation treatment start will be the anchor so analysis day values are negative for the first participation. This will be different from SDTM study day values as shown above in the AE example.

The Treatment Emergent and On Treatment flags in ADAE are based on the second participation treatment start date per analysis needs and thus ADAE.TRTEMFL will not match SUPPAE.AETRTEM values. If there are any differences between these variable values, it should be explained in the Analysis Data Reviewer's Guide (ADRG). In addition, records with null TRTEMFL values will need to be explained in the ADRG, along with the negative day values.

*adae.xpt*

| Row | USUBJID | SUBJID | AEDECOD | TRTSDT | ASTDT | AENDT |
|---|---|---|---|---|---|---|
| 1 | AAA 001 1001 | 1001 | Anaemia | 1-Sep-2023 | 3-Jan-2023 | 5-Jan-2023 |
| 2 | AAA 001 1001 | 1001 | Irritable bowel syndrome | 1-Sep-2023 | 10-Jan-2023 | 15-Jan-2023 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 3 | AAA 001 1001 | 1003 | Headache | 1-Sep-2023 | 10-Aug-2023 | 12-Aug-2023 |
| 4 | AAA 001 1001 | 1003 | Cough | 1-Sep-2023 | 10-Oct-2023 | 20-Oct-2023 |

| Row | ASTDY | AENDY | APERIOD | TRTP | TRTEMFL | ONTRTFL | SAFRFL |
|---|---|---|---|---|---|---|---|
| 1 | -240 | -238 | 1 | STUDY DRUG 10mg | | | |
| 2 | -233 | -228 | 1 | STUDY DRUG 10mg | | | |
| 3 | -21 | -19 | 2 | STUDY DRUG 30mg | | | Y |
| 4 | 40 | 50 | 2 | STUDY DRUG 30mg | Y | | Y |

In BDS datasets like ADVS, baseline is defined based on the selected participation for analysis. In cumulative SDTM where DM.RFSTDTC comes from the first participation, the VSLOBXFL will likewise be defined based on that participation regardless of analysis needs. However ADVS.ABLFL is more flexible so can be defined based on the second participation in the below example. ANLzzFL is also analysis-based so ANL01FL is only populated for the second participation in the below example. These variables, along with SAFRFL provide the required subsets for TLF analysis.

*advs.xpt*

| Row | USUBJID | SUBJID | TRTSDT | PARAMCD | AVISIT | AVAL | BASE | ADT | ADY |
|---|---|---|---|---|---|---|---|---|---|
| 1 | AAA 001 1001 | 1001 | 1-Sep-2023 | DIABP | SCREENING | 75 | | 3-Dec-2022 | -271 |
| 2 | AAA 001 1001 | 1001 | 1-Sep-2023 | DIABP | DAY 1 | 72 | | 1-Jan-2023 | -242 |
| 3 | AAA 001 1001 | 1001 | 1-Sep-2023 | DIABP | 1 MONTH | 78 | | 1-Feb-2023 | -211 |
| 4 | AAA 001 1001 | 1001 | 1-Sep-2023 | DIABP | 3 MONTHS | 80 | | 30-Mar-2023 | -154 |
| 5 | AAA 001 1001 | 1003 | 1-Sep-2023 | DIABP | SCREENING | 81 | 84 | 5-Aug-2023 | -26 |
| 6 | AAA 001 1001 | 1003 | 1-Sep-2023 | DIABP | DAY 1 | 84 | 84 | 1-Sep-2023 | 1 |
| 7 | AAA 001 1001 | 1003 | 1-Sep-2023 | DIABP | 1 MONTH | 83 | 84 | 1-Oct-2023 | 31 |
| 8 | AAA 001 1001 | 1003 | 1-Sep-2023 | DIABP | 3 MONTHS | 82 | 84 | 30-Nov-2023 | 91 |

| Row | AVISIT | ABLFL | ANL01FL | SAFRFL | APERIOD | TRTP |
|---|---|---|---|---|---|---|
| 1 | SCREENING | | | | 1 | STUDY DRUG 10mg |
| 2 | DAY 1 | | | | 1 | STUDY DRUG 10mg |
| 3 | 1 MONTH | | | | 1 | STUDY DRUG 10mg |
| 4 | 3 MONTHS | | | | 1 | STUDY DRUG 10mg |
| 5 | SCREENING | | | Y | 2 | STUDY DRUG 30mg |
| 6 | DAY 1 | Y | Y | Y | 2 | STUDY DRUG 30mg |
| 7 | 1 MONTH | | Y | Y | 2 | STUDY DRUG 30mg |
| 8 | 3 MONTHS | | Y | Y | 2 | STUDY DRUG 30mg |

Two alternative approaches (subset vs. cumulative) were discussed in this scenario to handle re-enrollment where a single participation needs to be analyzed, but these do not hold well when all participations need to be summarized together. A more efficient way to handle these multiple participations using the DC and ADPL domains will be discussed in the next scenario.

