SESUG Paper 166-2018

Tips on Developing SDTM Datasets for Complex Long-Term Safety Studies

Yunzhi Ling and Helen Wang, Sanofi, Bridgewater, NJ

ABSTRACT

There is an increase tendency in clinical research practice to combine multiple studies' extension phases into one long-term safety (LTS) study, as the cost of conducting one LTS study is much lower than maintaining individual LTS study for each core study separately, from clinical operation point of view. However, such combination significantly increases the complexity of LTS study's Clinical Study Report (CSR) programming handling, especially in Study Data Tabulation Model (SDTM) datasets development. This paper illustrates the key areas that need to pay special attentions, in SDTM development comparing with that of a typical study, in order to effectively facilitate LTS study's Analysis Data Model (ADaM) datasets creation and potential datasets pooling for integrated analyses.

INTRODUCTION

In late-stage pre-market and post-approval clinical investigations, long-term safety study is conducted to evaluate a drug's safety profile after subjects completed core phase study. In traditional clinical research practice, each core phase study has its own LTS study, with core and extension study operational data collected in either one database or two separate databases. Because a LTS study usually runs in years, with multiple interim analyses conducted at project milestones, the cost of keeping multiple LTS studies for one compound can be very high. When multiple core studies' extension phases are combined into one LTS study, clinical operational cost is usually much lower as there is only one study to be conducted and one operational database to be setup and maintained.

However, this combination increases the complexity of LTS study SDTM datasets creation to facilitate LTS study CSR reporting as LTS study data are coming from various sources:

- LTS study operational database.
- Individual core studies' SDTM and/or ADaM datasets to select certain data that only collected in core studies, e.g. demographics and baseline characteristics, baseline records and core studies' End of Treatment (EOT)/End of Study (EOS) records of finding domains, because:
    - Core studies' baseline values of endpoints are generally defined as the corresponding baseline values for LTS study.
    - Core studies' EOT/EOS data, the last available measurements before subjects exposed to LTS study treatment(s), are considered as LTS study's own baseline (or week 0 records).
    - Core studies' ADaM datasets are sometimes needed when EOT/EOS records that are only identified and defined in ADaM datasets of core studies (e.g. EOT/EOS records selected for core studies' CSR analysis when multiple records existed in the EOT/EOS visit window).

Following the CDISC SDTMIG and FDA eCTD Guidance, LTS study SDTM datasets being the source datasets submitted to FDA, should contain all data for LTS study CSR reporting, to show data traceability and completeness. Given multiple data sources to be incorporated, LTS study SDTM datasets setup becomes most challenge and requires extensive planning and considerations. After LTS study SDTM datasets are properly generated, ADaM datasets creation using SDTM datasets as the sole sources are relative straightforward and not very different from a typical study's ADaM datasets creation from SDTM datasets.

Based on two recent projects' LST study SDTM datasets development and FDA eCTD submission experiences, with one compound has been approved, this paper summarizes the special considerations in LTS study SDTM datasets development and related eCTD components preparation, compared with a typical study's SDTM datasets creation.

GENERAL PLANNING AT SET-UP STAGE

The following is an example of study design of a LTS study. Patients who completed core study A, B, C, or D are eligible to roll over into the LTS study. LTS study started after study A was completed. So patients who rolled over from study A have a separate 1-3 weeks screening period. Patients who rolled over from study B, C, or D enrolled into LTS study on the same day as the EOT visit of the core Study.

Study A: Patients who completed study A (complete EOS visit) are eligible to rollover into LTS study. There is a separate screening period for those patients who rolled over into LTS study.

