

Considerations When Representing Multiple Subject Enrollments in SDTM

Kristin Kelly, Pinnacle 21
Mike Hamidi, CDISC

ABSTRACT

In clinical trials, it has become more common for a study design to allow subjects to re-enroll in the same study or subsequent studies within a submission. For studies that allow subjects to re-screen for the same study, it may be difficult to determine how to represent the data for multiple enrollments in SDTM. There are a number of approaches seen in industry, but many pose issues. An example of this is creating multiple records in DM with the same USUBJID to represent each enrollment. Though this may seem the most straightforward approach, many tools used at FDA are configured to expect one record per subject and thus, the data may not readily load into their tools. Another approach is to assign different USUBJID values for the same subject within a study and across studies. This also creates issues for review because it is difficult to track the same subject across studies. This paper will focus on examples from industry as well as proposed solutions for representing this data in SDTM.

INTRODUCTION

When a subject is permitted to re-screen or re-enroll in a clinical trial or program, there are few scenarios that can occur depending on the study protocol. These are the following listed from most to least common:

- 1) Subjects that enroll in more than one study.
- 2) Subjects that are screened more than once within the same study.
- 3) Subjects that enroll more than once and are treated within the same study.

There are also other scenarios that can happen that are not outlined in the protocol such as a subject that changes sites during a trial with the approval of the sponsor. Another might be a subject that enrolls at different sites unbeknownst to the investigator and would be in violation of the exclusion criteria for the trial.

Regardless of whether the re-enrollment is planned or unplanned, it is at the discretion of the sponsor how to collect enough data that will support past instances of a subject in a study, e.g. Informed consent dates, previous STUDYID and SUBJID, etc. Please note that the actual data management process of how to do this is beyond the scope of this paper.

Currently, the SDTMIG does not have guidance or standards in place to handle this scenario of multiple subject instances. Because of this, sponsors have chosen different ways to represent this data and some are problematic because the data is not SDTM-compliant or does not load into regulatory review tools. This paper will provide examples on how to model the collected data within the SDTM standard as it is now. This is intended to guide sponsors on what to do for their imminent submissions to ensure conformance until the SDTM standard is updated.

BACKGROUND

Per the FDA Technical Conformance Guide (TCG), each subject should be assigned the same unique identifier (USUBJID) across the entire submission even if they participate in more than one study. This concept is also stated in the SDTMIG v3.2. In addition to USUBJID, all subjects should have a SUBJID assigned that is unique within the study. It is also stated in the TCG that 'if a subject is screened more than once in a trial, then the subject's SUBJID should be different' for each screening attempt.¹

Though it may seem at first glance that there is a one-to-one relationship between USUBJID and SUBJID, this is not the case when collecting data for re-screened subjects or subjects that enroll in multiple studies. As described above, even if there are multiple instances of the subject across the submission, the same USUBJID value should be maintained but the SUBJID values should be different. To better understand this concept, it may be helpful to think of the same USUBJID being assigned to the same person or 'warm body' regardless of how many times they appear in a submission. Each person will still have the same characteristics each time they are screened or enrolled, i.e. the same name, initials, race, ethnicity, sex, etc. And then SUBJID is simply a number assigned to that same person that is different each time they screen for a clinical trial.

Another consideration to bear in mind is that the TCG states in Section 4.1.1.3, 'in the DM domain, each subject should have only one single record per study.'¹ The SDTMIG v3.2 contains similar language in Assumption #2 for the DM domain, that if a subject changes sites (or participates more than once) in the same study, 'the sponsor must decide how to populate variables such as USUBJID, SUBJID, and SITEID based on their operational and analysis needs, but only one DM record should be submitted for the subject.'²

Combining these concepts when dealing with multiple enrollments can become difficult when it comes time to prepare the submission data in a manner that is clear, SDTM-conformant, and conducive to easily load into review tools.

WHAT DO SPONSORS DO (THAT MIGHT NOT WORK SO WELL)?

It is a common misconception that most study designs do not allow for multiple enrollments, especially within the same study. Also, that it happened less than 20% of the time, if one was going to apply the 80/20 rule. In actuality, these scenarios occur often enough that sponsors have asked for submission guidance from both regulatory agencies as well as standards development organizations, e.g. CDISC, on how to model this data. The following are some examples seen in industry for representing multiple enrollments.

