PharmaSUG 2024 - Paper SS- 333

Lead-in and extension trials, how we documented data point traceability

Hanne Ellehøj, H. Lundbeck A/S, Denmark;
Veeresh Namburi, H. Lundbeck A/S, Denmark

ABSTRACT

When generating data packages for extension trials, we have faced several challenges.

In our case, participants had completed a lead-in trial and had the option to continue into an extension trial. There were two independent Electronic Data Capture (EDC) databases built, one for each trial.

When we studied the Statistical Analysis Plan (SAP), we realized that some of the data points collected in the lead-in trial were needed in the reporting of the extension trial. For that we considered our options: we could transfer records from the lead-in trial Study Data Tabulation Model (SDTM) to the extension trial SDTM or from the lead-in trial Analysis Data Model (ADaM) to the extension trial ADaM.

The solution chosen was to take data points from the lead-in trial to the extension trial at the ADaM level. In this paper, we share the reflections we had, the decisions we made, and how we implemented and documented both data points and metadata traceability.

INTRODUCTION

Recently, we have worked on both a lead-in trial and its extension. The two trials were built with independent EDC databases and individual study numbers. As per trial design, the participants in the lead-in trial had the option to enroll into an extension trial on the last visit. The last visit in the lead-in trial would thereby be the same as the first visit in the extension trial. This meant that the assessments taken on this visit belonged to both trials, see Figure 1.

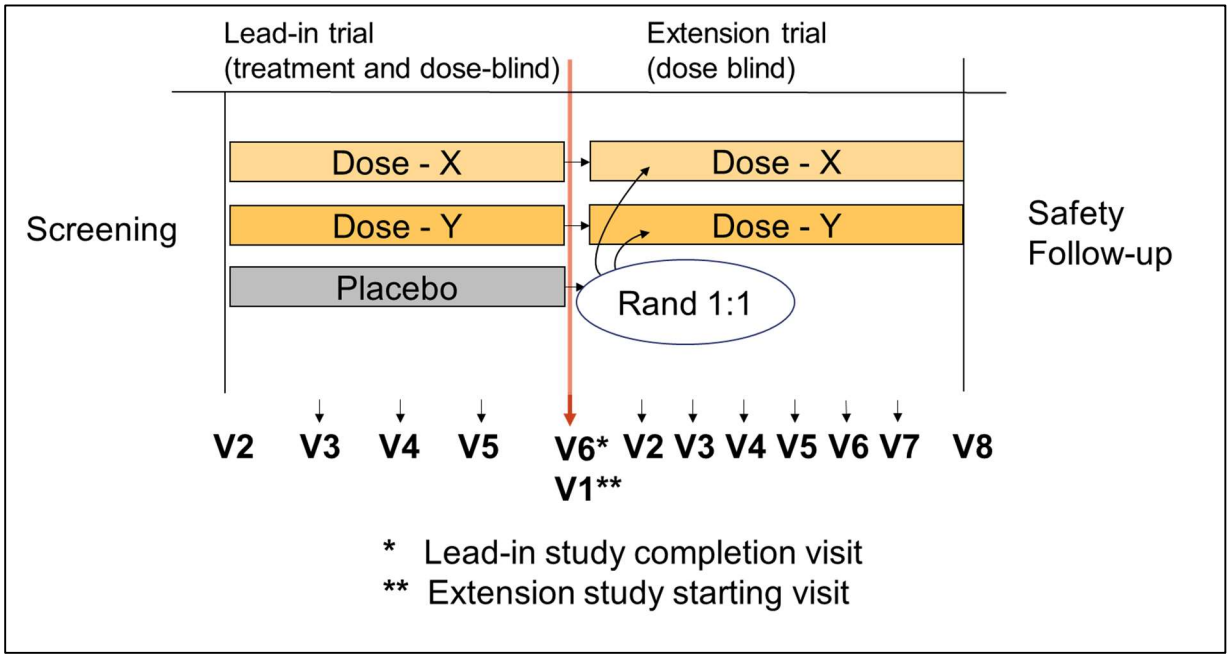


Figure 1: Concept of our lead-in and extension trials; Visit 6 in the lead-in trial took place on the same day as the Visit 1 in extension trial. Assessments on this day belonged to both the lead-in and in the extension trial.

