PharmaSUG China 2023 - Paper DS-147

Exploration on Demographic as Collected (DC) Domain to Handle Multiple Screenings in SDTM

Xuejie Zhou, Liming Xie, Qiuan Hu and Shanshan Ma, BeiGene, Inc.

ABSTRACT

For multiple screenings within a single study, United States (US) Food and Drug Administration (FDA) position is that multiple screenings and subsequent enrollment should be included in a custom domain with a similar structure to Demographics (DM) domain. The custom domain – Demographic as Collected (DC) domain, has been proposed by Clinical Data Interchange Standards Consortium (CDISC) Submission Data Standards (SDS) Multiple Subject Instances (MSI) team and will be included in SDTMIG v4.0. Though it has not been formally released by CDISC yet, DC domain is adopted as a custom domain in BeiGene. Considering the limited implementation guides on DC domain, we would like to share our experience on the exploration on DC domain in SDTM to dispel the clouds and see the sun.

INTRODUCTION

Multiple screenings are becoming more common in clinical trials to enroll potential subjects, which presents challenges in data collection, tabulation, and analysis. These challenges include determining which case report form (CRF) data needs to be re-entered, ensuring a one-to-one link for the same subject, and presenting the data in SDTM and ADaM.

While the US FDA recommends including the primary screening or enrollment in DM domain and additional screenings in a custom domain with a similar structure. This recommendation has not yet been incorporated into CDISC SDTM implementation guides, making it a challenging process.

Starting in 2022, CDISC SDS MSI Team has proposed the custom DC domain to address this issue and model the FDA's position on multiple screenings and enrollments. DC domain might be included in the release of SDTMIG v4.0.

Despite the lack of clear implementation guides, BeiGene has adopted the DC domain to represent multiple screenings in SDTM and gained valuable experience in the process. We would like to share our insights to shed light on the use of DC domain and remove any confusion surrounding its implementation.

FDA RECOMMENDATION ON MULTIPLE SCREENINGS

Starting from FDA Study Data Technical Conformance Guide v4.2 2018, the US FDA clarifies the recommendation on multiple screenings in Section 4.1.1.3 SDTM Domain Specifications for DM Domain (Demographics) as below:

> DM Domain (Demographics)
>
> In the DM domain, each subject should have only one single record per study.
>
> Screen failures, when provided, should be included as a record in DM with the ARM, ARMCD, ACTARM, and ACTARMCD field left blank. For subjects who are randomized in treatment group but not treated, the planned arm variables (ARM and ARMCD) should be populated, but actual treatment arm variables (ACTARM and ACTARMCD) should be left blank.
>
> 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.

Along with the recommendation for Unique Subject ID (USUBJID) and Subject ID (SUBJID) in Section 4.1.1.2 SDTM General Considerations as below, both sections consist of FDA's current thinking on handling multiple screenings in SDTM.

> Subject Identifier (SUBJID)
>
> The variable SUBJID uniquely identifies each subject that participates in a study. If a single subject is screened and/or enrolled more than once in a study, then the subject's SUBJID should be different for each unique screening or enrollment. For a study with multiple screenings and/or multiple enrollments per subject, SUBJID should be included in other related domains besides DM even though it may cause validation errors. It is recommended to include a table linking each SUBJID for a single subject to that subject's USUBJID with any additional necessary explanation included in the relevant RG.
>
> Unique Subject Identifier (USUBJID)
>
> The variable USUBJID is an identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product. Each individual subject should be assigned a single unique identifier across the entire application. This is in addition to the subject ID (SUBJID) used to identify subjects in each study and its corresponding study report. An individual subject should have the exact same unique identifier across all data sets, including between SDTM and ADaM data sets. Subjects that participate in more than one study should maintain the same USUBJID across all studies. It is important to follow this convention to enable pooling of a single subject's data across studies (e.g., a randomized control trial and an extension study).

The recommendation in short is to:

1) Include a custom domain with a structure like DM domain for additional screenings.
2) Include SUBJID in other related domains besides DM domain even though it causes validation errors.

