



DS03: A case of mistaken ID

Reimagining rescreenings and re-enrollments

Matt Methereil

I Introduction

- When a subject is able to re-enroll across multiple studies within a single project, it can be difficult to keep track of this one subject while maintaining CDISC compliance and making it clear which observations were recorded pre- and post-rescreening.
- This paper proposes the solution we arrived upon for a three-study project with multiple rescreenings within and between each study.

The Problem

- The SDTM IG v3.3 is clear that for each subject *“only one DM record should be submitted for the subject”*.
- Conversely, the FDA Technical Conformance Guide states that *“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.”*

A Solution

- These positions are incompatible unless we have multiple different values for SUBJID between datasets, so to square this circle we are proud to suggest a new variable: **SUBJIDC**, *Subject Identifier as Collected*.

USUBJID vs SUBJID vs SUBJIDP vs SUBJIDC

1 per project 1 per study 1 per screening

- Our approach begins by standardizing the value of USUBJID as the earliest study- and subject identifiers that a given subject received.
- If they can only pass screening once, the subject ID that successfully screens is kept as SUBJID in DM, and as a supplemental variable in all other datasets for easy reference. Otherwise, the 'final' assigned subject ID is kept.
- All other subject identifiers are stored in SUPPDM as Previous Subject Identifiers (SUBJIDP, SUBJIDP2 etc).
- Then the Subject identifier captured on a given CRF page or eDT is stored as SUBJIDC (Subject Identifier as Captured) in the relevant SUPP domain.

USUBJID vs SUBJID vs SUBJIDP vs SUBJIDC

1 per project 1 per study 1 per screening

Variable	In DM?	In other domains?	Derivation
USUBJID	Yes	Yes	The earliest study and earliest subject identifier
SUBJID	Yes	SUPPxx	The final subject identifier
SUBJIDC	No	SUPPxx	The subject identifier at the time of a given observation
SUBJIDPy	SUPPDM	No	Previous subject identifier. (extra variables can be added as appropriate)

This allows us to quickly identify if a subject has rescreened, and if a given observation is from their current, or a former screening:

SUBJID = SUBJIDC: Observation collected during or after the final screening period.

SUBJID ≠ SUBJIDC: Observation collected during a prior screening.

DM.SUBJIDP populated: Subject rescreened.

DM.SUBJIDP not populated: Subject has not rescreened.

I **STUDYID** vs **STUDYIDP** vs **STUDYIDC**

1 per study 1 per screening

If subjects can enroll for multiple studies, and perhaps some observations need to be copied between these studies, it can be useful to treat the study identifier in a similar way.

- We standardize the value of STUDYID as the identifier of the study we are reporting. Therefore a subject may be associated with multiple STUDYIDs across a project.
- All study identifiers associated with previous subject identifiers are stored in SUPPDM as Previous Study Identifiers (STUDYIDP, STUDIDP2 etc).
- Then the Study identifier captured on a given CRF page or eDT is stored as STUDYIDC (Study Identifier as Captured) in SUPPDM.

I **STUDYID** vs **STUDYIDP** vs **STUDYIDC**

1 per study

1 per screening

Variable	In DM?	In other domains?	Derivation
STUDYID	Yes	Yes	The final study identifier
STUDYIDC	No	SUPPxx	The study identifier for a given observation
STUDYIDP/ STUDIDPy	SUPPDM	No	Previous study identifier. (extra variables can be added as appropriate)

This allows us to quickly identify if a subject has re-enrolled, and if a given observation is from their current, or a former study:

STUDYID = STUDYIDC: Observation collected during the current study.

STUDYID $\neq$ STUDYIDC: Observation collected during a different study.

SUPPDM.STUDYIDP/STUDIDPy = DM.STUDYID: Subject rescreened within current study.

SUPPDM.STUDYIDP/STUDIDPy $\neq$ DM.STUDYID: Subject rescreened from a different study.

Examples

Demographics

STUDYID	DOMAIN	USUBJID	SUBJID		STUDYIDP	SUBJIDP
STUDYA	DM	STUDYA-10001	10001			
STUDYA	DM	STUDYA-10002	10100		STUDYA	10002
STUDYA	DM	STUDYA-10003	10003			

STUDYID	DOMAIN	USUBJID	SUBJID		STUDYIDP	SUBJIDP
STUDYB	DM	STUDYA-10003	20002		STUDYA	10003
STUDYB	DM	STUDYB-20001	20001			
STUDYB	DM	STUDYB-20003	20100		STUDYB	20003

In STUDYA, the subject on row 2 re-entered the study, gaining a new SUBJID.

In STUDYB, the first row is a subject from STUDYA, and row 3 re-entered from within STUDYB.

Example: EG

Sometimes we may use observations captured under previous user or study IDs.

