Paper DH06

The Challenges of Using Parent Study Data in a Long-Term Extension Study

Kirsty Wall, GSK, London, UK
Joanne Shoreson, GSK, London, UK

ABSTRACT

Imagine you need to set up the programming to report a Long-Term Extension (LTE) study that was designed to recruit subjects from multiple different parent studies. How would you go about the complicated task of pulling all the parent study data (e.g. labs, AEs, baselines) into the LTE study for reporting?

We will discuss different options along with our chosen solution, considering the assignments of SUBJID and USUBJID in this situation, use of the DQ domain, what variables to include in ADSL to reference the parent studies and how to merge the parent study data, whilst being mindful of potential treatment unblinding issues and the need for CDISC compliance.

INTRODUCTION

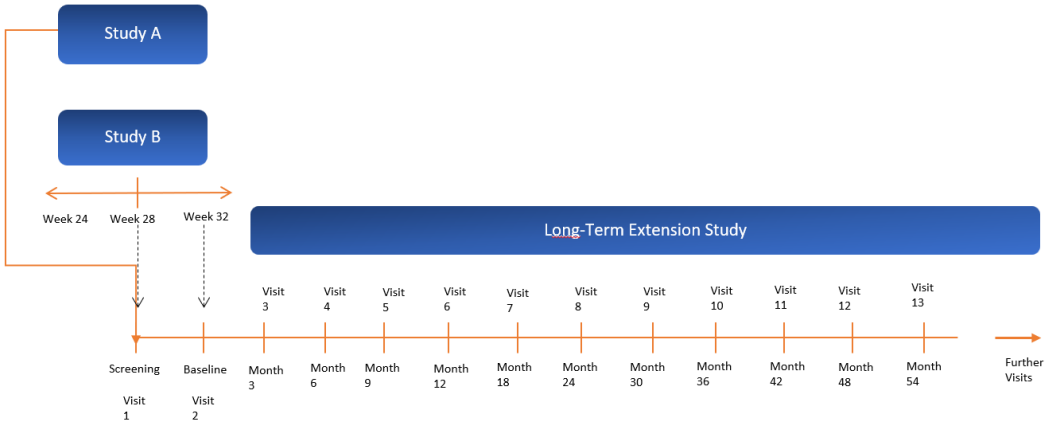
LTE studies play a crucial role in clinical research. Their purpose is to evaluate the long-term safety, efficacy and tolerability of long-term exposure to a new drug. They require a new protocol and hence their own CSR and corresponding Study Data Package, including datasets and data displays. However, they also come with unique challenges when it comes to data collection and merging.

The SDTMIG and FDA Study Data Technical Conformance Guide both state an individual subject should have the same subject identifier across all SDTM and ADaM datasets. It is important to follow this convention so each subject can be uniquely identified across all studies. Given the multiple data sources to be incorporated, merging study data becomes more challenging and requires extensive planning and consideration. Data like Demographics, Medical History etc. along with baseline measurements from the parent studies often need to be incorporated into the extension study. In this paper, we will focus on five study examples and the challenges faced when merging the parent study data.

EXAMPLES

EXAMPLE 1

The first example is of a study design with subjects that completed one of two parent studies and subsequently enrolled into a long-term extension study.



Subjects entered the LTE at the End of Study from Study A and at Week 28 from Study B. The dose level was different across the two parent studies, and three clinical databases were used for data collection. The study team faced challenges when determining baseline values due to numerous clinical databases.

The SDTM.DM domain contained one row per subject in line with CDISC guidelines. DM.SUBJID was assigned as 'MULTIPLE' and mapped to SUBJID1, SUBJID2 in SUPPDM. AE and CE SDTM domains were based only on LTE data. For example, a subject's AE that occurred in the parent study and continued in the LTE study, could potentially be entered into both the parent and LTE study databases. Special attention was required by sites to enter an event into the correct study's database to avoid duplicates.