However, regarding the custom domain implementation, there is no instruction until CDISC SDS MSI Team proposed DC domain in 2022. Additionally, it is not anticipated that the inclusion of custom domains and SUBJID in other domains will align with the CDISC SDTMIG. These deviations from standard practices may result in validation issues that need clarification in the clinical Study Data Reviewers' Guide (cSDRG). Furthermore, the FDA recommendation does not align with the guidance provided by other regional health authorities, such as the Japan Pharmaceuticals and Medical Devices Agency (PMDA) and China National Medical Products Administration (NMPA). This presents additional challenges when preparing for FDA, PMDA, and NMPA submissions concurrently.

CDISC PROPOSED CUSTOM DC DOMAIN

DC domain proposed by CDISC SDS MSI team is a special-purpose domain like DM domain, with a set of essential standard variables that describe each participation of a subject in a clinical trial. This domain is required when representing individual subjects having multiple participations and/or foci of study-specific interests within a trial.

Here, participation in a study is determined by the individual's engagement with protocol. Typically, individuals have a single engagement – either enroll in the study or do not. If they have multiple engagements, each is a distinct participation. There are three general scenarios for multiple participations:

1) Multiple enrollments – the protocol allows subjects to enroll in a study multiple times (multiple screen successes), each enrollment a distinct participation, as the example #1 in Figure 1.
2) Multiple screenings:

a. The protocol allows subjects to engage with a protocol, be identified as is a screen failure, and to subsequently re-engage one or more times with the protocol. The subject consistently remains a screen failure, as the example #2 in Figure 1.

b. The protocol allows subjects to engage with a protocol, be identified as is a screen failure, and to subsequently re-engage one or more times with the protocol. In one of these subsequent engagements, the subject is enrolled in the study (a screen success), as the example #3 in Figure 1.

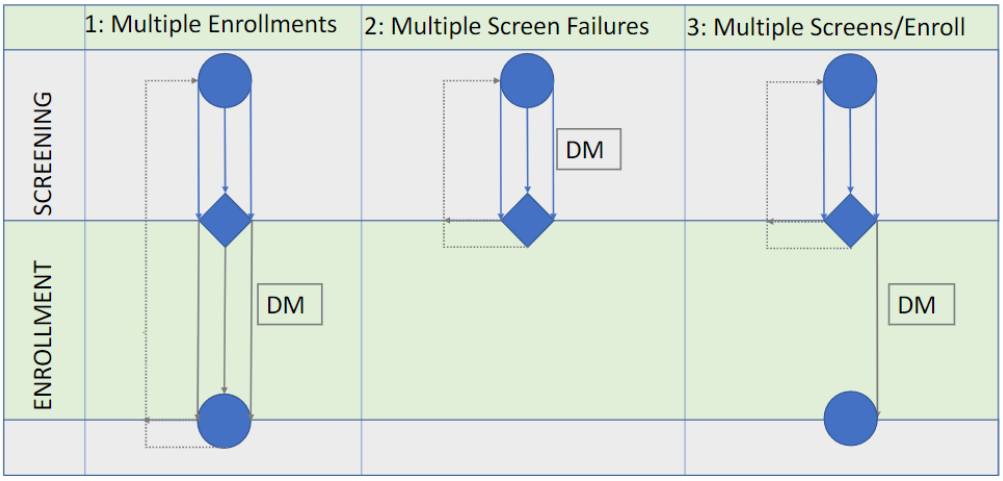


Figure 1. Multiple Subject Enrollments and Screenings

In this paper, we will focus on the scenario of multiple screenings, since it is more common in clinical trials.

To comply with FDA recommendation, the key design of DC domain is as below in Figure 2:

1) USUBJID remains the same for multiple instances of participation.
2) SUBJID represents each participation of the same subject, which means there could be multiple records per subject in DC domain.