STUDYID	DOMAIN	USUBJID	EGSEQ	EGTESTCD	EGSTRESC	SUBJID	STUDYIDC	SUBJIDC
STUDYB	EG	STUDYA-10003	1	INTP	NORMAL	20002	STUDYA	10003
STUDYB	EG	STUDYA-10003	2	INTP	NORMAL	20002	STUDYB	20002
STUDYB	EG	STUDYA-10003	3	INTP	NORMAL	20002	STUDYB	20002
STUDYB	EG	STUDYA-10003	4	INTP	NORMAL	20002	STUDYB	20002
STUDYB	EG	STUDYB-20001	1	INTP	NORMAL	20001	STUDYB	20001
STUDYB	EG	STUDYB-20001	2	INTP	NORMAL	20001	STUDYB	20001
STUDYB	EG	STUDYB-20001	3	INTP	NORMAL	20001	STUDYB	20001
STUDYB	EG	STUDYB-20003	1	INTP	NORMAL	20100	STUDYB	20003
STUDYB	EG	STUDYB-20003	2	INTP	NORMAL	20100	STUDYB	20100
STUDYB	EG	STUDYB-20003	3	INTP	NORMAL	20100	STUDYB	20100

Alternate solutions

DC Domain: Demographics as Collected

- CDISC working group on Multiple Subject Instances are suggesting a new domain, with a similar structure to DM but with one row per screening/enrollment.
- This would prevent too much being held in SUPPDM. If variables such as age, site, arm etc vary between screenings, this could become a problem.
- DC could co-exist with our suggested SUPPDM variables, as they provide quick orientation of observations, and in simpler cases DC may be unnecessary.

I DC: Demographics as collected

The proposed DC domain allows information collected at each screening to be grouped together on a single line.

In this example RFICDTC is collected at every screening, whereas SEX is only collected at the first, and RFSTDTC is only collected at the last:

STUDYID	DOMAIN	USUBJID	SUBJID	DCSEQ	RFICDTC	RFSTDTC	SEX
STUDYB	DC	STUDYA-10003	10003	1	2020-01-01		M
STUDYB	DC	STUDYA-10003	20002	2	2021-01-01	2021-01-28	
STUDYB	DC	STUDYB-20001	20001	1	2021-02-01	2021-02-28	F
STUDYB	DC	STUDYB-20003	20003	1	2021-03-01		F
STUDYB	DC	STUDYB-20003	20100	2	2021-04-01	2021-04-28	

DC and DM working together

The proposed solution uses an incrementing SUBJID in DC, keeping the final one in DM:

STUDYID	DOMAIN	USUBJID	SUBJID		STUDYIDP	SUBJIDP
STUDYB	DM	STUDYA-10003	20002		STUDYA	10003
STUDYB	DM	STUDYB-20001	20001			
STUDYB	DM	STUDYB-20003	20100		STUDYB	20003

STUDYID	DOMAIN	USUBJID	SUBJID	DCSEQ	RFICDTC	RFSTDTC	SEX
STUDYB	DC	STUDYA-10003	10003	1	2020-01-01		M
STUDYB	DC	STUDYA-10003	20002	2	2021-01-01	2021-01-28	
STUDYB	DC	STUDYB-20001	20001	1	2021-02-01	2021-02-28	F
STUDYB	DC	STUDYB-20003	20003	1	2021-03-01		F
STUDYB	DC	STUDYB-20003	20100	2	2021-04-01	2021-04-28	

I DC and DM working together

My proposal would keep SUBJID constant in DC and increment SUBJIDC

STUDYID	DOMAIN	USUBJID	SUBJID		STUDYIDP	SUBJIDP
STUDYB	DM	STUDYA-10003	20002		STUDYA	10003
STUDYB	DM	STUDYB-20001	20001			
STUDYB	DM	STUDYB-20003	20100		STUDYB	20003

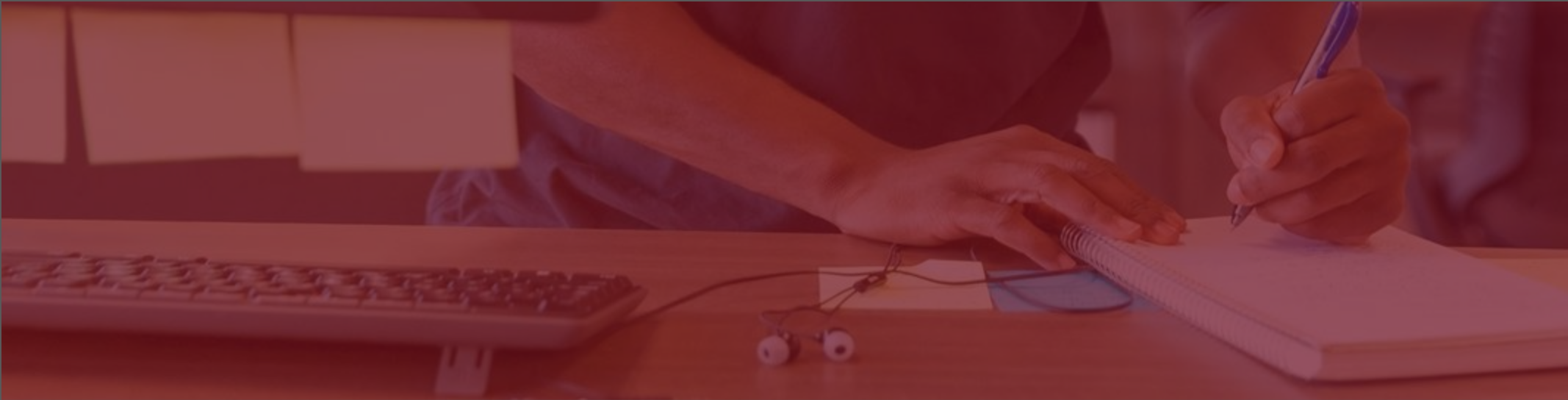
STUDYID	DOMAIN	USUBJID	SUBJID	DCSEQ	RFICDTC	RFSTDTC	SEX	SUBJIDC
STUDYB	DC	STUDYA-10003	20002	1	2020-01-01		M	10003
STUDYB	DC	STUDYA-10003	20002	2	2021-01-01	2021-01-28		20002
STUDYB	DC	STUDYB-20001	20001	1	2021-02-01	2021-02-28	F	20001
STUDYB	DC	STUDYB-20003	20100	1	2021-03-01		F	20003
STUDYB	DC	STUDYB-20003	20100	2	2021-04-01	2021-04-28		20100