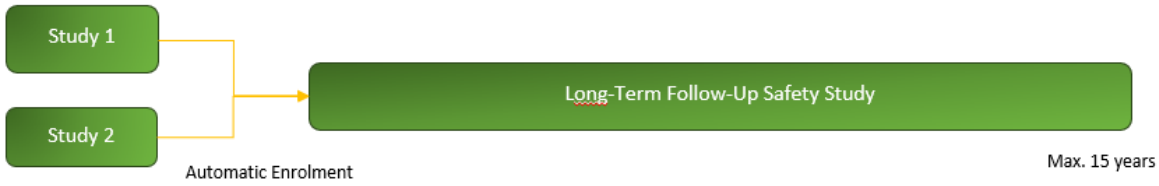
The Week 28 Visit from Study B was mapped to the Screening Visit for the LTE and the Week 32 Visit from Study B was mapped to the baseline visit for LTE. Screening and baseline data was missing in the LTE SDTMs for those subjects who entered the LTE from Study B as this data was not technically collected but still needed to be brought into the LTE for traceability. The SDTM programmer used the same USUBJID as the parent study to facilitate traceability and data integration. The data integration was performed at the ADaM level.

The ADaM Laboratory dataset (ADLB) was created with multiple rows for the same subject with both parent and LTE study data included. The records from Week 28 and Week 32 of the parent Study B were duplicated and the additional records were re-labelled as Screening and Baseline LTE records.

STUDYID	PSTUDYID	USUBJID	PSUBJID	SUBJID	PARCAT1	PARAMCD	AVISIT	AVAL
Study A (111111)		111111.789789		789789	HEMATOLOGY	LB0088	Screening	0.05
Study A (111111)		111111.789789		789789	HEMATOLOGY	LB0088	Baseline	0.06
Study A (111111)		111111.789789		789789	HEMATOLOGY	LB0088	Week 28	0.19
Study A (111111)		111111.789789		789789	HEMATOLOGY	LB0088	Week 32	1.41
Study B (222222)		222222.345345		345345	HEMATOLOGY	LB0088	Screening	0.01
Study B (222222)		222222.345345		345345	HEMATOLOGY	LB0088	Baseline	0.6
Study B (222222)		222222.345345		345345	HEMATOLOGY	LB0088	Week 28	**0.87**
Study B (222222)		222222.345345		345345	HEMATOLOGY	LB0088	Week 32	**1.45**
LTE (333333)	Study A (111111)	111111.789789	111111.789789	678678	HEMATOLOGY	LB0088	Screening	1.13
LTE (333333)	Study A (111111)	111111.789789	111111.789789	678678	HEMATOLOGY	LB0088	Baseline	1.27
LTE (333333)	Study A (111111)	111111.789789	111111.789789	678678	HEMATOLOGY	LB0088	Month 3	1.32
LTE (333333)	Study A (111111)	111111.789789	111111.789789	678678	HEMATOLOGY	LB0088	Month 6	2.54
LTE (333333)	Study B (222222)	222222.345345	222222.345345	456456	HEMATOLOGY	LB0088	Screening	**0.87**
LTE (333333)	Study B (222222)	222222.345345	222222.345345	456456	HEMATOLOGY	LB0088	Baseline	**1.45**
LTE (333333)	Study B (222222)	222222.345345	222222.345345	456456	HEMATOLOGY	LB0088	Month 3	1.98
LTE (333333)	Study B (222222)	222222.345345	222222.345345	456456	HEMATOLOGY	LB0088	Month 6	3.02

EXAMPLE 2

Example 2 is a study design with subjects that completed one of two parent studies and then enrolled into a Long-Term Follow-Up (LTFU) study.



Subjects who completed or withdrew from the parent study were automatically enrolled in the LTFU study for a maximum of 15 years in accordance with FDA and EMA guidance. Subjects kept the same unique Subject ID transitioning from the parent study to the LTFU for traceability. Data was summarized by parent study for the final analyses.

Each subject's demographic data was taken from the data collected in the parent study and all subject data from the completed parent studies was integrated at the SDTM level. The integrated SDTM datasets were used as source data for the ADaM datasets. Data from the Adverse Event (AE) and Medical History (MH) domains was not recollected, these domains were replicated in the LTFU study and only new AE events were reported in the LTFU.