## SCENARIO TWO – ALL PARTICIPATIONS SUMMARIZED

The second scenario is one where all participations are summarized. This may occur in a planned re-enrollment protocol design where all enrollments need to be analyzed, but also may occur for unplanned multiple enrollments, which for various reasons, it would be deemed necessary to include all enrollments in the analysis. Since some of the aspects of the below approach are not a part of published CDISC standards at this time, it is recommended to discuss it with the applicable regulatory agency prior to submission. It would need to be clearly described in the data reviewer's guides in order to facilitate regulatory review.

### SDTM Approach

In cases of multiple participations, the DC domain is the proposed solution to satisfy the TCG request of a custom domain structured like the DM domain to represent multiple participations in a single study. Under this scenario, where all participations are included in analysis, the DC domain (or similar) is necessary. The structure of the proposed draft DC domain is almost identical to the DM domain with the exception of the additional variable DCSEQ which provides the sequential order of the participations. This participant enrolled two times within the study so has two records in DC with the same USUBJID, with SUBJID and DCSEQ identifying the unique participations.

*dc.xpt*

| USUBJID | SUBJID | DCSEQ | RFSTDTC | RFENDTC | RFICDTC | BRTHDTC | ARMCD | ARM |
|---|---|---|---|---|---|---|---|---|
| AAA 001 1001 | 1001 | 1 | 2023-01-01 | 2023-03-30 | 2022-12-03 | 1998-01-01 | A | STUDY DRUG 10 mg |
| AAA 001 1001 | 1003 | 2 | 2023-09-01 | 2023-11-30 | 2023-08-05 | 1998-01-01 | C | STUDY DRUG 30 mg |

The DM domain would still include only one record for each participant. In the case of multiple participations, the TCG recommends that DM should contain the primary participation as determined by the sponsor. If all enrollments are summarized, choosing a primary one may be challenging. One can be chosen arbitrarily, for example the first, or all enrollments can be collapsed into one DM record so a participant's entire participation in the study is summarized. When deciding how to summarize or select a participation for the DM domain, keep in mind how this impacts EPOCH and study day derivations in the other SDTM domains. These SDTM variables don't need to be used for analysis, however they can aid in interpretation and review of the tabulation datasets.

The example displayed below is where all participations were summarized into one record in DM. The first SUBJID is selected, and RFSTDTC and RFICDTC come from the first participation where DCSEQ=1. RFENDTC however is populated with the value from DC.RFENDTC where DCSEQ=2. Setting ARMCD to AC could be misleading as this is not a crossover study with a planned sequence of treatments, therefore the first ARMCD, A, is chosen.

*dm.xpt*

| USUBJID | SUBJID | RFSTDTC | RFENDTC | RFICDTC | BRTHDTC | ARMCD | ARM |
|---|---|---|---|---|---|---|---|
| AAA 001 1001 | 1001 | 2023-01-01 | 2023-11-30 | 2022-12-03 | 1998-01-01 | A | STUDY DRUG 10 mg |

The adverse events domain under this scenario contains all AEs collected for both participations. The inclusion of SUBJID is necessary to distinguish which events are associated with each participation. Since DM.RFSTDTC comes from the first participation in this example, AESTDY and AEENDY are all relative to the first participation treatment start, which is less intuitive for the data of the second participation. EPOCH is more flexible, so is defined within each participation rather than relative to only the first, which is possible with the DC domain. In the Scenario One cumulative approach above, EPOCH was defined based on the first participation because the DC domain was not implemented. Additionally, SUPPAE.AETRTEM is defined within each participation so the AE that occurred during the second participation but prior to second treatment start date will not have AETRTEM populated.

