

University of Hertfordshire
DATA MANAGEMENT OVERVIEW
Sponsored/Co-sponsored
Clinical Trials Support Network (CTSN)

Standard Operating Procedure for Clinical Trial Data Management at the
University of Hertfordshire

SOP Number: gSOP-40-02	Effective Date: 10/08/22
Version Number: 2	Review Date: 3 years (or as required)

1. BACKGROUND

This is a University of Hertfordshire standard operating procedure.

This document sets out the procedures to be followed by all UH staff who are involved with data management processes to ensure that all clinical trial data shall be recorded, handled, and stored in a way that allows its accurate reporting, interpretation, and verification in accordance with the approved clinical trial protocol, the principles of Good Clinical Practice (GCP) and with the UH policy on data management.

Where there are potential conflicts between different collaborating organisations' SOPs, project level working instructions should be developed, to determine precedence.

2. PURPOSE

- To describe the process for developing, validating, and managing the databases used in UH sponsored/co-sponsored trials.
- To provide guidance for the management of trial related data and therefore ensure that all data is collected, verified, and analysed to assure that trial data is accurate.
- To provide guidance for the processing of clinical trial data (entry, uploading, cleaning and query management) and production of final dataset(s) ready for analysis.

3. APPLICABLE TO

Any UH employee involved with, but not limited to, the management of data for clinical research including Chief Investigators (CI), Principal Investigators (PI), Consultants, Clinical Trial Pharmacists, Trial Managers, Monitors, Statisticians, Statistical programmers, Research Assistants, Data Managers & students.

“ICH GCP guidelines state that only appropriately qualified individuals should supervise trial data handling, verify the data and conduct the statistical analyses (ICH 5.5)”.

4. RESPONSIBILITIES

The Data Manager (DM) is responsible for liaising with the statistician about the design of the database and for the development and approval of the study database. The Data Manager is also responsible for the initiation and maintenance of the Data Management Plan (DMP).

The Statistician is responsible for the overall design and validity of the database and will provide test output for database validation.

The Trial Manager (TM) is responsible for the review of the database and assisting in database testing. During study conduct the Trial Manager will consider circumstances that may be leading to high rates of errors.

The Chief Investigator has overall responsibility for ensuring the database meets the requirements of the study.

5. PROCEDURE

Once a project has been funded the data manager (DM) will establish contact with the CI/PI or TM to gather, refresh and confirm their data management requirements and to outline the Data Management processes and possibilities.

The following precursors and documents are required:

- A fully approved trial protocol or trial protocol awaiting regulatory approval.
- Trial Risk Assessment where appropriate.
- Data Management Plan (DMP) where appropriate (see Appendix 3 - Data management plan template).
- Evidence of database specific training for the system in use, for developers and users.
- A validated system for capturing electronic trial data.

The latest versions of the documents must be requested from the CI/TM. The documents and their locations should be referenced in the DMP. It is best practice to have a DMP for every trial. The development of the DMP is led by the DM. The following should be considered in a DMP:

- a. Study team members and roles
- b. Data Management team members and roles for this study
- c. Other relevant Data Management procedures
- d. Data sources
- e. Systems
- f. Validation
- g. Query process
- h. Quality assurance
- i. Quality control
- j. Handling protocol non-compliance
- k. Pharmacovigilance
- l. Training and documentation
- m. Database users
- n. Location of data
- o. Archiving

There will be discussion between DM and CI to note what is possible and what is required noting that the most stringent requirements will apply to data management systems for CTIMPs and Medical Device trials. This may involve tailoring the specification of the Trial Data Management System (TDMS) according to cost or time constraints and consideration should be given to the possible phasing of delivery.

The DMP will be developed and enhanced during the period leading up to the start of recruitment and data collection. All signed-off versions will be retained online in the study data management folder.