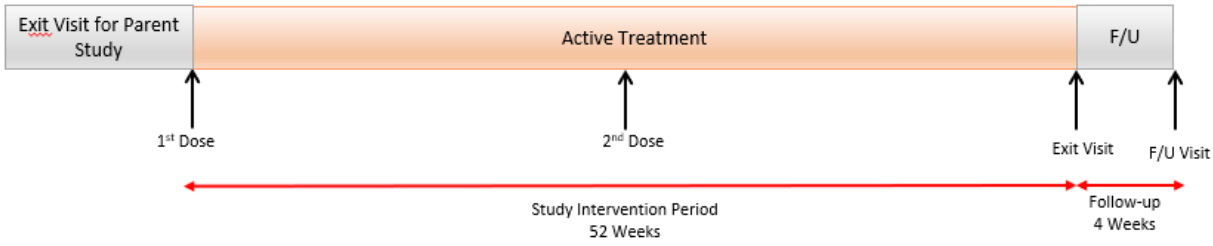
STUDYID	USUBJID	AESEQ	AESTDTC	AEENDTC	AETERM	AEBODSYS
123456	123456.000001	1	15/02/2023	17/02/2023	Headache	Nervous System
123456	123456.000001	2	26/04/2023	28/04/2023	Nausea	Gastrointestinal
456789	123456.000001	3	18/06/2023	25/06/2023	Rash	Skin and Subcutaneous
456789	123456.000001	4	19/09/2023	Ongoing	Diarrhea	Gastrointestinal
456789	123456.000001	5	23/06/2023	Ongoing	Headache	Nervous System

AESEV	AEACN	AEREL	AEOUT	TRTEMFL	PERIOD
Mild	None	Not related	Recovered	Y	Parent
Moderate	Drug withdrawn	Not related	Recovered	Y	Parent
Severe	Dose reduced	Related	Recovered	Y	LTFU
Mild	None	Related	Ongoing	Y	LTFU
Mild	None	Not related	Recovered	Y	LTFU

Once the SDTM datasets were merged consistency checks were performed to ensure there were no discrepancies in the study data variables between the parent and LTFU datasets.

EXAMPLE 3

The following is an example of a 12-month LTE study to evaluate the long-term safety and efficacy of study intervention. Subjects who completed either of the parent studies were given the opportunity to enrol in this open-label long-term extension study regardless of which treatment they received in the previous studies. Subjects received two doses of add-on study intervention over a period of 52 weeks followed by a four-week follow-up period.



For subjects who took active treatment in a parent study, their baseline data was taken from their baseline data in their parent study. For subjects who took placebo in a parent study, their baseline data was taken from their exit visit data in their parent study. Baseline derivations were over-written in the LTE ADaM ADLB dataset and parent study data was not retained in the LTE ADaM ADLB dataset. The study team experienced a challenge maintaining the blind for the parent studies as the LTE was an open-label design. To overcome this issue, an additional unblinded team was set up for unblinded data review as required during the LTE e.g. Protocol Deviations review. USUBJID was carried over from the parent studies into the LTE and data was integrated at the ADaM level. Only new Adverse Events recorded after the start of the LTE treatment period were reported.

EXAMPLE 4

This example is from a project with multiple extension studies, where some subjects participated in a total of five studies. The final study (200010) was a Long-Term Access Programme (LTAP), which was set up to collect additional safety data and to allow subjects with severe disease to have continued access to treatment pre-approval.

Previous Study Participation Prior to LTAP

Route	*Study 1*	*Study 2*	*Study 3*	*Study 4*	*Study 5*	*Most Recent Previous Study*
1	*100001*	*200010*				*100001*
2	*100002*	*200010*				*100002*
3	*100003*	*200011*	*200010*			*200011*
4	*100004*	*200012*	*300013*	*200010*		*300013*
5	*100005*	*200012*	*300013*	*200010*		*300013*
6	*100006*	*200011*	*300014*	*200010*		*300014*
7	*100007*	*200012*	*300013*	*300014*	*200010*	*300014*
8	*100008*	*200012*	*300013*	*300014*	*200010*	*300014*
9	*100009*	*200010*				*100009*