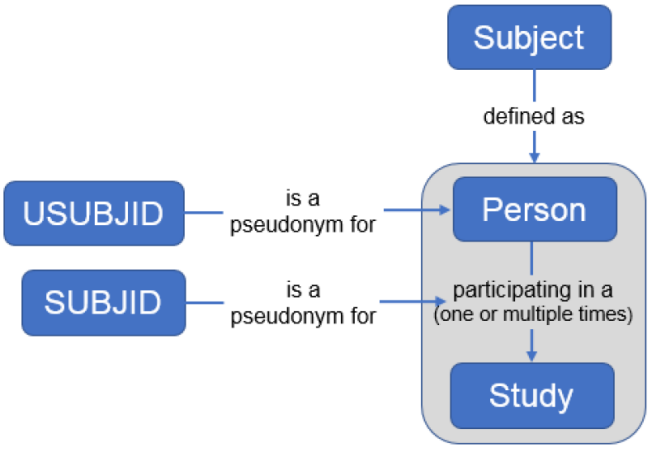


Figure 2. Relationships of USUBJID and SUBJID

DC DOMAIN METADATA

DC Domain data set level metadata as below Table 1:

Table 1. DC Domain Data Set Level Metadata

Domain	Description	Class	Structure	Purpose	Keys
DC	Demographics as Collected	SPECIAL PURPOSE	One record per subject participation	Tabulation	STUDYID, USUBJID, SUBJID

DC Domain variable level metadata is similar with DM domain, with the following new variables added in Table 2:

Table 2. DC Domain New Variables

Variable Name	Variable Label	Type	Controlled Terms or ISO Format	Role	CDISC Notes	Core
FOCID	Focus of Study-Specific Interest	Char		Identifier	Identification of a focus of study-specific interest on or within a subject or specimen as specified in the protocol for which a measurement, test, or examination was performed. The value in this variable should have inherent semantic meaning.	Perm
DCSEQ	Sequence Number	Num		Identifier	Sequence number to ensure uniqueness of subject records within DC. May be any valid number. When possible, it is recommended that the assigned values of DCSEQ reflect the chronological order of the subject participation.	Req
DCDTC	Date/Time of Collection	Char	ISO8601		Date/time of demographic data collection.	Perm
DCDY	Study Day of Collection	Num			Study day of collection measured as integer days.	Perm

DC DOMAIN ASSUMPTIONS

1. DC domain is required when representing data that support one or more individual subjects having multiple participations. However, when there are no subjects with multiple participations, DC domain must not be submitted for regulatory purposes.

2. DC domain is to meet the same assumptions and usage requirements as DM domain except as noted explicitly in the following DC domain assumptions, which means most of the variables would follow the same requirements as DM domain.

3. When used, DC domain represents all distinct participations for all subjects regardless of how many participations the subject has, including the "primary" participation if such a concept is applicable.

1) DC domain must contain all subjects represented in DM domain. This will result in a record in DM domain duplicating a record in DC domain for subjects with one record in DC domain. This is appropriate and supports the understanding that when DC domain is used, DM domain will consist of records derived from DC domain. How multiple participations details are derived or summarized within a single DM record is to be described in the Define-XML and clinical Study Data Reviewer's Guide based on protocol/sponsor convention.

2) Every subject participation in a study must have a distinct subject participation identifier (SUBJID).

IMPLEMENTATION EXAMPLE

Prior to implementing DC domain, we strongly advise gaining an understanding of how subject data, particularly demographics data, will be collected for multiple screenings in the Electronic Data Capture (EDC) system. This approach may differ across various companies.

At BeiGene, subjects will be viewed as unique individuals during each screening and given one unique subject number in the EDC system. All collected CRF forms from previous screening assessments will be re-entered during re-screening assessments, while the previous screening number will be collected to maintain a one-to-one link for the same subject. For example, if a subject failed screening once but was enrolled during a second screening assessment, their previous screening information will be collected as below Table 3:

Table 3. An Example of Subject with Rescreening

Subject Number	Was this subject screened previously?	If Yes, previous Screening Number
SCR-001	No	
SCR-002	Yes	SCR-001

SDTM CREATION ORDER

Typically, the creation of SDTM follows a specific order as below:



When incorporating the DC domain, it is essential to create it after the Trial Design Model but before the DM domain, as shown below:



DC DOMAIN

All information collected on CRFs and previously mapped to the DM domain would now be mapped to the DC domain. For example, mapping race information in a straightforward way:

