PharmaSUG 2022 – Paper DS-002

Subjects Enrolled in Multiple Cohorts within a Study- Challenge and Ideas

Peng Du and Hongli Wang, Innovent biologics, Inc.

ABSTRACT

Recently, an increasing number of study protocols have allowed the same subject to participate in more than one cohort within a study or multiple studies within a submission. One of the reason could be recruiting subjects especially health subjects and subjects with non-life threaten disease, has gained a lot of challenge during the COVID pandemic season.

At the meantime, the concepts of SUBJID and USUBJID are clearly defined by FDA in the Study Data Technical Conformance Guideline (TCG, MAR 2022). USUBJID should be same for each subject throughout the compound but SUBJID can be different. In this case, collecting and mapping all these demographic information presents a great challenge for our stats programmers and data management.

The CDISC presentation from MSI team has proposed the structure of DC domain at SDTM level. Even so, there are still some remaining questions such as how to handle it at ADaM level and what is the impaction to other SDTM datasets.

In this paper, how to handle the subjects enrolled in multiple cohorts or multiple studies at the programming level will be discussed and examples will be provided to address these questions. In addition, suggestions to handle SDTM datasets such as TS, DM, DC, finding domains and ADaM datasets such as ADSL will be highlighted as well.

INTRODUCTION

In the recent clinical trials, a number of protocols allow one subject enroll more than one cohort/part within the study because subjects have increasing concerns regarding to COVID pandemic. The protocol language could be “Subjects may enroll in more than one cohort, however the last dose given in previous part should be at least XX days before a study drug is administered in a subsequent part in the study.”. This language could be applied in most phase I studies with only health subjects such as Pharmacokinetics and Pharmacodynamics (PK/PD), Drug-Drug Interaction (DDI), Bioavailability (BA) or any studies focus on the rare disease. In another hand, TCG has clearly defined “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” [1]. USUBJID is also defined as “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” [1]. In a simple word, SUBJID should be different for the same subject from different studies or from different cohorts/parts within a study due to the multiple screening/enrollments. Though, the USUBJID should be same for the same subject across the same product or submission package. Per FDA request and CDISC suggestions, the way to handle SDTM and ADaM could be different in different scenarios.

In this paper, two scenarios will be illustrated and suggestions/examples related to each scenario will be given as well.

Scenario 1 – Same Subject enrolled in multiple cohorts/parts within a study.

At SDTM level

A mockup clinical example is shown in Table 1. Firstly, the subject (SUBJID=001) was enrolled in the study 101 cohort A1 with ICF signed on 2020-01-01, passed the A1 screening on 2020-01-15 and finished the cohort A1 last visit on 2020-03-06. About a month later, the same subject was enrolled in cohort B1. A new SUBJID 121 was assigned after the subject passed the B1 screening on 2020-04-15.

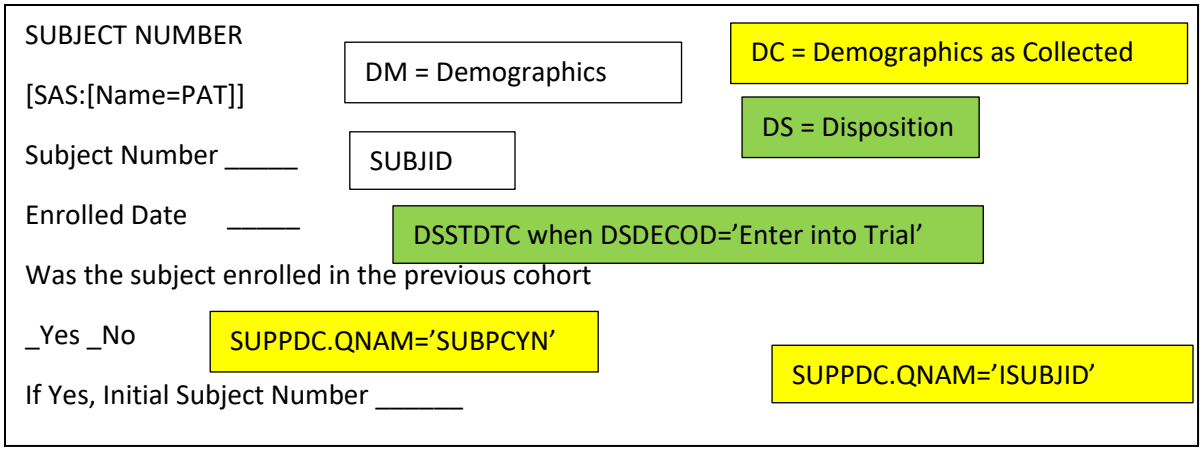
In addition, this patient finished the cohort B1 last visit on 2020-06-06. Lastly this subject was enrolled in cohort C1 with a new SUBJID 217 on 2020-07-01, passed the C1 screening on 2020-07-15 and finished Cohort C1 on 2020-09-06. In this study, the same subject was given with three different SUBJID due to the multiple enrollments. However the USUBJID 01-001 is same from three cohorts