ADaM

- All of the previous slides have talked about SDTM datasets, but if SUBJIDPx and STUDYIDP/STUDIDPx were added to the core variables that typically already include STUDYID, USUBJID and SUBJID, then we can perform the same comparisons by keeping STUDYIDC and SUBJIDC in each relevant ADaM dataset.
- We can then easily include this information in listings

An example listing

Subject ID	Treatment	Visit Day/Date	ECG - Interpretation
20001	Age = 30 years, Sex=Female, Weight = 60 kg, BMI = 25.5 kg/m ²		
20001	Treatment A	-28/2021-02-01	Normal
		1/2021-02-28	Normal
		85/2021-05-23	Normal
20002	Age = 40 years, Sex=Male, Weight = 70 kg, BMI = 27.5 kg/m ²		
10003	Treatment A	-394/2020-01-01	Normal
20002	Treatment A	-28/2021-01-01	Normal
		1/2021-01-28	Normal
		85/2021-04-22	Normal
20100	Age = 50 years, Sex=Female, Weight = 80 kg, BMI = 30.0 kg/m ²		
20003	Treatment B	-28/2021-04-01	Normal
20100	Treatment B	1/2021-04-28	Normal
		85/2021-07-21	Normal



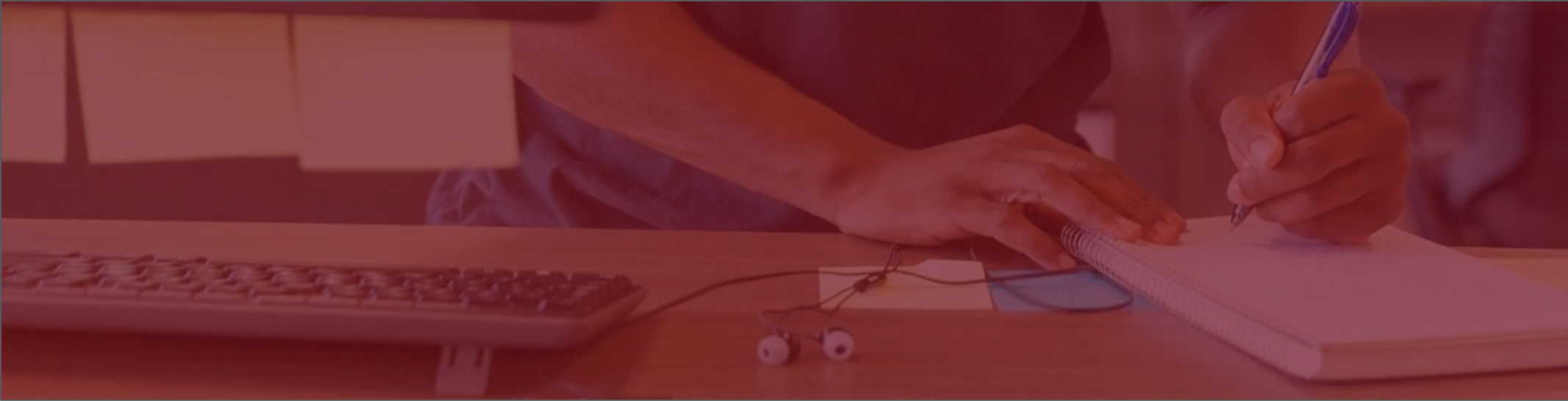
Further reading

[Updates on Handling Multiple Enrollments and Screenings Subjects in SDTM](#)

Italian CDISC User Group Network Annual Meeting

[Considerations When Representing Multiple Subject Enrollments in SDTM](#)

Kristin Kelly, Pinnacle 21 & Mike Hamidi, CDISC



Thank you

Questions?

matthew.metherell@phastar.com