Subjects could enter the LTAP from a previous study via nine different routes, as shown in the table above, with a permitted gap of up to six months. The study consisted of screening and baseline visits, followed by open-label treatment at two dose levels, one for adults and one for paediatrics. Older/heavier children were permitted to switch to a higher dose, creating a third treatment group. Minimal baseline characteristics were captured within this study and the required data was copied from the study that each subject was enrolled in immediately prior to the LTAP (see the 'Most Recent Previous Study' column above). Note that there are six parent (feeder) studies immediately preceding this study. The source data was the LTAP study SDTMs, plus the ADaM datasets from previous studies and also some GSK legacy datasets from older studies which were originally reported pre-CDISC. The total Extent of Exposure was reported across all previous studies where active treatment was received, i.e. not counting time on placebo, so this information was retrieved from each of the corresponding study ADaM (or equivalent) datasets.

In this example, the programmers created a PRVTRIAL dataset which recorded all the previous subject and study information, as this wasn't collected in the CRF. This was then incorporated into the SDTM Subject Characteristics (SC) domain so that all previous study IDs and subject IDs were present for each subject. USUBJID was mapped to the study and subject IDs from the first enrolled clinical trial for each subject.

Extract from SDTM.SC domain:

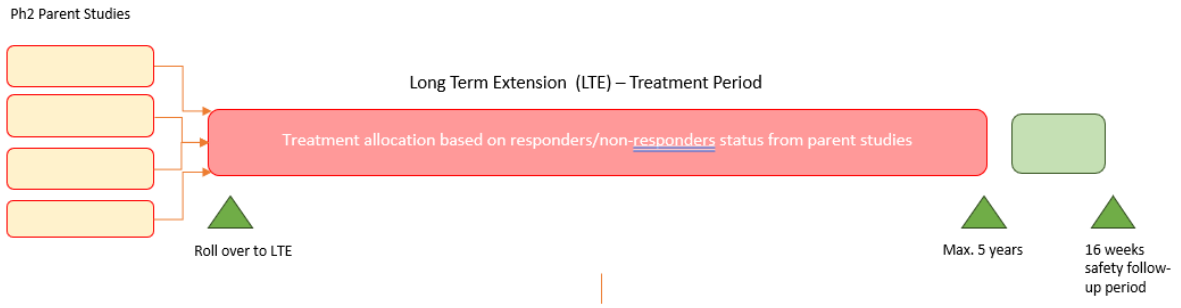
STUDYID	USUBJID	SCSEQ	SCGRPID	SCTESTCD	SCTEST	SCORRES
200010	100001.000001	1	100001.000001	PTYN	Was Subject in Previous Clinical Trial	Y
200010	100001.000001	2	100001.000001	PTSTUDY	Previous Study ID	100001
200010	100001.000001	3	100001.000001	PTSUBJID	Previous Subject ID	000001
200010	100002.000001	1		PTYN	Was Subject in Previous Clinical Trial	Y
200010	100002.000001	2	100002.000001	PTSTUDY	Previous Study ID	100002
200010	100002.000001	3	100002.000001	PTSUBJID	Previous Subject ID	000001
200010	100003.000001	1		PTYN	Was Subject in Previous Clinical Trial	Y
200010	100003.000001	2	100003.000001	PTSTUDY	Previous Study ID	100003
200010	100003.000001	3	100003.000001	PTSUBJID	Previous Subject ID	000001
200010	100003.000001	4	200011.000011	PTSTUDY	Previous Study ID	200011
200010	100003.000001	5	200011.000011	PTSUBJID	Previous Subject ID	000011
200010	100004.000001	1		PTYN	Was Subject in Previous Clinical Trial	Y
200010	100004.000001	2	100004.000001	PTSTUDY	Previous Study ID	100004
200010	100004.000001	3	100004.000001	PTSUBJID	Previous Subject ID	000001
200010	100004.000001	4	200012.000007	PTSTUDY	Previous Study ID	200012
200010	100004.000001	5	200012.000007	PTSUBJID	Previous Subject ID	000007
200010	100004.000001	6	300013.000033	PTSTUDY	Previous Study ID	300013
200010	100004.000001	7	300013.000033	PTSUBJID	Previous Subject ID	000033

Using this approach enables the viewer of the data to see the path a subject took and allows for merging with data from any previous study as required. The manual creation of the previous study information introduced a risk of data errors, which could have been reduced if each of the extension studies had included a CRF page to collect the appropriate previous study information.