During development and during the period of the study the PI/CI or TM should keep the DM informed of any changes to system requirements that affect the DMP and the DBM will make the necessary amendments. All amendments should go through the validation process.

The DM will keep the DMP up to date.

Any revisions to the DMP during the lifetime of the study must be agreed and signed off by the DM and the CI or TM.

5.1 Data Capture Tools

The preferred method of data capture should be agreed:

- Paper CRFs (pCRFs) – are completed at participating sites and sent to the coordinating site for entry into the database.
- Electronic CRFs (eCRFs) –: data is entered directly into the database by participating sites without the step of a paper CRF.

Other data (e.g., lab data) may be imported directly from source into the database. Electronic imports must be validated as detailed in the Data Management Plan and the results documented. Data transferred from other sources must be managed securely.

5.2 Design and Use of Database for Statistical Analysis

The choice of database is dictated by the risk assessment and GCP regulations. For CTIMPs only a validated system can be used (gSOP-42 Data Management System Validation).

5.3 Database Development

Once the CRF has been designed and approved the database for the statistical analysis should be developed by the study's statistician and data manager.

The following sections will provide a description of the processes to be followed. The steps include:

- Design
- Build
- Test phase
- Live phase
- Data Entry and cleaning phase
- End of trial phase

5.4 Database Design

A database should be designed to ensure that it captures all the information that is required according to the protocol. The study schedule usually documents what data is to be captured at which time points during the trial and helps form the basis for developing the database. The database should directly reflect the content of the Case Report Form (CRF) and only include the parameters which will be included in the final analysis. The database should include:

- Dataset names
- Variable names
- Variable formats

5.4.1 Data Dictionary

A data dictionary may be created which provides a description of the data to be collected during the study including data types and value ranges. It also describes how data is divided into categorized sets and the relationship between these sets. This should be agreed between the statistician, data manager and study team. This is a working document and may change during the development of the TDMS. Any significant changes should be discussed with the appropriate study team members.

5.4.2 Functional Specification

The Functional Specification contains:

- Description with illustrations of the user interface, where data is entered, reviewed and

updated.

- Description of standard data validation to be performed.
- Description of any data transformations that take place.
- Description of any non-standard functionality and the circumstances where it occurs.

The Functional Specification is a descriptive document to be used as a guide by programmers implementing the system, consideration should be given to the fact that the approvers are unlikely to be computer experts and therefore computing jargon should be avoided.

The Functional Specification should also be usable as a user reference and training manual for data entry staff.

The Functional Specification should be approved by the TM responsible for data entry before the TDMS goes live.

When the Data Dictionary and the Functional Specification have been approved by all reviewers the DM will produce the Test Plans which can be based on the Functional Specification.

5.4.3 Coding

The CRF responses must be coded before they can be entered onto the statistical database, therefore the codes should be determined before data entry begins, preferably when the database is being designed. It is important to ensure that codes are in place for the following eventualities: not done, not applicable and unknown to represent where data is missing and that these are recorded in the database coding manual.

5.4.4 Audit trail

The database should ensure that an electronic audit trail for the data is maintained with appropriate password protection to prevent unauthorised access to the data. The database should be maintained according to user requirements with appropriate levels of access. A list identifying individuals permitted to make changes to the data should be produced and filed in the Trial Master File.

The database should enable adequate backup of the data. If the study involves blinding, the data entry and processing systems should allow for this to be maintained.

5.5 Database Build

Once the design phase has been completed the build phase can begin. All database systems require an appropriate level of validation. The database should be validated to confirm that it is fit for purpose.

The statistician will review the database coding manual, Statistical analysis plan and CRFs.

Appropriate edit checks should be set up for the study. They may include but are not limited to:

- Presence checking especially with reference to required fields
- Range checking
- Type check
- Length check
- Logic check

Once an appropriate specification for the data edits checks have been specified, the code will be applied to a range of dummy data (dummy edit checks) with known data errors. The outcome of the dummy edit check will be compared against the known errors and recorded. Any amendments necessary will be made, and the dummy edit check repeated.