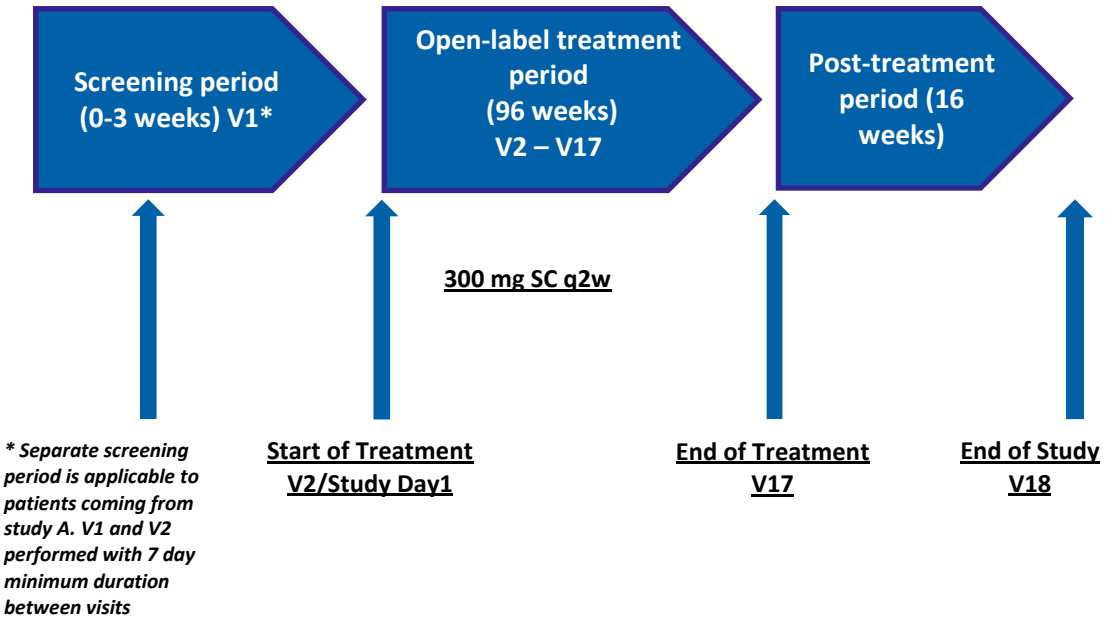

Figure 1: Study Flow Chat for Patients Who Completed Study A and Rolled Over Into LTS Study

Study B, C, and D: Patients who completed planned treatment period of the study (complete EOT visit) are eligible to rollover into LTS study. There is no separate screening period for those patients who are eligible to rollover into LTS study.

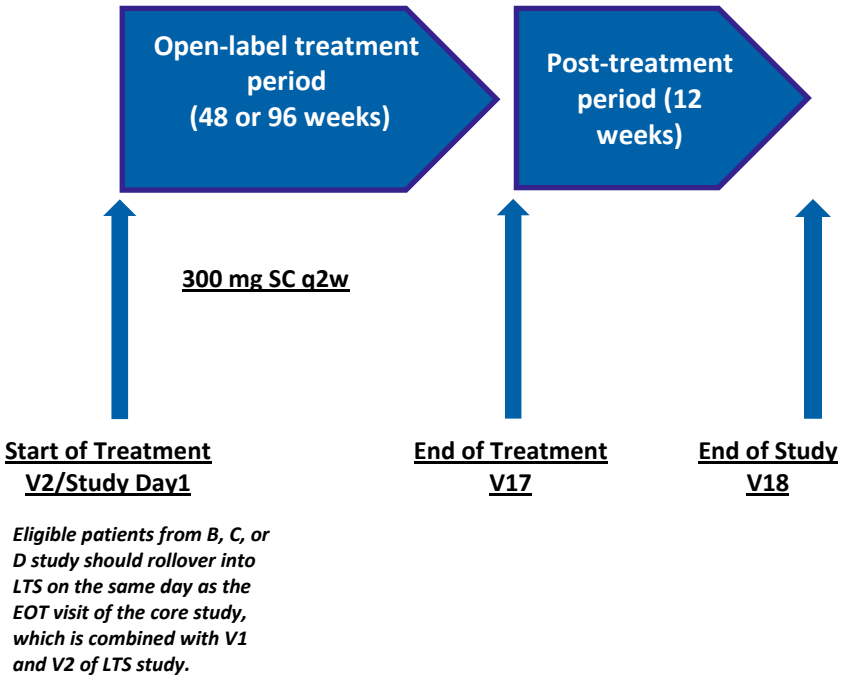

Figure 2: Study Flow Chat for Patients Who Completed Study B, C, or D and Rolled Over Into LTS Study

According to CDISC principles, SDTM datasets submitted to the FDA should contain all data collected during the course of a study. ADaM datasets created afterwards have all derivations traceable back to SDTM datasets. As the only source data submitted to FDA, LTS study SDTM datasets need to be self-sufficient to support CSR analyses or data exploration by FDA reviewers. Thus, in addition to all the considerations in a typical single study's SDTM datasets creation, LTS study SDTM datasets not only include all collected data from LTS study operational database, but also need to:

- Include certain data that only collected in core studies, e.g. demographic, baseline characteristics, baseline disease characteristics, medical or surgical history, prior medications, etc.
- Use the same USUBJID as core phase study to facilitate subject level data traceability and dataset integration.
- Obtain baseline and EOT/EOS records from core studies for finding domains. As shown in the study design flow charts above, the EOT visit of some of the core studies (such as B, C, and D study) is on the same day as the combined visit 1 and 2 of LTS study prior to expose to LTS study drug. So the measurements collected at EOT visit of core studies are often considered as the week 0 or LTS study own baseline records. For core studies (such as study A) that completed prior to the start of LTS study, the EOS visit is on the same day as the combined visit 1 and 2 of LTS study prior to expose to LTS study drug, so the measurements collected at EOS visit of core studies are often considered as the week 0 or LTS study own baseline records.
- Add Visit and USUBJID combination for the records obtained from core study to SV domain to avoid Pinnacle 21 violations.