Race (check all that apply)

SUPPDC.QVAL = AMERICAN INDIAN OR ALASKA NATIVE where QNAM = RACEAIAN

American Indian or Alaska Native — RACE = AMERICAN INDIAN OR ALASKA NATIVE

SUPPDC.QVAL = ASIAN where QNAM = RACEA — RACE = ASIAN

Asian

SUPPDC.QVAL = BLACK OR AFRICAN AMERICAN where QNAM = RACEBAA

Black or African American — RACE = BLACK OR AFRICAN AMERICAN

SUPPDC.QVAL = NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER where QNAM = RACENHPI

Native Hawaiian or Other Pacific Islander — RACE = NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER

SUPPDC.QVAL = WHITE where QNAM = RACEW — RACE = WHITE

White

Not Reported — RACE = NOT REPORTED

In cases where multiple screenings occur, it is common to collect previous screening numbers. This information must be mapped to the DC Supplemental Domain.

Previous screening number	SUPPDC.QVAL where QNAM = DCPRESCN

The following example shows how the previous screening number would be mapped in DC domain:

Table 4. An Example of Mapping Previous Screening Number as SUPPDC.DCPRESCN

STUDYID	DOMAIN	USUBJID	SUBJID	SUPPDC.DCPRESCN
STUDY-A	DC	STUDY-A-086001-001	086001-001	
STUDY-A	DC	STUDY-A-086001-001	086001-002	086001-001

As above example, USUBJID is derived using STUDYID and SUBJID assigned during the first screening assessments, in order to comply with FDA recommendation that each individual subject should be assigned a single unique identifier across the entire application.

The following example shows how the key variables would present in DC domain:

Table 5. An Example of Key Variables in DC Domain

USUBJID	SUBJID	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC	RFICDTC
STUDY-A-086001-001	086001-001					2022-04-20
STUDY-A-086001-001	086001-002	2022-06-16T10:38	2022-09-09	2022-06-16 T10:38	2022-09-09	2022-05-23

DM DOMAIN

With assumption #2 above, DM domain would be the subset of DC domain. In BeiGene, we will keep the last participation record per subject from DC domain, which means that –

1) DM domain is a derived domain directly from DC domain.
2) Except DOMAIN variable reassigned the value, the values of all other variables will be the same as DC domain.

Table 6. An Example of Key Variables in DM Domain

USUBJID	SUBJID	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC	RFICDTC
STUDY-A-086001-001	086001-002	2022-06-16T10:38	2022-09-09	2022-06-16 T10:38	2022-09-09	2022-05-23

OTHER DOMAINS

For all other domains, SUBJID would be added to identify each participation per subject, as below example of Laboratory Findings (LB) domain.

Note that SUBJID would be used to merge SDTM domains, instead of USUBJID.

Table 7. An Example of Subject Records in LB Domain

DOMAIN	USUBJID	SUBJID	LBSEQ	LBTESTCD	LBTEST	VISIT	LBDTC
LB	STUDY-A-086001-001	086001-001	1	HGT	Hemoglobin	SCREENING	2022-04-20
LB	STUDY-A-086001-001	086001-001	2	WBC	Leukocytes	SCREENING	2022-04-20
LB	STUDY-A-086001-001	086001-002	3	HGT	Hemoglobin	SCREENING	2022-05-23
LB	STUDY-A-086001-001	086001-002	4	WBC	Leukocytes	SCREENING	2022-05-23