A copy of the edit check specifications document and any amendments should be filed in the Trial Master File.

5.6 Quality Assurance and Validation

Test Plans should be written documenting how testing of the database will be undertaken. The test phase is used to ensure the requirements in the design phase are implemented correctly and the database functions as specified. There should be one test for internal testing by Data Management and one test for user testing. 'Standard' tests such as navigation between pages working correctly and data loading and saving should be included. Each specific operation in the specification (e.g., checking that a follow-up date is within a certain range compared to randomization) will be subject to its own tests. The Test Plan will, by default be produced by copying the relevant sections of the Functional Specification and adding for each function a box to indicate that it has been tested correctly, or a note to state the nature of the failure. The level of detail and testing will vary according to the risk assessment and the DMP.

5.6.1 Test system Deployment

Testing is an essential, mandatory part of the data management process. When the Data Dictionary and the Functional Specification have been approved by all reviewers, and the system has been built and undergone initial development testing, the DM will deploy the system for user testing.

The DM will install the system in the Test Environment and will work through the Test Plan noting the success or failure of each test. A System Test report will be produced. Any failures must be fixed and the tests re-run. All test plans and associated results should be stored in the Data Management filing system.

The DM will then set up a login account on the Test environment for one or more system user tester.

Once the system tester is satisfied with the Test system the TM should sign the system acceptance form. The system acceptance form should also be signed by the study statistician.

The database should then be finalised and approved. Once approved the database can be made live.

The TM will provide the DM with a list of Users and roles to which they should be assigned in the data management system. This Users Log must be kept up to date throughout the lifetime of the study. The DM will set up the required accounts and send details to the individual users. Passwords will be sent separately to login details and must only be sent to the user in question with a reminder that they must not be divulged. When required, system specific guidance and training materials should be provided to end users.

The data management system will now be under Change Control. Any bugs reported and enhancements requested will be tracked along with a status.

5.7 Data Entry and Cleaning

The single data entry method is the most commonly used method for UH sponsored/co-sponsored trials. Once the data has been entered onto the eCRF/database checks should be completed to ensure the accuracy of the data recorded. For each trial, the process for data entry should be standardised and documented.

For UH sponsored/co-sponsored CTIMPs, data verification is arranged by another member of the project team, through source data verification, as stated in the protocol/monitoring plan. The person responsible for verification will be named on the delegation log and the statistical analysis plan.

Where data queries arise from CRF review, the corrected CRF should be entered onto the database by a delegated individual. For multicentre studies where the CRFs are being sent to a coordinating centre for data entry a copy should be kept in the site file. It is important for the coordinating centre data manager to keep a log of all the CRFs received.

5.7.1 Manual Data Entry (onto database)

- Forms and data fields should be checked to ensure sites are following completion guidance.
- If data is missing from SAE forms, the PI/Local Researcher will be contacted to ensure reporting timelines are adhered to as specified in Study Protocol and gSOP-02 Adverse Event Reporting.
- The database/worksheet to which data from paper CRF's or questionnaires must bear one to one correspondence to the original.
- Text and/or numerical data should be entered as seen, regardless of spelling although obvious errors may be corrected (e.g., date errors at beginning of a new year) if the error is recorded.
- Where there are issues with legibility on forms or CRF's completed at local sites, site personnel should be contacted to resolve and draw attention to the need for legibility.

5.7.2 Data Cleaning

To ensure the accuracy of the data set for the statistical analysis an important process is the 'cleaning' and validation of the data. This process should be done as defined in study protocol and Quality Management and Monitoring Plan (QMMP) if applicable.

5.8 Maintenance and support

Once the trial data management system has gone "live" there is a requirement for a structure to be in place to deal with issues arising during study conduct. This would include the management of bug reports and any updates and enhancements. (See gSOP-43 Trial Data Management System Maintenance and Support).

