SOP 003_04

Title	Database Set-up
SOP Code	SOP 003_04
Effective Date	30-June-2023

Site Approval/Authorization to Adopt

Name and Title of Local Personnel (Type or print)	Signature	Date dd/Mon/yyyy

1.0 PURPOSE

This Standard Operating Procedure (SOP) describes the tasks necessary to generate, test, implement and approve a database set-up for a study to ensure accurate, reliable, complete and secure electronic data.

2.0 SCOPE

This SOP is applicable to all studies databases requiring set-up at the site and to those personnel responsible for study database set-up such as all Data Management and Information Technology (IT) personnel unless otherwise stated in the study contract or Data Management Plan.

3.0 RESPONSIBILITIES

The Sponsor-Investigator or Qualified Investigator (QI)/Investigator, Data Management personnel and IT Systems Support personnel (when applicable) are responsible for ensuring that the processes involved in all database system set-up meet all the applicable regulatory, International Conference on Harmonization (ICH) Good Clinical Practice (GCP), Sponsor, and local requirements.

Any or all parts of this procedure may be delegated to appropriately trained study team members, but remain the ultimate responsibility of the Sponsor, Sponsor-Investigator and/or Qualified Investigator (QI)/Investigator.

4.0 DEFINITIONS

Annotation (of the eCRF): A representation of the variables and tables in the paper CRF or data entry screens.

Coded data (de-identified): Single code: A participant's data are assigned a random code. Direct identifiers are removed from the dataset and held separately. The key linking the code back to direct identifiers is available only to authorized members of the research team. Double or multiple codes: Two or more codes are assigned to the same participant's data held in different datasets (e.g. health administrative data, clinical data, genetic samples and data). The key connecting the codes back to participants' direct identifiers is held by a third party (such as the data holder) and is not available to the researchers. Coded data refers to data that are at least single coded. (Adapted from CIHR Best Practices for Protecting Privacy in Health Research – September 2005) Data dictionary: The repository for all the required objects for creating and maintaining data collection, validation, and extraction operations. Database set-up: A collection of data entry screens that defines, through Data Management software, the database structure. The database set-up will be defined according to the study protocol. For paper based Case Report Form (CRF), the database set-up will be defined according to the CRF and the study protocol.

Direct identifiers: These are variables such as name and address, health insurance number, etc., that provide an explicit link to a respondent. (CIHR Best Practices for Protecting Privacy in Health Research – September 2005) Edit check: An auditable process, of assessing the content of a data field against its expected logical, format, range, or other properties that is intended to reduce error. NOTE: Time-of-entry edit checks are a type of edit check that is run (executed) at the time data are first captured or transcribed to an electronic device at the time entry is completed of each field or group of fields on a form. Back-end edit checks are a type that is run against data that has been entered or captured electronically and has also been received by a centralized data store. (CDISC) Identifiable data: Any element or combination of data elements that allows direct or indirect identification of an individual (i.e. via direct identifiers or indirect identifiers). (Adapted from CIHR Best Practices for Protecting Privacy in Health Research – September 2005)

Indirect identifiers: These are variables such as date of birth, sex, initials, marital status, area of residence, occupation, type of business, etc. that, in combination, could be used to identify an individual. (CIHR Best Practices for Protecting Privacy in Health Research – September 2005)

User Acceptance Testing (UAT): A formal means by which one verifies that the system meets the required business functions by emulating normal use conditions.

See also, “CDISC Clinical Research Glossary, Version 8.0” and “N2 Glossary of Terms”.

5.0 PROCEDURE

4.1 Database Set-up

5.1.1 Determine data variables to be captured at each specific visit according to the protocol and user requirements that may be beyond protocol.

• Determine the event schedule, which specifies the segments and study variables expected as per the study protocol

• Create the data dictionary which identifies software version, variable names, type (e.g. character, date, time, derived, numeric) and associated attributes.

• Review and approval of the data dictionary

• Define the applicable code-lists and the group of fields that will reflect the data storage

• Annotation of the CRF/eCRF (where applicable)

• Review and approval of the annotation of CRF/eCRF

• Create database user's manual

5.1.2 Ensure planning of database set-up adheres to protocol timelines.

5.1.3 Ensure requirements are defined for data transfers and integration with other systems, such as, but not limited to, laboratory databases or medical dictionary linkages (MedDRA, WHOdrug).

5.1.4 Design the data entry screens and ensure they are user-friendly and flexible for data entry. Take into consideration the data collection method being used when setting-up the database.

5.1.5 Define items/fields to be coded and the standard dictionaries to be used (e.g. MedDRA, WHOdrug), if applicable.

5.1.6 Create specifications for database access authorization (user roles, e.g. data entry, and accounts).

5.1.7 Identify and program the edit checks and validation rules

• Perform UAT of the database and edit checks against user requirements or specifications, with expected and unexpected data in a testing environment.

5.1.8 Ensure privacy legislation provisions (local, provincial, national, and international, as applicable to the study) are met within the database for identifiable data (including coded data).

5.1.9 Ensure database finalization and agreement of approval by Data Management and the Sponsor (if applicable) are complete. Document appropriately that test and production environments match with UAT.

6.0 REFERENCES

Health Canada, Food and Drug Regulations, Part C, Division 5, Drugs for Clinical Trials Involving Human Subjects, (Schedule 1024), June 20, 2001.

Health Canada, Guidance for Industry, Good Clinical Practice: Consolidated Guideline, ICH Topic E6, 1997.

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada, Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, December 2014.

Department of Justice (Canada), Personal Information Protection and Electronic Documents Act (PIPEDA), updated 2006.

Pharmaceutical Inspection Convention, Pharmaceutical Inspection Co-operation Scheme, Annexe 11, Computerised Systems.

CDISC Clinical Research Glossary, Version 8.0, Glossary. December 2009. Canadian Institutes for Health Research, Privacy Advisory Committee, CIHR Best Practices for Protecting Privacy in Health Research, September 2005.

US Food and Drug Administration, Code of Federal Regulations, Title 21, Volume 1:

• Part 11, Electronic Records; Electronic Signatures, (21CFR11).

• Part 50, Protection of Human Subjects, (21CFR50).

• Part 56, Institutional Review Boards, (21CFR56).

US Department of Health and Human Services, Office of the Secretary, 45 CFR Parts 160 and 164, Standards for Privacy of Individually Identifiable Health Information.

US Department of Health and Human Services. Food and Drug Administration. Office of the Commissioner. Guidance for Industry, Computerized Systems Used in Clinical Investigations. Guideline. May 2007.

Official Journal of the European Communities, Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995.

Official Journal of the European Communities, Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001. Medical Dictionary for Regulatory Activities (MedDRA), Maintenance and Support Services Organization (MSSO).

The Society for Clinical Data Management, GCDMP Committee, Good Clinical Data Management Practices. December 2009 Ed. WHO Drug Dictionary, Uppsala Monitoring Centre (UMC).