SCENARIO #1 – SUBJECTS THAT RE-SCREEN OR RE-ENROLL WITHIN THE SAME STUDY

When sponsors design studies where subjects are permitted to re-screen for the same study until they are eligible for enrollment, it typically depends on the trial indication and the availability of subjects. If it is a diabetes trial, perhaps some subjects do not meet the criteria for glucose levels at an initial screening visit but on another day, they are able to be re-tested and meet the requirements of the study. Another example would be when the trial indication is a rare disease where the number of subjects with the condition may be limited and thus, the study design allows either re-screening or in some cases, re-enrollment after they gone through the trial once.

In March 2014, there was a PhUSE Working Group³ that was formed to help create solutions for modeling this type of data in SDTM. The proposed solution was clear, straightforward, and implementable but it would require an update to the SDTM standard in order to be conformant. Since then, some sponsors have implemented this proposal in their submission data.

The recommendations were the following:

- 1) Design the CRF to handle collection of data for previous screenings/enrollments
- 2) Create more than one record in DM for each screening/enrollment event per subject
 - a. The subject would have the same USUBJID but the SUBJID values would be different
- 3) Add SUBJID to the general observation class domains as a permissible Identifier in order to differentiate records specific to each instance in the trial³

In an example study, XYZ1047, for an investigational drug developed for treatment of Hepatitis C, subjects were able to re-screen more than once until they became eligible for enrollment. This was a Phase III, global study where 3500 subjects were planned to be enrolled. Subject 505, an identifier assigned at the initial visit, failed screening twice before being enrolled in the study. This subject was

then subsequently randomized and treated. Based on the recommendations presented above, this is what the sponsor submitted for this subject in DM. Please note that only relevant variables for the example are included:

STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTC	RFENDTC	RFICDTC	SITEID	ARMCD
XYZ1047	DM	XYZ1047-002-0725	002-0505				2017-04-23	002	SCRNFAIL
XYZ1047	DM	XYZ1047-002-0725	002-0618				2017-04-30	002	SCRNFAIL
XYZ1047	DM	XYZ1047-002-0725	002-0725	2017-05-23	2018-10-23		2017-05-14	002	XYZ1047

In the DM domain, the subject was assigned the same USUBJID for each record which was also used in all subject-level domains. The SUBJID values for each screening event are different as per the guidance in the TCG and the SDTMIG. Only the DM record where the subject was randomized and treated have RFSTDTC/RFENDTC populated as expected.

Applying this approach in DM triggers the following validation rule⁴:

FDA Validator Rule ID	FDA Validator Message	Publisher	Publisher ID	Business or Conformance Rule validated	Validator Rule
SD0083	Duplicate USUBJID in DM within STUDYID	CDISC	CG0151	Identifier used to uniquely identify a subject across all studies	The value of Unique Subject Identifier (USUBJID) variable must be unique for each subject across all trials in the submission. In the Demographics (DM) domain, there should only be one record per subject.

This rule is looking for more than one record in the DM domain within one study. It does not currently check across studies. The sponsor fully explained that the records flagged by the rule were re-screened subjects and that the approach that was followed was one proposed by the PhUSE WG. The approach taken seems straightforward and clean even though it is not in conformance with the SDTMIG and the TCG, but the issue is explained appropriately in the reviewer’s guide. Also, in this example, the sponsor chose to submit the corresponding SUBJID to differentiate data collected during each screening event in SUPPQUAL domains rather than adding it to the parent subject-level domains as this is not a permitted variable in other domains except for DM.

Everything is in place until.....the NDA is submitted and the study is loaded into FDA review tools. The tools are expecting one record per subject in DM and the study data will not be accepted into the tool when there are duplicate USUBJIDs in DM. This issue does not make the data unreviewable, but it does add time to the review schedule because alternative approaches need to be used to review the data outside the tools or a workaround provided to be able to load the data. Bottom line: Unless the sponsor has checked with their review division: Don’t Do This! At least until the SDTM standard, the TCG, and the regulatory review tools are updated to handle this data modeling.

SCENARIO #2 – SUBJECTS THAT ENROLL IN MORE THAN ONE STUDY

The scenario where subjects enroll in more than one study whether it be an extension of a base trial or another primary study is the one seen most often when handling multiple enrollments. As described earlier, current guidance holds that each subject or ‘warm body’ be assigned the same USUBJID regardless of how many studies in which they participate. While this seems an easy tenet to follow, some

sponsors have difficulty either tracking previous enrollments or do not have the ability to go back to assign a unique identifier due to tool restrictions or timeline constraints. These two reasons may or may not be mutually exclusive. There are cases where the studies being submitted were conducted before CDISC standards were required for submissions and this type of information was not collected. Or the study drug was purchased from another sponsor and not all the information needed to track these subjects is available.