*ae.xpt*

| USUBJID | SUBJID | AEDECOD | AESTDTC | AEENDTC | AESTDY | AEENDY | EPOCH | SUPPAE. AETRTEM |
|---|---|---|---|---|---|---|---|---|
| AAA 001 1001 | 1001 | Anaemia | 2023-01-03 | 2023-01-05 | 3 | 5 | TREATMENT | Y |
| AAA 001 1001 | 1001 | Irritable bowel syndrome | 2023-01-10 | 2023-01-15 | 10 | 15 | TREATMENT | Y |
| AAA 001 1001 | 1003 | Headache | 2023-08-10 | 2023-08-12 | 222 | 224 | SCREENING | |
| AAA 001 1001 | 1003 | Cough | 2023-10-10 | 2023-10-20 | 283 | 293 | FOLLOW-UP | Y |

The VS domain will also contain all records for both participations, using SUBJID to distinguish which records are associated with which participation. VSDY, similar to the example above in AE, is numbered relative to the first participation. The SDTM IG v3.3 variable, VSLOBXFL is defined as the last non-missing record prior to DM.RFXSTDTC, and therefore only one record can be marked as Y within USUBJID[6]. The IG states that the authoritative baseline for statistical analysis is found in ADaM, so differences between this baseline definition and the analysis baseline definition are permitted. EPOCH again is defined within each participation.

*vs.xpt*

| USUBJID | SUBJID | VSTESTCD | VSORRES | VSDTC | VSDY | VSLOBXFL | VISIT | EPOCH |
|---|---|---|---|---|---|---|---|---|
| AAA 001 1001 | 1001 | DIABP | 75 | 2022-12-03 | -28 | | SCREENING | SCREENING |
| AAA 001 1001 | 1001 | DIABP | 72 | 2023-01-01 | 1 | Y | DAY 1 | TREATMENT |
| AAA 001 1001 | 1001 | DIABP | 78 | 2023-02-01 | 32 | | 1 MONTH | FOLLOW-UP |
| AAA 001 1001 | 1001 | DIABP | 80 | 2023-03-30 | 89 | | 3 MONTHS | FOLLOW-UP |
| AAA 001 1001 | 1003 | DIABP | 81 | 2023-08-05 | 217 | | SCREENING | SCREENING |
| AAA 001 1001 | 1003 | DIABP | 84 | 2023-09-01 | 244 | | DAY 1 | TREATMENT |
| AAA 001 1001 | 1003 | DIABP | 83 | 2023-10-01 | 274 | | 1 MONTH | FOLLOW-UP |
| AAA 001 1001 | 1003 | DIABP | 82 | 2023-11-30 | 334 | | 3 MONTHS | FOLLOW-UP |

*Note: For VSTESTCD=DIABP, VSTEST is 'Diastolic Blood Pressure'.*
*Note: The SDTM IG v3.3 variable, VSLOBXFL is a new required variable defined as the last non-missing record before exposure. VSBLFL became permissible but was required in IG 3.2.*

### ADaM Approach

To handle multiple participations in ADaM, the ADPL dataset has been proposed[7], which is similar to ADSL, but like DC can contain multiple rows per USUBJID. An NSV, ORDER, is used to indicate the sequence of participations, similar to DC.DCSEQ. This dataset could be used in place of ADSL to create other ADaM datasets, although ADSL would still be a required dataset for a submission package. The benefit of using ADPL for creating other ADaM datasets is it contains each combination of USUBJID and SUBJID so is more straightforward to use to merge or join with SDTM domains.

*adpl.xpt*

| USUBJID | SUBJID | ORDER | TRT01P | TRTSDT | TRTEDT | EOSDT | RFICDTC |
|---|---|---|---|---|---|---|---|
| AAA 001 1001 | 1001 | 1 | STUDY DRUG 10 mg | 01JAN2023 | 01FEB2023 | 30MAR2023 | 2022-12-03 |
| AAA 001 1001 | 1003 | 2 | STUDY DRUG 30 mg | 01SEP2023 | 01OCT2023 | 20NOV2023 | 2023-08-05 |

If ADPL is used to create other ADaM datasets, subject-level analysis timing variables should be defined in ADPL. Since each unique participation has separate record, fewer variables need to be defined to represent periods and phases compared to ADSL. Thinking through to how records in other datasets would have APHASE and APERIOD populated, it would be more useful to have participation number also identified in the APHASE values, i.e. FIRST SCREENING, FIRST TREATMENT, FIRST FOLLOW-UP and so on. In this way, the same variables, APHASE1, APHASE2, and APHASE3 could be used across all participations. If ADPL is not used to create other ADaM datasets, then the APHASEw variables would not be needed.