EXAMPLE 5

The last example is of a study design with subjects that completed one of four Phase 2 parent studies and enrolled into a Long-Term Extension (LTE) Study. In this example, a subject's treatment allocation in the LTE was based on the subject's responder/non-responder status from the parent studies. Responders stayed on the same treatment as in the parent study (either 1 of 4 active doses or placebo) and non-responders moved to the highest active dose.

Study Design Overview



- Blinding to dose and frequency maintained to avoid unblinding bias, until respective Ph2 parent study was analysed or reported.
- Participants from one or more Ph2 parent studies could roll over to LTE.

A Previous Clinical Trial Participation form was incorporated into the eCRF for the LTE study to collect a subject's Subject ID and Site ID from the parent study. This data was populated in the Subject Characteristics (SDTM.SC) dataset.

Example eCRF Page SC Domain

`SC (Subject Characteristics)`

Previous Trial Involvement (PRVTRIAL)

Previous Trial Involvement Status	`SCCAT=PREVIOUS TRIAL INVOLVEMENT`
Participated in a Previous Clinical Trial/Study for the Same Study Intervention Given in this Trial/Study	Yes No `SCORRES when SCTESTCD=PTYN`

Previous Trial Involvement Details - 1	
Previous Study Identifier/Protocol Number/Identifier	`SCORRES when SCTESTCD=PTSTUDY`
Previous Center Number	`SCORRES when SCTESTCD=PTCENTER`
Previous Participant Identifier	`SCORRES when SCTESTCD=PTSUBJID`

The SC dataset was used to assign the Unique Subject ID (USUBJID) across all datasets. The SDTM programmer created the SC domain for the LTE study (123456) with four records per subject from the four different parent studies:

Extract from SDTM.SC domain:

STUDYID	USUBJID	SCSEQ	SUBJID	SCTESTCD	SCTEST	SCORRES	SCDTC
123456	111222.000001	1	000002	PTYN	Was Subject in Previous Clinical Trial	Y	2020-03-20
123456	111222.000001	2	000002	PTSTUDY	Previous Study ID	111222	2020-03-20
123456	111222.000001	3	000002	PTSUBJID	Previous Subject ID	000001	2020-03-20
123456	111222.000001	4	000002	PTCENTER	Previous Center Number	xxxx	2020-03-20
123456	333444.000001	1	000003	PTYN	Was Subject in Previous Clinical Trial	Y	2022-03-24
123456	333444.000001	2	000003	PTSTUDY	Previous Study ID	333444	2022-03-24
123456	333444.000001	3	000003	PTSUBJID	Previous Subject ID	000001	2022-03-24
123456	333444.000001	4	000003	PTCENTER	Previous Center Number	xxxx	2022-03-24
123456	555666.000001	1	000004	PTYN	Was Subject in Previous Clinical Trial	Y	2023-04-26
123456	555666.000001	2	000004	PTSTUDY	Previous Study ID	555666	2023-04-26
123456	555666.000001	3	000004	PTSUBJID	Previous Subject ID	000001	2023-04-26
123456	555666.000001	4	000004	PTCENTER	Previous Center Number	xxxx	2023-04-26
123456	777888.000001	1	000005	PTYN	Was Subject in Previous Clinical Trial	Y	2021-06-18
123456	777888.000001	2	000005	PTSTUDY	Previous Study ID	777888	2021-06-18
123456	777888.000001	3	000005	PTSUBJID	Previous Subject ID	000001	2021-06-18
123456	777888.000001	4	000005	PTCENTER	Previous Center Number	xxxx	2021-06-18

The SDTM programmer created a Demographic Qualifiers (DQ) SDTM dataset for the LTE study that contained all the subjects from the parent study and the LTE. There were two records for each subject, SUBJID contained the subject number from the parent study on one record and subject number from the LTE on the other record. The same USUBJID was used across all SDTM datasets in the LTE. The new Subject ID given in the LTE study was only populated in SUBJID not USUBJID to ensure subjects only had one unique subject identifier across all domains as per CDISC and FDA guidelines.