5.9 Management of Amendments to CRFs and Database

All amendments involving CRF collection in clinical trials require authorisation by the TM as per gSOP-09.

The study DM should review the amendment to assess the impact on CRF design and study database. The investigator should be advised of any potential changes required to CRF and study database.

Any changes to a database should be controlled and a clear audit trail should be present.

5.10 End of Trial Phase

Once the trial data is as complete and clean as possible to be it is placed in a state of hard lock. For activities after database lock see gSOP-44 Database closedown and manipulation of data after export.

6. RELATED DOCUMENTS

- UPR IM16 Appendix III Data Management Policy-University Guide to Research Data Management
- gSOP-33 Risk Assessment and Risk Rating
- gSOP-15 Case Report Form Design
- gSOP-42 Data Management System Validation
- gSOP-43 Data Management System Maintenance and Support
- gSOP-44 Data Management Closedown and Manipulation of Data after Export
- MHRA, 2012, Good Clinical Practice Guide

7. APPENDICES

- Appendix 1- Definitions
- Appendix 2- Data Management Process Flow Chart
- Appendix 3 – Data Management Plan Template

8. VERSION HISTORY/REVISIONS

Version Number	Effective Date	Reason for Change
02	10/08/22	Periodic review

9. AUTHORSHIP & APPROVAL

Author

Signature



Date 23/07/2022

**Pro Vice-Chancellor (Research & Enterprise) Approval
Professor J M Senior**

Signature



Date 08/08/2022

10. AGREEMENT (MOVE ON TO SEPARATE SHEET BEFORE PRINTING)

Please detach and retain within your training files

I have read and understood the contents and requirements of this SOP (ref gSOP-040-02) and accept to follow University policies implementing it.

Recipient

Signature:Date:

Name & Position:

Appendix 1: Definitions

Adverse Event (AE)

Any untoward medical occurrence in a subject to whom a medicinal product has been administered, including occurrences which are not necessarily caused by or related to that product.

Case Record Form (CRF)

A printed, optical, or electronic document designed to record all of the protocol required information to be reported to the sponsor on each trial subject”.

Chief Investigator (CI)

A Registered Physician, Dentist, Pharmacist or Registered Nurse who has overall responsibility for the conduct of the trial.

Clinical Trial

A research study in which one or more human subjects are prospectively assigned to one or more interventions (which may include placebo or control) to evaluate the effects of those interventions on health outcomes.

Clinical Trial of Investigational Medicinal Product (CTIMP)

A study that looks at the safety or efficacy of a medicine/food stuff/placebo in humans as defined by the Medicines for Human Use Regulations (2004).

Delegated Individual (DI)

An individual delegated by the PI to carry out their task(s).

Good Clinical Practice (GCP)

As defined in the Regulations.

International Conference on Harmonisation (ICH)

The ICH produced a series of guidelines in 1996, E6 being the guideline on Good Clinical Practice, otherwise known as (ICH-GCP).

Principal Investigator (PI)

A Registered Physician, Dentist, Pharmacist or Registered Nurse who has responsibility for the conduct of the trial at a host site.

Serious Adverse Event (SAE) or Serious Adverse Reaction (SAR)

Any untoward medical occurrence or effect that at any dose results in:

- Death
- Is life-threatening*
- Requires hospitalisation or prolongation of existing hospitalisation
- Results in persistent or significant disability or incapacity
- Is a congenital anomaly or birth defect
- Is an important medical event

* “life-threatening” refers to an event in which the patient was at risk of death at the time of the event; it does not refer to an event which hypothetically might have caused death if it were more severe

Site File

Site Files are held by the Principal Investigator at sites and contain copies of the essential documents, local approvals, signed consent forms and completed data forms.