*adpl.xpt analysis timing variables*

| Row | USUBJID | SUBJID | ORDER | APHASE1 | PH1SDT | PH1EDT |
|---|---|---|---|---|---|---|
| 1 | AAA 001 1001 | 1001 | 1 | FIRST SCREENING | 03DEC2022 | 31DEC2022 |
| 2 | AAA 001 1001 | 1003 | 2 | SECOND SCREENING | 05AUG2023 | 31AUG2023 |

| Row | APHASE2 | PH2SDT | PH2EDT | APHASE3 | PH3SDT | PH3EDT |
|---|---|---|---|---|---|---|
| 1 | FIRST TREATMENT | 01JAN2023 | 01FEB2023 | FIRST FOLLOW-UP | 02FEB2023 | 30MAR2023 |
| 2 | SECOND TREATMENT | 01SEP2023 | 01OCT2023 | SECOND FOLLOW-UP | 02OCT2023 | 20NOV2023 |

ADSL on the other hand would need to be mapped like DM, with one record per USUBJID and all participations captured as new variables (horizontally) rather than as new rows (vertically). Even though ADSL is a required dataset for a submission, if ADSL is not used for creating other ADaM datasets, the decision on which variables to include depends on what would be most useful for review and not necessarily driven by analysis need. For example, the periods and phases for each participation may not need to be defined as they could become very cumbersome. In this example there would be six sets of APHASEw variables (APHASEw, PHwSDT, PHwEDT), for a total of 18 variables for phases alone.

ADSL in a multiple participation model could utilize a mix of existing ADaM variables and additional NSVs to represent all the participations. Existing variables like TRTxxP/TRTxxA/TRxxSDT/TRxxEDT and APxxSDT/APxxEDT which allow for the concept of multiple treatment/analysis periods could be used. These variables are defined for trial designs with multiple treatment periods, like crossover or an extension study, but they are valid also for multiple participations as these are new treatment/analysis periods. NSV would be useful for example to represent the SUBJID of each participation. In this example the first participation is SUBJID, second participation is SUBJID2. Another NSV, MULENRFL, is used as an identifier flag for a multiple enrolled participant. Depending on what is considered necessary to include in ADSL, many other variables may need to be defined for each participation, for example any baseline variables, covariates, or population flags.

*adsl.xpt*

| Row | USUBJID | SUBJID | SUBJID2 | TRTSDT | TRTEDT | TRT01P | TR01SDT |
|---|---|---|---|---|---|---|---|
| 1 | AAA 001 1001 | 1001 | 1003 | 01JAN2023 | 01OCT2023 | STUDY DRUG 10 mg | 01JAN2023 |

| Row | TR01EDT | TRT02P | TR02SDT | TR02EDT | MULENRFL |
|---|---|---|---|---|---|
| 1 | 01FEB2023 | STUDY DRUG 30 mg | 01SEP2023 | 01OCT2023 | Y |

The ADAE dataset is created by merging ADPL and SDTM.AE, so that the subject-level variables are populated for the corresponding participation. ASTDY and AENDY then are based on TRTSDT within each participation, so the values differ from SDTM variables AESTDY and AEENDY. APHASE is defined based on the applicable APHASEw in ADPL by comparing ASTDT with each PHwSDT and PHwEDT. APERIOD and TRTA are populated based on the participation, rather than whether

the particular record is on treatment, which is a sponsor decision. The ADaM Structure for Occurrence Data (OCCDS) IG v1.1 defines new variables TREMwFL and ONTRxxFL which can be used for cases with multiple treatment periods or analysis needs[8]. These variables could be useful in this situation, however TRTEMFL and ONTRTFL can still be defined within each participation, and APERIOD used to select a particular period, if needed, for analysis. This allows flexibility to either select a single participation for analysis using APERIOD, or to analyze all participations together in the same table.

