Study Data Reviewer's Guide (SDRG)

1. Purpose of This Document

This Study Data Reviewer's Guide (SDRG) provides supplemental information to assist reviewers in understanding the structure, content, and derivation of the SDTM datasets generated for a mock clinical study. The SDRG describes key derivations, assumptions, and data handling decisions that are not fully apparent from the datasets alone.

The datasets were created for training and demonstration purposes and do not represent data from real subjects.

2. Study Overview

This is a simulated Phase II clinical study evaluating the safety of Drug ABC in adult subjects. Subjects were followed for approximately eight weeks. Safety-related data collected included demographics, adverse events, and laboratory assessments.

The following SDTM datasets are included:

- DM (Demographics)
- AE (Adverse Events)
- LB (Laboratory Test Results)

3. General Data Conventions

- All SDTM datasets follow CDISC SDTM naming conventions.
- Character date variables use ISO 8601 format (YYYY-MM-DD).
- Study day variables (–DY) are derived relative to the reference start date (RFSTDTC).
- Sequence variables (–SEQ) are assigned sequentially within each subject.
- Only variables present in the datasets are described in this document.

4. Demographics (DM)

4.1 Dataset Description

The DM dataset contains one record per subject and provides subject-level identifiers and reference dates used across all other SDTM domains.

4.2 Key Variables and Derivations

- **STUDYID**
  Assigned as a constant value ("ABC123") for all records.

- **DOMAIN**
  Set to “DM”.
- **SUBJID**
  Derived from the raw subject identifier and standardized to a three-character format with leading zeros.
- **USUBJID**
  Unique subject identifier derived by concatenating STUDYID and SUBJID.
- **RFSTDTC**
  Reference start date derived from the first dose date in the raw data and converted to ISO 8601 format. This date serves as the anchor for all study day derivations.

4.3 Comments

- DM was created prior to all other SDTM datasets.
- RFSTDTC is used consistently across AE and LB for timing derivations.

5. Adverse Events (AE)

5.1 Dataset Description

The AE dataset contains one record per adverse event per subject and describes adverse events occurring during the study.

5.2 Key Variables and Derivations

- **STUDYID**
  Constant value (“ABC123”).
- **DOMAIN**
  Set to “AE”.
- **USUBJID / SUBJID**
  Merged from DM.
- **AETERM**
  Mapped directly from the raw adverse event description.
- **AESTDTC**
  Adverse event start date converted to ISO 8601 format from normalized raw dates.
- **AEENDTC**
  Adverse event end date converted to ISO 8601 format when available. Missing for ongoing events.
- **AESTDY**
  Derived as (AESTDTC − RFSTDTC + 1) when both dates are present.
- **AESEV**
  Standardized from raw severity values.
- **AEREL**
  Standardized from raw relationship-to-study-drug values.

- **AESER**
  Seriousness flag derived as “Y” or “N” based on raw seriousness indicators.
- **AESEQ**
  Sequential identifier assigned per subject in chronological order of adverse event start date.

5.3 Comments

- Events occurring prior to RFSTDTC result in negative AESTDY values.
- AESEQ uniquely identifies each adverse event record within a subject.

6. Laboratory Test Results (LB)

6.1 Dataset Description

The LB dataset contains laboratory test results collected during the study. Each record represents one laboratory test for a subject at a given visit.

6.2 Key Variables and Derivations

- **STUDYID**
  Constant value (“ABC123”).
- **DOMAIN**
  Set to “LB”.
- **USUBJID / SUBJID**
  Merged from DM.
- **LBTESTCD**
  Laboratory test short code mapped from raw data.
- **LBTEST**
  Laboratory test name mapped from raw data.
- **LBORRES**
  Original laboratory result as collected.
- **LBORRESU**
  Original unit of measure.
- **LBSTRESC**
  Standardized character result derived from LBORRES.
- **LBSTRESN**
  Standardized numeric result derived when LBORRES contains numeric values.
- **LBDTC**
  Laboratory collection date converted to ISO 8601 format from normalized raw dates.
- **LBDY**
  Derived as (LBDTC − RFSTDTC + 1) when both dates are present.
- **VISIT**
  Visit name mapped from raw data.
- **VISITNUM**
  Numeric visit identifier mapped from raw data.

- **LBSEQ**
  Sequential identifier assigned per subject in chronological order of laboratory collection date.

6.3 Comments

- Laboratory results with non-numeric values have missing LBSTRESN.
- Missing or partial laboratory dates result in missing LBDY.
- LBSEQ uniquely identifies each laboratory record within a subject.

7. Data Validation and Quality Checks

- Subject identifiers were verified for consistency across DM, AE, and LB.
- Study day derivations were reviewed relative to RFSTDTC.
- Sequence variables were confirmed to be unique within subject.
- Date conversions were validated to ensure correct ISO 8601 formatting.

8. Limitations

- This is a simulated dataset created for educational and demonstration purposes.
- Only selected SDTM variables are included.
- Define-XML is not provided.

9. Conclusion

The DM, AE, and LB datasets were generated using standard SDTM principles, with explicit handling of non-standard raw data, consistent subject-level anchoring, and transparent derivation of timing and sequence variables. This SDRG documents the key decisions required to interpret the datasets accurately.