The Regulations

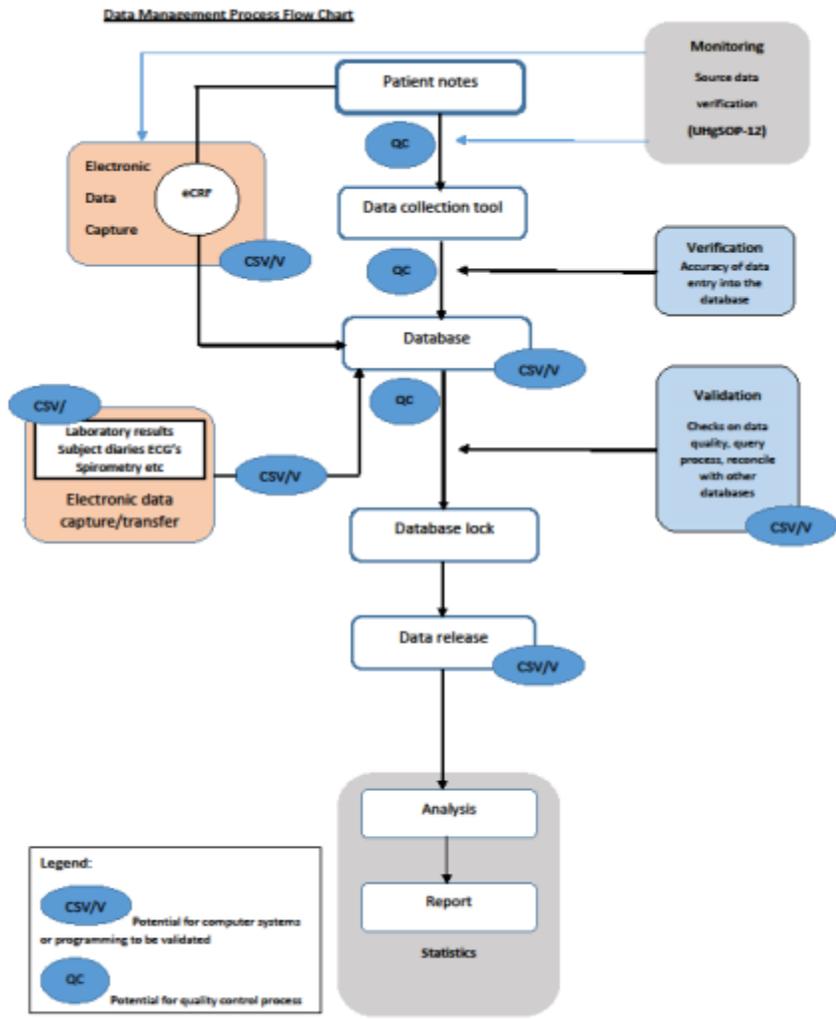
Medicines for Human Use (Clinical Trial) Regulations 2004 transposed the EU Clinical Trials Directive into UK legislation, as Statutory Instrument 2004 no 1031. This became effective on the 1st May 2004. An amendment to implement Directive 2005/28/EC was made to the Regulations as Statutory Instrument 2006 no 1928.

Trial Master File

The Trial Master File (TMF) will be held at the principal site by the sponsor, Chief Investigator or at the co-ordinating Centre. The TMF should contain all essential documents defined as documents which individually and collectively permit evaluation of the conduct of a trial and the quality of the data produced. A TMF should be set up at the beginning of a trial and maintained up to date throughout the trial until trial conclusion.

For trials currently running, it is recommended that Section 8 of the ICH-GCP Guideline is followed as guidance to meet statutory requirements. However, some of the documents listed may not be available or applicable in many non-commercial trials. The appropriate documentation will vary according to the trial and sponsor requirements.

Appendix 2: Data Management Process Flow chart



Appendix 3 - Data Management Plan Template

The purpose of a Data Management Plan (DMP) is to define and provide instructions regarding the conduct of data management processes and the flow of data from source data to data release for analysis.