The following is an example where the sponsor was unable to assign a unique USUBJID for subjects that were in multiple studies because there was no data collected for previous enrollments in each study. Subject 202 enrolled in and completed Study ABC709. This subject was then eligible to enroll in another study for this treatment, ABC710. The subject was treated and completed this study as well.

Study ABC709 – dm.xpt

STUDYID	USUBJID	SUBJID	RFSTDTC	RFENDTC	BRTHDTC	SEX	RACE	ARMCD
ABC709	ABC709-08-0202	08-0202	2016-10-08	2017-04-25		1955	F	WHITE	ABC_DRUG

Study ABC710 – dm.xpt

STUDYID	USUBJID	SUBJID	RFSTDTC	RFENDTC	BRTHDTC	SEX	RACE	ARMCD
ABC710	ABC710-14-0560	14-560	2017-10-14	2018-08-23		1955	F	WHITE	ABC_DRUG

Because this subject was not assigned the same USUBJID, the reviewer did not readily know that each study enrollment was the same person. The only way to determine this would be to compare the subject’s characteristics such as BRTHDTC (where just having the year of birth collected is not helpful), SEX, RACE, ETHNIC etc. But even comparing these characteristics may not be enough. The subject’s height and weight may need to be included and these details may also not be sufficient to identify the records as being from the same subject. Not being able to track a subject throughout a submission could be problematic for safety and efficacy analyses, as well as determination of long-term effects for a drug. It is important for sponsors to employ good data collection practice for previous enrollments as well as establishing a process for assigning the same USUBJID to the same person. Bottom line: If a sponsor has collected the relevant data to track a subject across studies, then an effort should be made to update the USUBJID values to be the same.

WHAT SHOULD SPONSORS DO (THAT MIGHT WORK BETTER)?

In a perfect world, the SDTM standard would have been updated to incorporate the proposed recommendations discussed earlier by allowing multiple records in DM as well as adding SUBJID as a permissible identifier to other subject-level domains. Or it could have been updated using some other proposed strategy that makes sense. But since standards development and regulatory acceptance of the standard evolve at a much slower pace than the perfect world, what can be done in the meantime for regulatory submissions that contain multiple enrollments and still adhering to the standard as it exists today?

SCENARIO #1 (RE-VISITED) – SUBJECTS THAT RE-SCREEN OR RE-ENROLL WITHIN THE SAME STUDY

Going back to the previous example study, XYZ1047, where the sponsor chose to create multiple records in DM, how could this be modeled so that is SDTM-conformant as well as easily loaded into regulatory review tools?

In the study, the subject failed screening twice and then was enrolled on the third attempt, treated, and then completed the study. Since there should be only one record per subject in DM, the instance where the subject was randomized and treated should be the record that is included in this domain.

dm.xpt

STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTC	RFENDTC	...	RFICDTC	SITEID	ARMCD
XYZ1047	DM	XYZ1047-002-0725	002-0725	2017-05-23	2018-10-23		2017-05-14	002	XYZ1047

USUBJID is assigned based on the last SUBJID assigned when the subject was enrolled and RFICDTC is assigned based on the date informed consent was signed at the last screening attempt. Previous SUBJID values assigned when the subject failed screening can be stored in the SUPPDM dataset. Subject characteristics that are present in DM such RACE, ETHNIC, SEX, BRTHDTC that will not change do not need to be duplicated in SUPPDM for each failed screening attempt.

suppdm.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL
XYZ1047	DM	XYZ1047-002-0725			PSUBJID1	Previous Subject Identifier 1	002-0505
XYZ1047	DM	XYZ1047-002-0725			PSUBJID2	Previous Subject Identifier 2	002-0618

Previous informed consent dates as well as the primary informed consent when the subject enrolled can be stored in the DS domain as is done typically in SDTM, i.e., RFICDTC in DM should have a corresponding record in DS. In order to differentiate between each event, the corresponding SUBJID can be stored in SUPPDS.

ds.xpt

STUDYID	DOMAIN	USUBJID	DSSEQ	DSTERM	DSDECOD	DSCAT	DSSTDTC
XYZ1047	DS	XYZ1047-002-0725	1	INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE	2017-04-23
XYZ1047	DS	XYZ1047-002-0725	2	INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE	2017-04-30
XYZ1047	DS	XYZ1047-002-0725	3	INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE	2017-05-14

suppds.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL
XYZ1047	DS	XYZ1047-002-0725	DSSEQ	1	SUBJID1	Subject Identifier 1	002-0505
XYZ1047	DS	XYZ1047-002-0725	DSSEQ	2	SUBJID2	Subject Identifier 2	002-0618
XYZ1047	DS	XYZ1047-002-0725	DSSEQ	3	SUBJID3	Subject Identifier 3	002-0725

For the other subject-level domains that may contain data from multiple screening attempts, the same approach can be taken where the corresponding SUBJID for the record can be stored in SUPPQUAL. The SUPP dataset can then be merged back onto the parent domain during review so that it is clear which subject instance in the trial the data was collected.