RELATIONSHIPS SCHEME

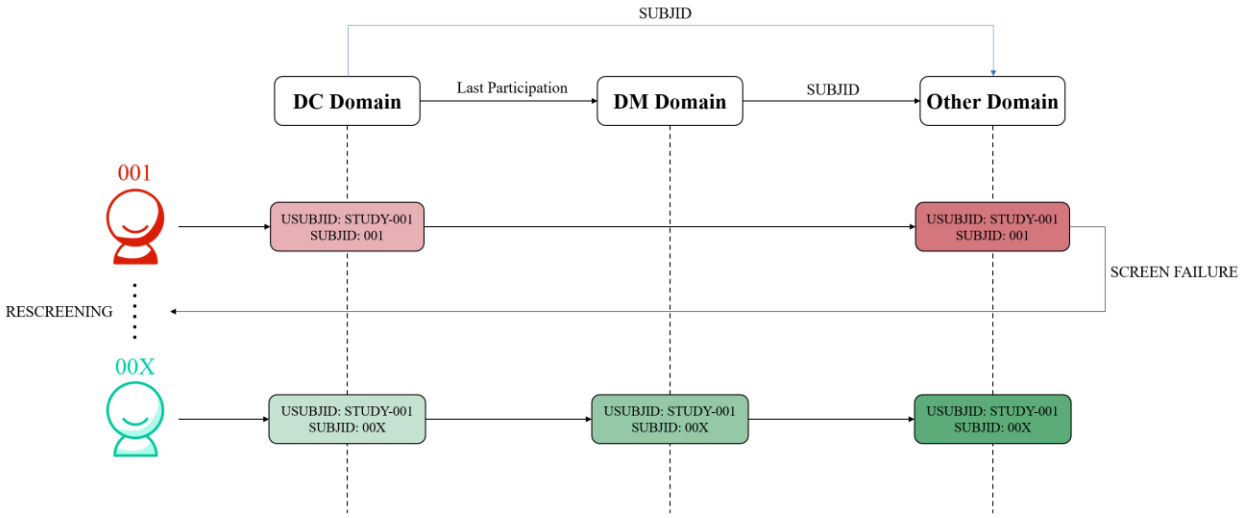


Figure 3. Relationships among DC, DM and Other Domains

IMPACTS ON ADAM AND TFL PROGRAMMING

Please note that when merging SDTM domains in downstream ADaM and TFL programming, it is preferable to use SUBJID instead of USUBJID. However, when study do not have re-screened patients yet, please don't use DC domain in ADaM and TFL programming, since DC domain is a data driven domain and would be removed when we know there are not re-screened patients at the timepoint of Last Patient First Visit.

CONCLUSION

Implementing DC domain to handle multiple enrollments and multiple screenings would be a preferable and more compliant approach according to FDA recommendations. However, currently it poses challenges due to limited implementation guides. In BeiGene, we have explored the implementation of the DC domain and shared our experiences in this paper. While we have identified and addressed some known issues, such as SDTM validation warnings, there may still be unknown issues that need to be discovered and resolved. We anticipate that CDISC will release the DC domain with comprehensive implementation guides in the near future.

REFERENCES

Website Updates on Handling Multiple Enrollments and Screenings Subjects in SDTM by Éanna Kiely. 7th Italian CDISC User Group Network Annual Meeting. https://wiki.cdisc.org/pages/viewpage.action?pageId=136085596&preview=%2F136085596%2F136085612%2FUpdates+on+Handling+Multiple+Enrollments+and+Screenings+Subjects+in+SDTM+-+ClinBuild+-+Eanna+Kiely+-+V1.0.pdf.

Multiple Enrollments / Screenings in SDTM and ADaM by Judith Köppert. https://wiki.cdisc.org/pages/viewpage.action?pageId=178260534&preview=%2F178260534%2F179336012%2F20230314_GCUG_02-03_Multiple+enrollments_JKoeppert.pdf

ACKNOWLEDGMENTS

Thanks to our managers (Leo Li, Cindy Song and Tony Guo) for their great support at the conference and programming colleagues for their help with preparing the paper.

RECOMMENDED READING

- *U.S. Food and Drug Administration's (FDA's) Study Data Technical Conformance Guide (TCG)*
- *CDISC Demographics for Multiple Participations (DC) - SDTM Draft Domains - Wiki.pdf*

CONTACT INFORMATION

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

Xuejie Zhou
BeiGene, Inc.
xuejie.zhou@beigene.com

Liming Xie
BeiGene, Inc.
liming.xie@beigene.com

Qiuan Hu
BeiGene, Inc.
qiuan.hu@beigene.com

Shanshan Ma
BeiGene, Inc.
shanshan.ma@beigene.com