Table 1. Subject Milestone in Study 101

Participation Cohort /SUBJID	Milestone	Date
A1/001	ICF signed/ A1 Screening Pass	2020-01-01/2020-01-15
	First dose	2020-01-20
	Last dose	2020-02-20
	Last safety follow-up	2020-03-06
B1/121	ICF signed/ B1 Screening Pass	2020-04-01/2020-04-15
	First dose	2020-04-20
	Last dose	2020-05-20
	Last safety follow-up	2020-06-06
C1/217	ICF signed/ C1 Screening Pass	2020-07-01/2020-07-15
	First dose	2020-07-20
	Last dose	2020-08-20
	Last safety follow-up	2020-09-06

Firstly, CRF should be used to collect the SUBJID in multiple cohorts. The questions to check if the subjects enrolled in the previous cohorts should be added. If the answer is “Yes”, then the subject ID from previous cohort or the subject ID from the initial cohort should be triggered as the following question (Sponsor can choose which subject ID should be collected). Back to the subject example, the answer to the question “Was the subject enrolled in the previous cohort” should be “No” in cohort A1 and the following question “If Yes, Initial Subject Number” should not be triggered. In cohort B1 and C1, the answer to the question “Was the subject enrolled in the previous cohort” should be “Yes” and the Initial Subject Number should be “001”.

Fig 1: SDTM annotated CRF page collect SUBJID. DC domain is proposed by CDISC MSI team[2].



Secondly, DC (Demographics as Collected) as shown in FIG1 should be used to store the subject level information collected from additional cohorts. This domain was presented by Eanna Kiely from MSI team [2]. DC domain shares the most variables' name and label as DM, such as STUDYID, DOMAIN, SUBJID RFSTDTC, RFENDTC, RFICDTC, BRTHDTC, RACE, ARM, ACTARM etc. However, DC domain also has an additional variable DCSEQ (label: Sequence Number). It is recommended that the value of DCSEQ should reflect the chronological order of the subject participation in the study [2]. Currently, the structure of DC allows multiple records for the same subjects and use DCSEQ to distinguish these records. The example shown in table 2 introduces some basic ideas regarding the structure of DC and DM domain.

Table 2. Example for DM and DC

DM Domain

Row	STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC
1	101	DM	01-001	001	2020-01-20	2020-02-20	2020-01-20	2020-02-20
CONT	RFICDTC	RFPENDTC	DTHDTC	DTHFL	SITEID	AGE	AGEU	SEX
1	2020-01-01	2020-03-06			001	35	YEARS	M
CONT	ARM	ARMCD	ACTARM	ACTARMCD	COUNTRY			
1	Cohort A1	CA1	Cohort A1	CA1	USA			

DC Domain

Row	STUDYID	DOMAIN	USUBJID	SUBJID	DCSEQ	RFSTDTC	RFENDTC	RFXSTDTC
1	001	DC	01-001	121	1	2020-04-20	2020-05-20	2020-04-20
2	001	DC	01-001	217	2	2020-07-20	2020-08-20	2020-07-20
CONT	RFXENDTC	RFICDTC	RFPENDTC	DTHDTC	DTHFL	SITEID	AGE	AGEU
	2020-05-20	2020-04-01	2020-06-06			001	36	YEARS
	2020-08-20	2020-07-01	2020-09-06			001	36	YEARS
CONT	SEX	ARM	ARMCD	ACTARM	ACTARMCD	COUNTRY		
1	M	Part B, Cohort B1	CB1	Part B, Cohort B1	CB1	USA		
2	M	Part C, Cohort C1	CC1	Part C, Cohort C1	CC1	USA		

SUPPDC DOMAIN

STUDYID	RDOMAIN	USUBJID	IDVAL	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
001	DC	01-001	DCSEQ	1	SUBPCYN	Enrolled in the Protocol Previously?	Y	CRF	
001	DC	01-001	DCSEQ	1	ISUBJID	Initial Cohort Subject ID	001	CRF	
001	DC	01-001	DCSEQ	2	SUBPCYN	Enrolled in the Protocol Previously?	Y	CRF	
001	DC	01-001	DCSEQ	2	ISUBJID	Initial Cohort Subject ID	001	CRF	

In Table 2, We suggest all the variables from DM parent domain should be included in DC domain and DC domain should include the extra variable DCSEQ. This approach could potentially help us to easily create the ADSL. The answers to the questions “Was the subject enrolled in previous cohort” and “The initial subject number” should be mapped to the SUPPDC domain.