Though this approach may not be viewed as the 'ideal' it does provide the following:

- 1) SDTM conformance
- 2) Adherence to the Technical Conformance Guide
- 3) Data that readily loads into regulatory review tools

Some 'not so ideal' points to consider for this approach:

- 1) If there are many previous dates stored in SUPPDM, this data is not machine-readable as a supplemental qualifier, e.g. 'Informed Consent Date 1', 'Informed Consent Date 2', etc.
 - a. Recommend storing these dates elsewhere perhaps in the domain in which they should reside such as DS, SC etc
- 2) Does not work well when a subject is treated and completes the study more than once (that may occur in trials for rare diseases or low subject availability)
 - a. Because only one record is in DM, it forces the sponsor to choose the 'primary' enrollment when each time a subject is treated and passes through the trial is important

SO...WHAT'S NEXT?

As discussed so far, proposals for modeling data for multiple subject enrollments that are not in conformance with the standard pose issues for regulatory review. In an effort to remedy this, an approach has been suggested that is SDTM-conformant that will facilitate review but there are some scenarios where the recommendation is a less than ideal approach. The ideal solution would require updates to the SDTM standards and hopefully will be able to handle many of the situations associated with multiple subject instances.

A FINAL THOUGHT ON THE FDA TECHNICAL CONFORMANCE GUIDE

In the TCG, most recently released in March 2019, there is language included on how to handle multiple enrollments that would still mandate having one record per subject in Demographics (DM):

'For subjects with multiple enrollments within a single study, the primary enrollment should be submitted in DM. Additional enrollments should be included in a custom domain with a similar structure to DM. Clarifying statements in the RG would be helpful.

For subjects with multiple screenings and no subsequent enrollment, include the primary screening in DM with additional screenings in a custom domain with a structure similar to DM.

*For subjects with multiple screenings and subsequent enrollment, include the enrollment in DM with screenings in a custom domain with a structure similar to DM.'*¹

The request is that for multiple enrollment data for subjects that a custom domain for a domain classified as Special Purpose be created. Per the SDTMIG, custom domains can only be created based on one of the three general observation classes: Events, Interventions, or Findings. Currently, following this request would also require modeling data that is not in keeping with the SDTMIG. Also, it is unclear what variables from DM should be included in this custom domain. Until such time as these requirements are provided and the standard is updated, it is recommended to apply the approach that would ensure conformance with the SDTM standard. It is also important to keep in mind that whatever approach is utilized, it is always encouraged to communicate with the review division for the submission.

CONCLUSION

Allowing multiple subject instances either within the same study or across studies has become quite common in clinical trials. Even though neither the SDTM standard nor regulatory agencies provide guidance on how to submit this data currently, there have been a few proposals from the FDA and a PhUSE Working Group. Though these approaches seem straightforward, they may not fully comply with the standard as it exists today. Submitting data that is not in conformance with SDTM may increase review times if the data cannot readily load into regulatory review tools. Until the SDTM standard is updated, it may be helpful for sponsors to adopt an approach that is in keeping with the standard in an effort to shorten review times and to get drugs to patients faster.

REFERENCES

[1] U.S. Food & Drug Administration, *Study Data Technical Conformance Guide* (current version at time of access – v4.3/ March 2019), Accessed March 2019 - <https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources>

[2] Study Data Tabulation Model Implementation Guide: Human Clinical Trials. Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Team. Version 3.2. November 2013

[3] PhUSE, *Ensuring USUBID is Unique for an Individual within a Submission* (current version at time of access – PP30, March 2015), Accessed January 2019 - <https://www.phusewiki.org/docs/CSS2015Presentations/PP30FINAL.pdf>

[4] U.S. Food & Drug Administration, *FDA Validator Rules* (current version at time of access – v1.3/October 2018), Accessed December 2018 - <https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources>

CONTACT INFORMATION

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

Kristin Kelly
Pinnacle 21
kkelly@pinnacle21.com

Mike Hamidi
CDISC
mhamidi@cdisc.org

Any brand and product names are trademarks of their respective companies