The EDC database for the extension trial was built with a database load feature, taking key data points from the lead-in trial to the extension trial. An example of this was demographic data such as date of birth, sex, ethnicity and race, which were transferred, so the site personnel were not burdened with reentering the data for those who chose to participate in the trial extension. Also, the same USUBJID was used across the trials and was unique per participant, ensuring we could identify subjects across trials. Another example of a load feature was adverse events and concomitant medications, where only ongoing items at the end of the lead-in trial were transferred to the extension trial.

CHALLENGES

We realized that, even though there was a database load-feature, more information from the lead-in trial was needed in the extension trial, according to SAP. This meant that several more data points had to be taken from the lead-in to the extension trial in the analysis datasets. We explored the options we had: Either we had to take records from the lead-in trial SDTM to the extension trial SDTM or take records from the lead-in trial ADaM to the extension trial ADaM. Further, we had to ensure that we were able to document traceability for all data points.

CONSIDERATIONS AND DECISIONS

THE SDTM LEVEL APPROACH

By retrieving records at the SDTM level, this strategy would have the advantage of making SDTM the primary source for future ADaM generation. The challenge with this option would be that SDTM for the extension trial would reflect data from different EDC databases. The source of the SDTM would be EDC, but it would remain unclear from which of the two EDC databases the data points originated. The challenge with this approach lies with the data point traceability.

THE ADAM LEVEL APPROACH – THE APPROACH SELECTED

As revealed earlier, we decided to select the relevant records from the lead-in trial ADaM datasets and adding them as a source to the ADaM datasets of the extension trial, see Figure 2. There are several advantages for taking records on ADaM level.

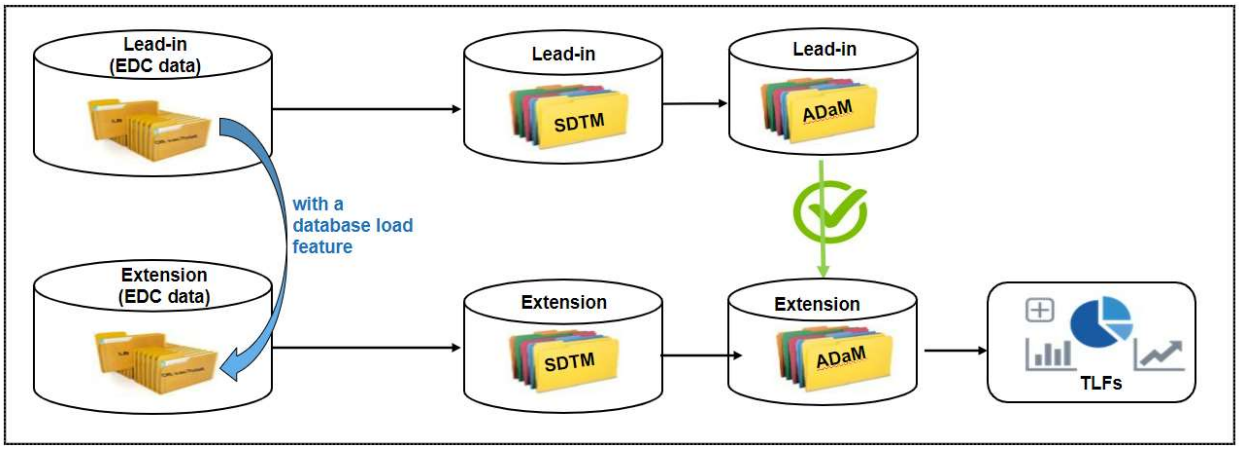


Figure 2: Dataflow for the lead-in and the extension trials. Lead-in trial ADaM datasets, together with SDTM from the extension trial, would serve as a source for ADaM in the extension trial.