*adae.xpt*

| Row | USUBJID | SUBJID | AEDECOD | TRTSDT | TRTEDT | ASTDT | AENDT |
|---|---|---|---|---|---|---|---|
| 1 | AAA 001 1001 | 1001 | Anaemia | 01JAN2023 | 01FEB2023 | 03JAN2023 | 05JAN2023 |
| 2 | AAA 001 1001 | 1001 | Irritable bowel syndrome | 01JAN2023 | 01FEB2023 | 10JAN2023 | 15JAN2023 |
| 3 | AAA 001 1001 | 1003 | Headache | 01SEP2023 | 01OCT2023 | 10AUG2023 | 12AUG2023 |
| 4 | AAA 001 1001 | 1003 | Cough | 01SEP2023 | 01OCT2023 | 10OCT2023 | 20OCT2023 |

| Row | ASTDY | AENDY | APERIOD | TRTA | APHASE | TRTEMFL | ONTRTFL |
|---|---|---|---|---|---|---|---|
| 1 | 3 | 5 | 1 | STUDY DRUG 10 mg | FIRST TREATMENT | Y | Y |
| 2 | 10 | 15 | 1 | STUDY DRUG 10 mg | FIRST TREATMENT | Y | Y |
| 3 | -21 | -19 | 2 | STUDY DRUG 30 mg | SECOND SCREENING | | |
| 4 | 40 | 50 | 2 | STUDY DRUG 30 mg | SECOND FOLLOW-UP | Y | |

The ADVS domain would be created by merging the ADPL and VS datasets, so ADY would be defined within each participation. ABLFL can be defined within each participation, as long as the BASETYP variable is included to distinguish each unique definition of baseline (in this case, participation). Analysis flags could be used across all participations, since APERIOD could be used to select a particular participation for analysis, so all participations could be summarized in the same table, similar to the approach with TRTEMFL and ONTRTFL from ADAE.

*advs.xpt*

| Row | USUBJID | SUBJID | TRTSDT | PARAMCD | AVAL | BASE | ADT | ADY | ABLFL |
|---|---|---|---|---|---|---|---|---|---|
| 1 | AAA 001 1001 | 1001 | 01JAN2023 | DIABP | 75 | 72 | 03DEC2022 | -28 | |
| 2 | AAA 001 1001 | 1001 | 01JAN2023 | DIABP | 72 | 72 | 01JAN2023 | 1 | Y |
| 3 | AAA 001 1001 | 1001 | 01JAN2023 | DIABP | 78 | 72 | 01FEB2023 | 32 | |
| 4 | AAA 001 1001 | 1001 | 01JAN2023 | DIABP | 80 | 72 | 30MAR2023 | 89 | |
| 5 | AAA 001 1001 | 1003 | 01SEP2023 | DIABP | 81 | 84 | 05AUG2023 | -26 | |
| 6 | AAA 001 1001 | 1003 | 01SEP2023 | DIABP | 84 | 84 | 01SEP2023 | 1 | Y |
| 7 | AAA 001 1001 | 1003 | 01SEP2023 | DIABP | 83 | 84 | 01OCT2023 | 31 | |
| 8 | AAA 001 1001 | 1003 | 01SEP2023 | DIABP | 82 | 84 | 30NOV2023 | 91 | |

| Row | BASETYP | AVISIT | ANL01FL | APERIOD | TRTA | APHASE |
|---|---|---|---|---|---|---|
| 1 | PERIOD 01 | SCREENING | | 1 | STUDY DRUG 10 mg | FIRST SCREENING |
| 2 | PERIOD 01 | DAY 1 | Y | 1 | STUDY DRUG 10 mg | FIRST TREATMENT |
| 3 | PERIOD 01 | 1 MONTH | Y | 1 | STUDY DRUG 10 mg | FIRST FOLLOW-UP |
| 4 | PERIOD 01 | 3 MONTHS | Y | 1 | STUDY DRUG 10 mg | FIRST FOLLOW-UP |
| 5 | PERIOD 02 | SCREENING | | 2 | STUDY DRUG 30 mg | SECOND SCREENING |
| 6 | PERIOD 02 | DAY 1 | Y | 2 | STUDY DRUG 30 mg | SECOND TREATMENT |
| 7 | PERIOD 02 | 1 MONTH | Y | 2 | STUDY DRUG 30 mg | SECOND FOLLOW-UP |
| 8 | PERIOD 02 | 3 MONTHS | Y | 2 | STUDY DRUG 30 mg | SECOND FOLLOW-UP |

## Tables, Figures and Listings

In tables and figures, multiple participations can be summarized together or separately depending on analysis needs. Thoughtful consideration should be made to the end use when designing ADaM to be analysis-ready. Multiple participations do challenge some traditional table designs such as demography or disposition, which are typically thought of as 'by participant' rather than 'by participation'. For example, if there are multiple participations analyzed together, and it happens that both are the same treatment arm, then a participant could be double counted in the same arm. Careful discussion is required with the study team to make sure all understand the implications of the decided analysis path.

