

In the fast-paced world of clinical research, efficiency is a key priority. Splitting SDTM domains into multiple datasets can serve as a valuable tool, streamlining the process of creating and reviewing large, complex data domains.

Reasons to Split

- Improved traceability from Case Report Forms (CRF) to datasets
- Simplified programming and validation
- Dataset is larger than 5GB
- Aid in streamlined reviews
- Strategic segmentation for targeted analyses
- Improved compliance with FDA regulations

Splitting Methods

- Split by purpose, for example QS split by different questionnaires
- Split by category or timepoints, for example LB split by category (Hematology, Chemistry) or time points (Baseline, Week 12)

Define-XML Documentation

- Document in Define XML: Splitting due to unique datasets
- Don't Document in Define: Splitting due to file size

Practical Tips

Example

A study includes multiple questionnaires, such as a Quality-of-Life Questionnaire and a Symptom Severity Questionnaire. These questionnaires have different structures, scales, and scoring methods. SDTM will be split with the Quality-of-Life Questionnaire mapped to the QSCS and the Symptom Severity Questionnaire mapped to QSHA.

Variables

- Use consistent naming and lengths for variables across datasets, so split datasets can be easily concatenated together.
- The value of DOMAIN should be consistent across datasets as if they had not been split (ex: DOMAIN='QS' in QSCS and QSHA)
- Variables requiring domain prefix must use DOMAIN as prefix (ex: --TESTCD should be QSTESTCD in both QSCS and QSHA)
- --SEQ must be unique with no repeat values within USUBJID for all records across the split datasets

Annotating CRF

Annotate with domain names not dataset names (use QS rather than QSCS/QSHA)

Dataset Directories

Split data due to large file size should be submitted in a separate sub-directory, and main (non-split) dataset in the SDTM directory. If splitting is due to unique dataset definitions, each split dataset should be included in the SDTM datasets directory.

Sources:

- SDTMIG v3.4 | CDISC: <https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-v3-4>
- FDA: Study Data Technical Conformance Guide

<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/study-data-technical-conformance-guide-technical-specifications-document>