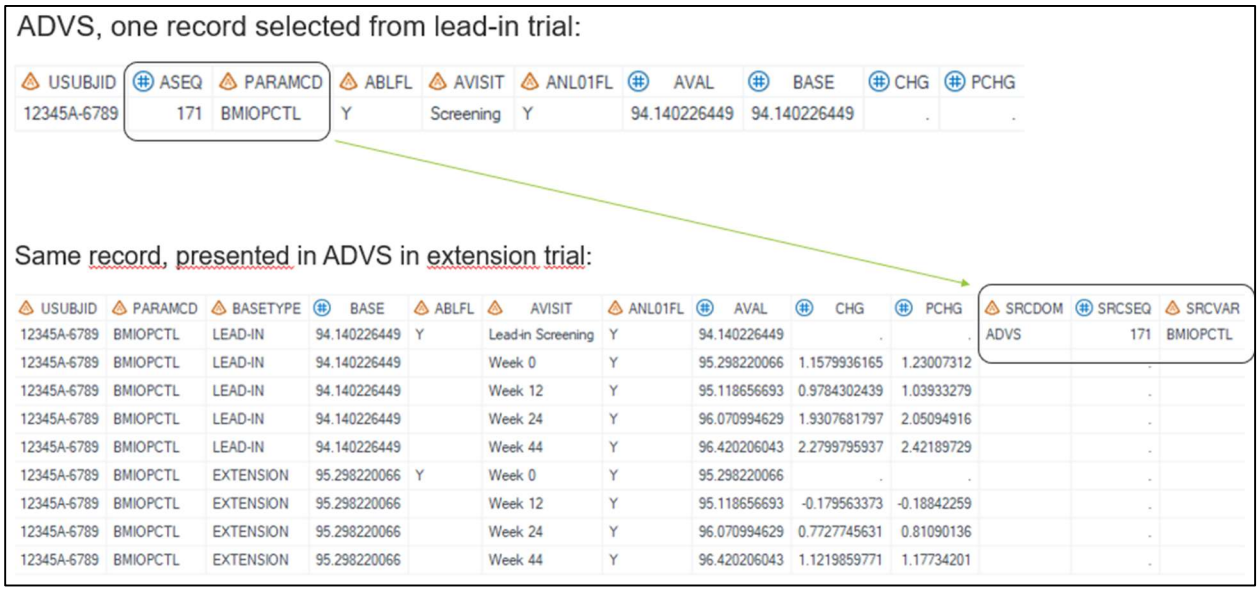
IMPLEMENTATION IN THE BDS DATASET

When it came to implementation for BDS datasets, it was clear that the ADaM level approach had more advantages than first anticipated. In the ADaM for the lead-in trial, both windowing and row-selection were already defined, created, and documented. So, we took advantage of this and used the ADaM variables PARAMCD, AVISIT, and ANL01FL to select which rows to use for the analysis datasets for the extension trial. With these selections, we could ensure that the data points selected were exactly the

same as those used for analysis in the lead-in trial. With this, we avoided re-doing the windowing and analysis flag generation on the data from the lead-in trial. Re-doing work could impose a risk of introducing errors, and the work would again require documentation.

IMPLEMENTATION IN ADVS DATASET

In ADVS for the extension trial, the variables SRCDOM (label=Source Data), SRCVAR (label=Source Variable), and SRCSEQ (label=Source Sequence Number) were used to document data point traceability so that it was clear which rows originated the lead-in trial. The SRCSEQ was assigned from the lead-in ADVS.ASEQ variable, and the SRCVAR was assigned from the lead-in ADVS.PARAMCD variable. With the usage of SRCDOM, SRCVAR, and SRCSEQ, the data point traceability was secured in the ADVS dataset for the extension trial. See Figure 3.



ADVS, one record selected from lead-in trial:

USUBJID	ASEQ	PARAMCD	ABLFL	AVISIT	ANL01FL	AVAL	BASE	CHG	PCHG
12345A-6789	171	BMIOPCTL	Y	Screening	Y	94.140226449	94.140226449	.	.

Same record, presented in ADVS in extension trial:

USUBJID	PARAMCD	BASETYPE	BASE	ABLFL	AVISIT	ANL01FL	AVAL	CHG	PCHG	SRCDOM	SRCSEQ	SRCVAR
12345A-6789	BMIOPCTL	LEAD-IN	94.140226449	Y	Lead-in Screening	Y	94.140226449	.	.	ADVS	171	BMIOPCTL
12345A-6789	BMIOPCTL	LEAD-IN	94.140226449		Week 0	Y	95.298220066	1.1579936165	1.23007312		.	
12345A-6789	BMIOPCTL	LEAD-IN	94.140226449		Week 12	Y	95.118656693	0.9784302439	1.03933279		.	
12345A-6789	BMIOPCTL	LEAD-IN	94.140226449		Week 24	Y	96.070994629	1.9307681797	2.05094916		.	
12345A-6789	BMIOPCTL	LEAD-IN	94.140226449		Week 44	Y	96.420206043	2.2799795937	2.42189729		.	
12345A-6789	BMIOPCTL	EXTENSION	95.298220066	Y	Week 0	Y	95.298220066	.	.		.	
12345A-6789	BMIOPCTL	EXTENSION	95.298220066		Week 12	Y	95.118656693	-0.179563373	-0.18842259		.	
12345A-6789	BMIOPCTL	EXTENSION	95.298220066		Week 24	Y	96.070994629	0.7727745631	0.81090136		.	
12345A-6789	BMIOPCTL	EXTENSION	95.298220066		Week 44	Y	96.420206043	1.1219859771	1.17734201		.	