In this particular complex example, TCG recommended SUBJID should be included in all SDTM domain except the trial domains [1]. Details should be discussed in the cSDRG and the explanation will be added to address the message from P21 report.

In addition, some special variables should be carefully considered in SDTM datasets:

TS: ACTSUB (Actual Number of Subjects) might be different than PLANSUB (Planned Number of Subjects) because one subject is enrolled in multiple cohorts. The actual number of subjects participated in the clinical trial should be less than or equal to the planned number of subjects.

DM: RFXSTDTC/ RFXENDTC are used to present the dosing date and time in the first cohort and RFSTDTC/RFENDTC should be sponsor-defined date time. In most case, RFSTDTC and RFENDTC are equal to RFXSTDTC and RFXENDTC, respectively. RFICDTC/RFPENDTC are used to present the initial Informed Consent form (ICF) obtained date and end of participation date in first cohort. In this example, we choose the map ICF date from the first cohort but not the sponsor considered as “Primary Cohort” into DM domain. Because Pinnacle 21 checks the ICF date in DM domain against to the date in other SDTM domains.

DC: RFSTDTC/RFXSTDTC/RFENDTC/RFXENDTC are used to present the dosing date/time in the rest of following cohorts. RFICDTC/RFPENDTC should be used to present the ICF date and end of participation date in associate cohort respectively.

DC: USUBJID in DC domain must be same as the USUBJID in DM domain for the same subject.

DC: DCSEQ are used to order the records for subjects who are enrolled in more than two cohorts.

DTHDTC DTHFL SITEID COUNTRY SEX RACE and ETHINIC are EXP/REQ in DM domain but perm in DC domain. It is highly recommended to keep these variables in DC in order to easily derive the ADSL. DTHDTC DTHFL SEX RACE and ETHINIC should be subject-level variables. These variables should be not changed due to a different screening process or at a different site. DTHDTC and DTHFL should be populated in all the previous cohorts if the subject passed away in the current cohort.

DC: AGE might be different between DM.AGE and DC.AGE or within DC domain for the same subject, because most AGE in CRF was collected as AGE at the screening in each cohort. It is possible the DC.AGE is greater than the DM.AGE.

Other SDTM domain except the Trials Domain: SUBJID need to be added into the SDTM domain.

Macro updates: some macros are widely used in the SDTM, ADaM and TFLs production. Such as calculating study day or generating EPOCH, --BLFL,--SEQ variables in SDTM. These macros may need to be updated to consider SUBJID as a key variable when merged with DM or SE domains.

At ADaM level:

CDSIC ADaM group is still working on the proposal. In the meanwhile, ADaM structure is not as restricted as SDTM datasets except ADSL. Sponsors could try a few different approaches with ADSL as mentioned below.

Approach 1. Keep a single ADSL, which would result in multiple records per USUBID but different SUBJID. This would need to be discussed in reviewer's guide and the message should be explained in P21 report.

Approach 2. Keep the ADSL records from DM datasets and create another ADSL-like structured ADaM dataset (such as ADPSL) for DC domain. This could be more challenging and time-consuming since to derive the other ADaM datasets, stats programmers have to set ADSL and ADPSL before merging with associated SDTM datasets.

Scenario 2: Same Subject enrolled in multiple studies within one product.

At SDTM level.

In some therapeutic area such as rare disease, it is very common that subjects are allowed to participate in multiple studies within one submission. For example, the subject (SUBJID=001) was firstly enrolled in the study 123 with ICF signed on 2021-01-01, passed the screening on 2021-01-15 (new SUBJID=171) and finished the last visit on 2021-03-06. Furthermore, the subject was enrolled in the study 223 with ICF signed on 2021-08-01, passed the screening on 2021-09-15 and finished the last visit on 2021-12-06. Study 123 could be a proof of concept (POC) study and 223 could be a pivotal study. Both studies are required to be submitted to FDA.

Table 3. One Subject Milestone in Study 101 and Study 223

Study/SUBJID	Milestone	Date
123/001	ICF signed/ Screening Pass	2021-01-01/2021-01-15
	First dose	2021-01-20
	Last dose	2021-02-20
	Last safety follow-up	2021-03-06
223/171	ICF signed/ Screening Pass	2021-08-01/2021-09-15
	First dose	2021-09-20
	Last dose	2021-11-20
	Last safety follow-up	2021-12-06

Firstly, CRF should be updated to add additional two questions as shown in Fig 2. In most studies, USUBJID is derived from study ID, site ID and subject number at the SDTM level. It is impossible to collect USUBJID directly from CRF. But the CRF is able to collect the previous study ID and subject number that the subject was assigned to. Once this information is collected in the raw data, it is much easier for programmers to identify the previous USUBJID.