Below is a table of treatment emergent AEs for the above discussed ADAE example which analyzes both participations in the same table. Since both participations have different treatment arms, the analysis would likely be straightforward when USUBJID is grouped by treatment arm since it would still result in the correct counts, whether or not USUBJID or SUBJID is used as the unique key.

*AE table double counting multiple participations with different treatment arms*

| System Organ Class<br>Preferred Term | Study Drug 10 mg<br>(N=10)<br>n (%) | Study Drug 30 mg<br>(N=10)<br>n (%) |
|---|---|---|
| Any adverse event | 1 (10) | 1 (10) |
| Blood and lymphatic system disorders | 1 (10) | 0 |
| Anaemia | 1 (10) | 0 |
| Respiratory, thoracic and mediastinal disorders | 0 | 1 (10) |
| Cough | 0 | 1 (10) |

If there are multiple participations within the same treatment arm analyzed in the same table, table programming needs to be updated to use SUBJID instead of USUBJID as a unique key. For the below example, the same participant re-enrolls to the same treatment arm and has an AE during each participation. If USUBJID is used as the key the data could be summarized as one AE for one participant, rather than two different AEs for two participations.

*adae.xpt*

| USUBJID | SUBJID | AEBODSYS | AEDECOD | APERIOD | TRTA | APHASE | TRTEMFL | ONTRTFL |
|---|---|---|---|---|---|---|---|---|
| AAA 001 1002 | 1002 | Nervous system disorders | Headache | 1 | STUDY DRUG 10 mg | FIRST TREATMENT | Y | Y |
| AAA 001 1002 | 1005 | Nervous system disorders | Headache | 2 | STUDY DRUG 10 mg | SECOND TREATMENT | Y | Y |

*AE table using USUBJID as unique key*

| System Organ Class<br>Preferred Term | Study Drug 10 mg<br>(N=10)<br>n (%) |
|---|---|
| Any adverse event | 1 (10) |
| Nervous system disorders | 1 (10) |
| Headache | 1 (10) |

*AE table using SUBJID as unique key*

| System Organ Class<br>Preferred Term | Study Drug 10 mg<br>(N=11)<br>n (%) |
|---|---|
| Any adverse event | 2 (18) |
| Nervous system disorders | 2 (18) |
| Headache | 2 (18) |

For listings, SUBJID would need to be included in addition to USUBJID in order to distinguish data from each participation. If each participation has a different treatment, then perhaps the listings could have a page break for each treatment, but the same participants' data would not be presented together when this approach is taken.

## SUBMISSION CHALLENGES

Each scenario will have particular challenges when it comes to submission. The first approach for Scenario One where data is subset to include one participation for analysis might have a challenge since the Trial Design Model datasets will include all participations per protocol while the subject-level SDTM domains will only contain the subset of data needed for analysis and reporting. If there is a gap for example, between the arms represented in the TA domain and the arms in the DM domain, the DC domain could be included to help clarify the full participation details for all participants. Similarly, the second cumulative approach for Scenario One could also benefit from the DC domain for clarification. A DC domain is therefore recommended in situations of multiple participations even if the analysis does not require it. However, this approach needs to be discussed with the applicable regulatory agency prior to submission.

With the second scenario, where all participations are included, a domain like DC is expected. As discussed above, it is essential to include SUBJID in all subject-level domains, but this needs to be explained in the Clinical Study Data Reviewer's Guide (cSDRG) since it will be marked as a conformance issue during P21 validation. Additionally, all cases of multiple participations should be described in the cSDRG. The ADPL dataset should be clearly described in the ADRG since it is not a typical ADaM dataset, and if it is used in place of ADSL to create other ADaM datasets, this should also be clearly explained. For Scenario Two, any approach would likely require discussion with the applicable regulatory agency prior to submission.