Figure 3: BDS dataset ADVS, for one participant, one parameter. The variables SRCDOM, SRCSEQ, and SRCVAR are documenting data point traceability. Please notice that only records originating from the lead-in trial have filled SRCDOM, SRCSEQ, and SRCVAR.

DOCUMENTATION IN DEFINE.XML

The metadata presented in define.xml gives the reviewer a clear overview of the source of the three variables SRCDOM, SRCSEQ, and SRCVAR. Figure 4 below show how the three variables were documented in the define.xml. Further, there was a section added in the ADRG describing how data was selected and presented. In this way, it was possible to see where these data points originated – both in define.xml and ADRG.

Variable	Label	Type	Length/Display Format	Source/Derivation/Comment
SRCDOM	Source Data	text	4	Filled only for rows originating from lead-in trial. Set to "ADVS".
SRCVAR	Source Variable	text	8	Filled only for rows originating from lead-in trial. Set SRCVAR=PARAMCD from lead-in trial.
SRCSEQ	Source Sequence Number	integer	3	Filled only for rows originating from lead-in trial. Set SRCSEQ=ASEQ from lead-in trial.

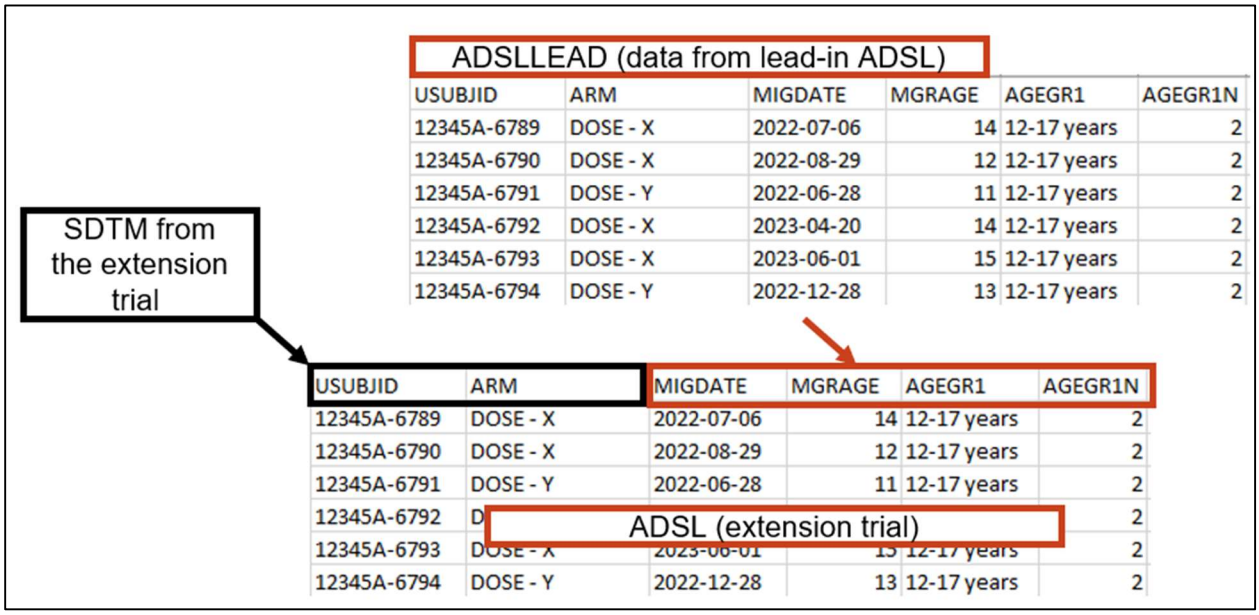
Figure 4: Variable Metadata for the three variables providing data point traceability in ADVS in the extension trial.

IMPLEMENTATION IN THE ADSL DATASET

The next challenge was on the ADSL dataset, which required us to choose an approach once more. The demographics and disease characteristics data points were collected only once per participant, at the start of the lead-in trial. However, as previously stated, the trial database was built with an automatic load-

feature that fetched data points from the lead-in trial to the extension trial. This meant that we had data available in the extension trial database/SDTM. That left us with a choice, as we could derive ADSL from SDTM or take the derived variables from the lead-in ADSL. Again, we looked at how the documentation of data traceability would be simplest, and this was the driver of our decision.