In addition, DC is not needed in scenario 2 because study 123 and 223 have a split DM domain. USUBJID in 223 should be same the USUBJID from study 123, which is 01-001. Using the same USUBJID could potentially provide a great help on pooled analysis requested by the regulatory agency such as ISS and ISE.

At ADaM level: if study 123 and 223 have a separate CSR report, then no additional process is needed at the ADaM level.

Fig 2: SDTM annotated CRF page to collect the information on subjects previous SUBJID in other studies.

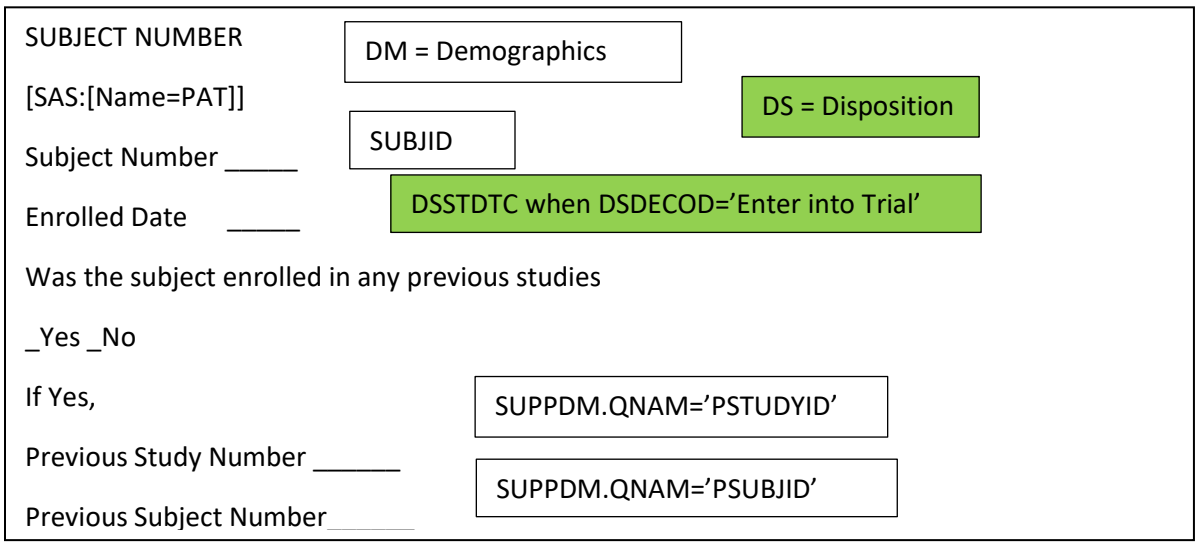


Table 4. Example for DM for study 123 and study 223

DM Domain for study 123

Row	STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC
1	123	DM	01-001	001	2021-01-20	2021-02-20	2021-01-20	2021-02-20
CONT	RFICDTC	RFPENDTC	DTHDTC	DTHFL	SITEID	AGE	AGEU	SEX
1	2021-01-01	2021-03-06			001	35	YEARS	M
CONT	ARM	ARMCD	ACTARM	ACTARMCD	COUNTRY			
1	Drug A	DA	Drug A	DA	USA			

DM Domain for study 223

Row	STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC
1	223	DM	01-001	171	2021-09-20	2021-11-20	2021-09-20	2021-11-20
CONT	RFICDTC	RFPENDTC	DTHDTC	DTHFL	SITEID	AGE	AGEU	SEX
1	2021-08-01	2020-12-06			001	35	YEARS	M
CONT	ARM	ARMCD	ACTARM	ACTARMCD	COUNTRY			
1	Placebo	PBO	Placebo	PBO	USA			

SUPPDM DOMAIN for study 223

STUDYID	RDOMAIN	USUBJID	IDVAL	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
223	DM	001			PSTUDYID	Enrolled in the Protocol Previously?	123	CRF	
223	DM	001			PSUBJID	Initial Cohort Subject ID	001	CRF	

CONCLUSION

Overall, DC (Demographics as Collected) domain should be used if the same subject has multiple enrollments in a study, or if the sponsor chooses to collect multiple screening in a study. This paper is intended to highlight features of DC domains and help the readers to better understand the difference between DM and DC domains. In addition, it provides a couple of examples of what DC domain looks like and when it is needed in the study.

REFERENCES

[1] U.S. Food & Drug Administration*, Study Data Technical Conformance Guide*,Accessed March 2022 - https://www.fda.gov/media/153632/download

[2] Éanna Kiely *Updates on Handling Multiple Enrollments and Screenings Subjects in SDTM,* 7th

Italian CDISC User Group Network Annual Meeting October 2020

CONTACT INFORMATION

Peng (Lucas) Du

pengp.du@innoventbio.com

Hongli Wang

hongli.wang@innoventbio.com