- Remind data management group to enter data into the right study database to avoid duplicates at beginning, such as the same Adverse Events (AEs) could be entered into both core and LTS study database.

SDTM DOMAIN LEVEL CONSIDERATIONS

1. DM DOMAIN

- LTS study's partial or complete set of demographic and subject characteristics data are obtained from core studies, depending on study CRF design.
- Use the same USUBJID as core phase study to enable subject level data traceability and dataset integration.
  It's common to have subjects switch sites during LTS study due to original site closure or changes, USUBJID changes when a subject switches site as USUBJID is usually formed by concatenating STUDYID, COUNTYID, SITEID, and SUBJID.
  USUBJID and SUBJID are defined in DM of LTS study. USUBJID is the core study's USUBJID and SUBJID is the latest SUBJID in the LTS study.
  For other domains, the USUBJID was obtained by merging with DM using SUBJID as SUBJID was available in LTS study operation database.
  For patients who switched sites, their USUBJID was inconsistent with SUBJID. This was explained in SDRG and a list of patients with site switched was listed in SDRG as well.
- All patients who signed informed consent form of LTS study were included in DM of LTS study. As the study was still on-going at interim lock, patients whose eligibility and treatment allocation were not yet available as of the data cutoff date were assigned as screen failure (ARMCD=SCRNFAIL). This was also explained in SDRG.

2. TV AND SV DOMAINS

Per SDTMIG, the TV domain describes the planned visits in a trial. Therefore unscheduled visits should not be included in this domain. For scheduled visits, the VISITNUM and VISIT values must match those found in all other domains.
SV domain should contain all visits defined in all other domains (both scheduled and unscheduled visits). For scheduled visits, the VISITNUM and VISIT values must match those found in the Trial Visits (TV) domain.

To be compliant with SDTMIG and avoid Pinnacle 21 violations,

- Add USUBJID and VISIT combination (both scheduled and unscheduled visits) of the baseline records obtained from core studies' SDTM datasets into LTS study SV domain;
- Add scheduled VISIT of the baseline records obtained from core studies' SDTM datasets into LTS study TV domain.
- Same thing needs to be done on TV and SV domains for the EOT/EOS records obtained from core studies' SDTM datasets as LTS study own baseline (or week 0) records.

The following example illustrates the VISIT 1 and Visit 2 records obtained from core studies and added into TV domain of LTS study as the baseline records were derived from VISIT1 and VISIT 2 records of core studies.

STUDYID	DOMAIN	VISITNUM	VISIT	VISITDY	TVSTRL
STUDY01	TV	-2	Core Study Visit 1	-365	Depends on core studies
STUDY01	TV	-1	Core Study Visit 2	-365	Depends on core studies

STUDYID	DOMAIN	VISITNUM	VISIT	VISITDY	TVSTRL
STUDY01	TV	1	Visit 1	-7	Day -7 to D-1
STUDY01	TV	2	Visit 2	1	On the same day as the start of the treatment epoch
STUDY01	TV	2.5	Visit 1/2	1	On the same day as the start of the treatment epoch

3. FINDINGS DOMAINS SUCH AS: LB, VS, EG, IS, ETC.

Data for finding domains are usually collected by visit. Based on SAP of LTS study, the baseline of core studies are usually defined as the baseline of LTS study. So the baseline records of core studies' need to be obtained from core studies and added into the corresponding domain of LTS study. Depend on SAP, the core studies' EOT/EOS records may also need to be obtained and added into LTS study SDTM datasets as LTS own baseline (or week 0 records).

- Obtain core studies' baseline records and add them into the corresponding domains of LTS study if core studies' baseline is also defined as the baseline of LTS study.
- Select core studies' EOT/EOS records if considered as LTS study own baseline (or week 0 records) from core studies' SDTM datasets and add them into LTS SDTM datasets. If there are multiple records for the week 0 analysis visit window (as some records were collected in LTS study own operation database), the records selected for LST CSR analysis are flagged as ANLxxFL='Y'. These records flagged for LTS CSR analysis might not be the same records obtained from core studies' SDTM datasets and used for core studies' CRS analysis.
- In the case that SAP requires that the records flagged for LTS CSR analysis for week 0 are the same as those used for core studies' CSR analysis for EOT/EOS analysis visit window, it is recommended to obtain EOT/EOS records from core studies' ADaM datasets to ensure the consistency with core studies' CSR analysis.
- Some derived variables on the records obtained from core studies' SDTM datasets such as --SEQ, EPOCH, --DY, ELEMENT, ETCD, need to be re-derived using LTS study reference date, e.g. RFSTDTC/RFENDTC.
- Add flag (--CORSTD=Y) to indicate records obtained from core studies not only to show traceability but also to ease ISS/ISE datasets integration when pool core studies and LTS study together. The records obtained from core studies' are flagged with --CORSTD='Y'. These records need to be dropped when creating integrated datasets from core studies and LTS study to avoid data duplication.