With the one-record-per-subject structure in ADSL, it was not possible to use data point traceability variables (SRCSEQ, SRCDOM, and SRCVAR), as done in BDS structure dataset. But once again, we wanted to take advantage of the already derived variables in the lead-in trial. We therefore decided to create an intermediate "pre-ADSL" dataset, which we named ADSLLEAD (label=Lead-in ADSL dataset). The intermediate "pre-ADSL" dataset and the SDTM from the extension trial, would be the source for the ADSL in the extension trial, see Figure 5. Figure 6 show how metadata traceability was documented in the define.xml for a subset of ADSL variables in the extension trial.



USUBJID	ARM	MIGDATE	MGRAGE	AGEGR1	AGEGR1N
12345A-6789	DOSE - X	2022-07-06	14	12-17 years	2
12345A-6790	DOSE - X	2022-08-29	12	12-17 years	2
12345A-6791	DOSE - Y	2022-06-28	11	12-17 years	2
12345A-6792	DOSE - X	2023-04-20	14	12-17 years	2
12345A-6793	DOSE - X	2023-06-01	15	12-17 years	2
12345A-6794	DOSE - Y	2022-12-28	13	12-17 years	2

Figure 5: Both SDTM from the lead-in trial and the intermediate ADSLLEAD dataset serves as sources for ADSL in the extension trial.

Variable	Label	Type	Length/Display Format	Source/Derivation/Comment	Origin	Core
STUDYID	Study Identifier	text	20	DM.STUDYID	Predecessor	Req
USUBJID	Unique Subject Identifier	text	19	DM.USUBJID	Predecessor	Req
MIGDATE	Date of migraine diagnosis per ICHD-3	text	10	Use ADSLLEAD.MIGDATE	Predecessor	Perm
MGRAGE	Age at migraine diagnosis (year)	integer	8	Use ADSLLEAD.MGRAGE	Predecessor	Perm
AGEGR1	Pooled Age Group 1	text	11	Use ADSLLEAD.AGEGR1	Predecessor	Perm
AGEGR1N	Pooled Age Group 1 (N)	integer	8	Use ADSLLEAD.AGEGR1N	Predecessor	Perm
STRATAR	Strata Used for Randomization	text	3	Select where CM.CMCAT contains "PREVENTIVE HEADACHE/MIGRAINE MED." and if CM.CMPRESP = "Y" and CM.CMOCCUR = "Y" then STRATAR = "Yes". Else STRATAR = "No".	Derived	Perm

Figure 6: Metadata for ADSL for the extension trial. Please notice that three types of sources are presented: predecessor SDTM, predecessor ADSLLEAD, and derived variables.

The approach of using an intermediate dataset for the ADSL generation for the extension resulted in one more ADaM dataset in the submission package, as the ADSLLEAD is intended to be part of the submission package and hence documented in the define.xml.

OCCURRENCE DATASETS AND OTHER DATASETS

As stated in the introduction, the built-in EDC functionality of automatically transferring adverse events and concomitant medications from the lead-in trial to the extension trial made the ADaM dataset

derivation straightforward as all data was already available at the extension trial SDTM level.

CONCLUSION

The examples shown in this paper demonstrate our reflections and decisions when handling data from the lead-in trial in the extension trials. We have provided examples on how we ensured that a reviewer could follow if a data point was copied from the lead-in trial or derived in the extension trial. We did this by implementing traceability variables (SRCSEQ, SRCDOM, and SRCVAR) in BDS structure datasets and by using an intermediate “pre-ADSL” dataset as a source for ADSL in the extension trial. The driver for our decisions was to have specific and simple data point and metadata traceability and to avoid re-work.

ACKNOWLEDGMENTS

We would like to thank our great colleagues across Lundbeck for the good collaborations, their valuable input, and for reviewing this paper.

REFERENCES

CDISC Analysis Data Model (ADaM) Examples of Traceability (Version 1.0 Final)
https://www.cdisc.org/standards/foundational/adam/adam-examples-traceability-v1-0

CONTACT INFORMATION

Hanne Ellehøj and Veeresh Namburi

H. Lundbeck A/S, Copenhagen, Denmark

hanq@lundbeck.com; venm@lundbeck.com