SDTM.DQ domain:

LTE STUDYID	USUBJID	SUBJID
123456	111222.000001	000001
123456	111222.000001	000002
123456	333444.000001	000001
123456	333444.000001	000003
123456	555666.000001	000001
123456	555666.000001	000004
123456	777888.000001	000001
123456	777888.000001	000005

A corresponding ADaM dataset ADDQ was created with multiple rows for the same subject which facilitated merging data from any study. This way integration of data was performed at the ADaM level. Variables in the Subject Level dataset (ADSL) displayed relevant information from the parent study e.g. previous subject and study ID so the required 'one record per subject' format was maintained.

ADSL

LTE STUDYID	LTE USUBJID	LTE SUBJID	PARENT STUDYID	PARENT SUBJID
123456	111222.000001	000002	111222	000001
123456	333444.000001	000003	333444	000001
123456	555666.000001	000004	555666	000001
123456	777888.000001	000005	777888	000001

Following this process, incorporation of lab data, for example, from the parent study into the LTE was performed at the ADaM level. ADaM datasets created in the LTE study included both parent and LTE study data in one dataset by merging the source datasets by USUBJID and other relevant variables to create e.g. the ADaM ADLB dataset. In this example, all derivations were traceable back to parent study SDTM datasets.

BLINDING CONCERNS

Blinding in LTE studies is crucial for maintaining the integrity of the study and ensuring unbiased results. Here are some examples of key concerns related to blinding in LTE studies:

ETHICAL CONSIDERATIONS

If a subject was randomized to the control arm of a parent study and had a positive outcome, it may pose a risk to re-randomize in the LTE if the subject then received the active arm treatment. Similarly, if a subject was randomized to an active treatment arm of a parent study with poor results, it could be considered unethical to enrol in the LTE study.

LONG-TERM CHALLENGES

As LTE studies often continue for many years, changes are likely as new information becomes available. Any change in the study design should be carefully monitored to avoid compromising the blind.

DATA INTERPRETATION AND MANAGEMENT

Unblinding can lead to biased data interpretation and reporting. One solution is to have a separate team review the unblinded data to preserve the integrity of the results.

MAINTENANCE OF BLINDING

Once the results of the parent studies are known, this could compromise the blind of the subjects in a LTE, e.g. in Example 5 where the treatment assignment for the LTE is based on that of the parent study. Here, a responder in the parent study remains on the same active dose and a non-responder switches to the highest active dose. There also may be lab tests where once the result is known it becomes obvious which treatment arm the subject is on, either in the LTE and/or the parent study.

CONCLUSION

The purpose of this paper was to provide example studies and discussion of points to consider when creating SDTM and ADaM datasets for LTE studies. Ensuring that data collection methods, variables and formats are consistent at the study set-up stage where feasible between the parent and LTE studies will lead to easier merging of data and improved dataset quality at the time of submission. We also recommend inclusion of a 'Previous Trial Involvement' CRF Page in an Extension Study, to capture the required information on prior Study IDs and Subject IDs.

Although we have shown that the design of LTEs is very varied and there are multiple ways to handle the data, our preferred option is Example 5. Our reasons are because no data is over-written in the Long-Term Extension study datasets, ensuring traceability and easier programming. The LTE study datasets exist as a standalone package, and it is clear within these datasets on the path each subject took prior to entering the extension study.

Whilst there are multiple methods that could be used, it is important when merging study data to consider early planning, standardising data collection and addressing any issues quickly. Documenting the process of how the data has been merged and any limitations of the process should be included to ensure transparency and maintain traceability between SDTM and ADaM. By using careful upfront planning, it is possible to effectively merge data from Long-Term Extension studies with parent study data ensuring a comprehensive and robust analysis.

CONTACT INFORMATION

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

Author Name: Kirsty Wall
Company: GSK
Address: 79 New Oxford Street, London, WC1A 1DG
Email: Kirsty.j.wall@gsk.com
Website: www.gsk.com