Below is an Example of SDTM VS for LTS study:

DOMAIN	USUBJID	VSTESTCD	VSORRES	VSBLFL	VISIT	VSDTC	VSDY	VSCORSTD	VSWK0FL
VS	111111-001-002-003	DIABP	90	Y	Baseline	2013-10-03	-168	Y	
VS	111111-001-002-003	DIABP	80		Visit 1/2	2014-03-20	1	Y	Y
VS	111111-001-002-003	DIABP	70		Visit 4	2014-04-15	27		
VS	111111-001-002-003	DIABP	70		Visit 6	2014-05-15	57		
VS	111111-001-002-003	DIABP	83		Visit 7	2014-06-12	85		
VS	111111-001-002-003	DIABP	85		Visit 8	2014-09-04	169		
VS	111111-001-002-003	DIABP	73		Visit 9	2014-11-27	253		

DOMAIN	USUBJID	VSTESTCD	VSORRES	VSBLFL	VISIT	VSDTC	VSDY	VSCORSTD	VSWK0FL
VS	111111-001-002-003	DIABP	87		Visit 10	2015-02-19	337		
VS	111111-001-002-003	DIABP	72		Visit 11	2015-05-14	421		

4. INTERVENTIONS DOMAINS SUCH AS EX AND CM

- EX domain, in general, only based on LTS operational data.
- CM domain includes up-versioned prior and concomitant medication from core studies as well as prior and concomitant medication collected in LTS own operational database. The CM records obtained from core studies are flagged with CMCORSDT='Y'.

5. EVENTS DOMAINS SUCH AS AE, MH, AND CE

- AE and CE domains should be only based on LTS operational database. It is important to pay particular attention to database cleaning and reconciliation. For example, a subject's AE that occurred in core study and continued in LTS study, can potentially be entered into both core and LTS study's database. Special attention is need from site and DM to enter an event into correct study's operational database to avoid duplicates in the integrated datasets.
- MH includes all the relevant medical (or surgical) history data during the lifetime of the patients, which are usually only collected in core studies, plus MH records if collected in LTS study. In summary, MH includes up-versioned medical (or surgical) history of core studies and medical (or surgical) history of LTS study if applicable. The MH records obtained from core studies are flagged with MHCORSDT='Y'.

SPECIAL CONSIDERATIONS AT SUBMISSION STAGE

- Set ORIGIN to "Derived" for data obtained from core studies to avoid providing related core studies' annotated CRF pages.
- LTS study specific Pinnacle 21 issues, are explained in SDRG Data Conformance Summary. Below are a few examples:
  - DM – no baseline results for some subjects in PC domain.
    For PK concentration, baseline is not defined for subjects who are in Placebo treatment group. As these subjects were in 'Placebo' treatment group in core studies, the baseline was not defined in core studies. Therefore the baseline was not defined for these patients in LTS study as the baseline of LTS study is the same as the baseline of core studies.
  - DM – no records for 'Screen Failure' subjects are found in IE domain.
    This is due to LTS study is still ongoing at interim lock. Subjects who have signed informed consent form, but their eligibility and treatment allocation are not yet available as of interim lock data cutoff date are set as 'Screen Failure' (per core study's convention), though they are not true Screen Failure subjects. This usually occurs only in LTS study due to long study duration and subjects' ongoing enrollment.
  - Duplicate records in MH domain.
    This is due to some MH events are collected in both core studies and LTS study. The records obtained from core studies are flagged with MHCORSTD='Y'.

CONCLUSION

This paper shared some tips and special considerations in creating SDTM datasets for LTS study of multiple core phase studies. These tips are useful and practical not only in helping ease LTS study SDTM datasets creation and potential datasets pooling for integrated summary analyses, but also in complying with the CDISC SDTMIG and submission guidelines. These hands-on experiences can be applied to similar LTS study SDTM datasets development to avoid Pinnacle21 violations and improve SDTM datasets quality at submission stage.

REFERENCES

The CDISC SDTM implementation guide (SDTMIG) (http://www.cdisc.org/sdtm)

Pinnacle21 (https://www.pinnacle21.com/)

FDA eCTD guidance (https://www.fda.gov/downloads/forindustry/datastandards/studydatastandards/ucm384744.pdf)

CONTACT INFORMATION

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

Yunzhi Ling	Helen Wang
Sanofi	Sanofi
55C-330A, 55 Corporate Drive	55C-330A, 55 Corporate Drive
Bridgewater, NJ 08807	Bridgewater, NJ 08807
908-981-6169	908-981-4752
Yunzhi.ling@sanofi.com	Helen.wang@sanofi.com