When multiple participation occurs, whether planned as part of the protocol design, or unplanned without the sponsor's prior knowledge, particular care should be taken in case a study has interim data cuts for submission, to ensure that USUBJID is not changing across those submissions for the same study and product. If a submission does occur and USUBJID is later found to be incorrect because accurate prior participation was not initially entered, this could compromise the quality and reliability of the study database. It is the sponsor's responsibility to ensure the completeness and accuracy of data. If a situation occurs where a previous submission contains an incorrect USUBJID, the sponsor must inform the regulatory agency and discuss what the proper remedy would be.

## CONCLUSION

Multiple participations pose challenges to our standard data collection, SDTM and ADaM datasets, and analysis and reporting. Accurate data collection to be able to assign the correct USUBJID is critical in being able to link participations together, whether planned or unplanned in protocol design. As study designs become more complex, the need for updating our standards becomes more apparent. The draft proposal of a DC domain forms the basis of this updated approach, but requires many downstream considerations in other SDTM domains, ADaM generation, and analysis which are yet to be standardized.

In this paper several approaches and scenarios were discussed, each with their own advantages and disadvantages. In the first approach to Scenario One where data is subset to a particular participation, the advantage is the SDTM, ADaM and TLFs can follow a simpler approach. In the second approach, all participation data is included however only one participation data was needed for analysis. The benefit of this approach is it gives an opportunity to list all prior participations in the listings however DM and ADSL become more complicated and have less traceability when DC domain is not implemented. The advantage of Scenario Two is when all participations need to be analyzed, it gives more flexibility to present multiple participations because of the DC and ADPL domains. However, this design is more complex to implement as it challenges our current structures and standards and requires discussion with any applicable regulatory agency ahead of time. Updated CDISC standards to clarify how to handle these situations with multiple participation in a single study will be a critical step in moving this topic forward and we look forward to what the CDISC Multiple Subject Instances Team will release in the future.

## REFERENCES


[1] Clinical Data Interchange Standards Consortium (CDISC). Study Data Tabulation Model Implementation Guide. Available at: https://www.cdisc.org/standards/foundational/sdtmig.

[2] Clinical Data Interchange Standards Consortium (CDISC). Analysis Data Model Implementation Guide. Available at: https://www.cdisc.org/standards/foundational/adam.

[3] US Department of Health and Human Services, Food and Drug Administration, Center for Biologics Evaluation and Research and Center for Drug Evaluation and Research. Study Data Technical Conformance Guide. Version 5.5. October 2023. Available at: https://www.fda.gov/media/153632/download.

[4] Éanna Kiely. CDISC. Updates on Handling Multiple Enrollments and Screenings Subjects in SDTM. October 2020. Available at: https://wiki-test.cdisc.org/download/attachments/113577697/Updates%20on%20Handling%20Multiple%20Enrollments%20and%20Screenings%20Subjects%20in%20SDTM%20-%20ClinBuild%20-%20Eanna%20Kiely%20-%20V1.0.pdf?version=1&modificationDate=1602659826225&api=v2.

[5] Clinical Data Interchange Standards Consortium (CDISC). Study Data Tabulation Model Implementation Guide: Human Clinical Trials. Version 3.2, November 2013. Available at: https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-v3-2.

[6] Clinical Data Interchange Standards Consortium (CDISC). Study Data Tabulation Model Implementation Guide: Human Clinical Trials. Version 3.3, November 2018. Available at: https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-v3-3.

[7] Elizabeth Dennis, Grace Fawcett, Sandra Minjoe. ADaM Datasets with Multiple Participations per Subject. PharmaSUG 2023. Available at https://www.pharmasug.org/proceedings/2023/SS/PharmaSUG-2023-SS-146.pdf

[8] Clinical Data Interchange Standards Consortium (CDISC). ADaM Structure for Occurrence Data (OCCDS) Implementation Guide. Version 1.1, November 2021. Available at: https://www.cdisc.org/standards/foundational/adam/adam-structure-occurrence-data-implementation-guide-v1-1.


## ACKNOWLEDGMENTS


We would like to acknowledge Helen Yang, Lisa Houterloot, Sandra VanPelt Nguyen and Liping Zhang for their review of this paper.


## CONTACT INFORMATION

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


Author Name: Emily Bernstein-Hanlon
Company: Pfizer
Email: emily.bernstein-hanlon@pfizer.com

Author Name: Sirisha Davuluri
Company: Pfizer
Email: Sirisha.Davuluri@pfizer.com