A DMP is recommended for large, long-running, multi-centre trials. However, for clarity of trial management processes, data flow and responsibilities it is recommended to put a proportionate DMP in place also for smaller, single-centre trials, based on a risk-assessment approach. Not all headings of this template are necessarily needed for all trials and additions of other headings should be introduced when necessary.

The scope of this DMP is not to define CRF design, data base design, data base validation and testing, plan of data analysis, data base security. These processes are defined and detailed in other SOPs. The DMP will detail trial specific processes or will just refer to SOPs if those are followed without any specifics.

The DMP will describe:

- 1) CRF workflow
- 2) Data entry
- 3) Data cleaning – data query process - quality checks
- 4) Definition and location of source data
- 5) Database lock
- 6) Data release
- 7) Reports
- 8) Transferring/downloading data/coding/PV data base reconciliation
- 9) Location of data and plan for data archiving

- Appendix 1- Definitions
- Appendix 2- Responsibilities and scope of work (responsible individuals/groups)
- Appendix 3 - Applicable SOPs and guidelines
- Appendix 4 - Required documentation to be produced

1. CRF Workflow

*Study specific handling/logistics for delivering CRFs to sites;
completion of CRFs; amendment and implementation process for CRFs
When and how are CRFs delivered for data entry, etc.*

2. Data Entry

*When, how, by whom will data be entered; Study specific entry guidelines (Give reference to Data Entry Guidance document if applicable)
Specify any data which is collected, but not entered into the actual trial data base – i.e. Sponsor SAE form data.*

3. Data Cleaning, data query process, quality checks

By whom, when and how?

Edit check specifications (can reference data validation plan)

Data management self-evident corrections (need for agreement - if at all)

Handling guidelines for queries and chasing of CRFs

Query process/flow and tracking including process for editing data

Different levels and steps of quality check- i.e. checks of data from source to CRF (SDV); check of data from CRF to data base (quality check of data entry); check

before data lock for completeness, accuracy;

define percentage of SDV, CRF data entry checks;

Any specific monitoring

4. Definition and Location of source data

Detail here what is considered source data and if any source data is directly entered into CRF pages; specific actual location of source data

5. Data base lock

Give details

6. Data release

Details of when data can be released, by whom data release can be requested, and relevant SOPs/Guidance

7. Reports

Timing and frequency of planned reports where interim data downloads are necessary –i.e., for DMECs, interim analysis or other justified purposes. List of standard reports and frequency of report

8. Transferring/ Downloading Data/Coding/PV database reconciliation

Details of any requirements and arrangements in this trial where data will be downloaded / transferred directly into the trial database as electronic data set from another source –for example laboratory systems etc. And details if any specific coding will take place (when and by whom) before data is entered into data base;

9. Location of data and plan for data archiving

Include here whether any of the data (source or CRF data) will be located outside the normal database, TMF/ISF e.g. in hospital systems/specialist software packages or in a standalone file.

Details of relevant SOPs to be followed; can include any specific study closeout checklist; details of format, how and where data will be archived – if known; otherwise just “suitable archiving method to be decided and agreed at a later date”

Appendix 1 - Responsibilities and scope of work

Team Member	Role	Contact Details
	Data entry	
	Quality checks	
	Data lock	
	Data release	
	Reports	
	Etc.....	

Appendix 2 Definitions – examples only - amend as required

Term	Definition
CI	Chief Investigator
CTIMP	Clinical Trial of an Investigational Medicinal Product
DM	Data Manager
CRF	Case report form
PI	Principal Investigator
Validations	Edit checks which are programmed into the data entry system
DE	Data Entry
AE	Adverse Event
DMP	Data Management Plan
SAE	Serious Adverse Event

Appendix 3 - Applicable SOPs and guidelines

Add current table listing all relevant SOPs

Appendix 4 - Required documentation to be produced

If applicable - Add here list of any specific logs/signature/forms needed