University of Leicester Research Governance Office

Standard Operating Procedures

SOP S-1036 UoL

Data Management Process for Research Sponsored by University of Leicester

Version 3.3 January 2024

Office Base
Research Governance Office
University of Leicester
Academic Department, Ground Floor
Leicester General Hospital
Gwendolen Road
Leicester
LE5 4PW

Effective Date: January 2024

This SOP will be implemented in line with this document's effective date for all UoL Sponsored research still in set up. For active clinical research that is already in the recruitment phase (or further) at the time of implementation, this SOP must be implemented within 3 months of the effective date.

1.0 Introduction and Scope

The purpose of this SOP is to provide guidance for managing data collected for research purposes and ensuring all data are collected, verified and analysed in the appropriate manner for all studies sponsored by the University of Leicester (UoL).

Data management processes are required to ensure that the data included in reports and/or publications of the study results are accurate and were captured in accordance with the approved protocol.

2.0 Definition

The data management process covers the design and production of the data capture tool (see SOP S-1039, Case Report Form (CRF) production), the design and construction of the database, the processing of the data (entry, uploading cleaning and query management) and production of the final dataset(s) ready for analysis. Quality control (QC) should be applied at each stage of data handling to ensure that all data are accurate, reliable and have been processed correctly.

3.0 Delegation

In accordance with GCP, data handling must be conducted by appropriately qualified and trained individuals. UoL as Sponsor delegates the responsibility for data management to the Chief Investigator (CI). Where this is further delegated to another member of the research team, this must be clearly documented on the Delegation of Authority and Signature Log (S-1010 Appendix 2). When delegated to an external organisation, or to another team or a Clinical Trials Unit (CTU), the roles and responsibilities, and where applicable timelines, must be clearly detailed in an appropriate agreement.

4.0 Computer System Validation

There are three key features of computer system validation which must be complied with and for which there must be documentary evidence. These are:

- Appropriate controls of the system are in place throughout the system's lifetime
- Documentation is available to support the application of the controls
- The system is fit for purpose and performs reliably and consistently as intended.

The database system is then used to create a study-specific database and data entry system. This should also be validated to confirm that the database system is fit for purpose specifically for the study.

There should be a mechanism in place for version control of the system and a formal process to manage any changes to this to ensure the validated state is maintained.

4.1 Choice of Data Management Database

UoL conducts a wide range of research studies ranging from simple qualitative studies to complex highly regulated Clinical Trials of Investigational Medicinal Products (CTIMPs). The type of database used should be proportionate to the risk associated with the study and this will be determined during the Sponsor Risk Assessment process. Where possible the services of a CTU should be used.

Whilst the type of database used (e.g., the use a validated database is particularly important in the case of CTIMPs) will be proportionate with the associated risk, the underlying principles remain the same in that the Sponsor must ensure that adherence to the principles of GCP which include “All clinical trial information should be recorded, handled and stored in a way that allows its accurate reporting, interpretation and verification” are followed.

4.2 Electronic Data Handling Systems

When using electronic data handling or remote electronic data systems, the following Sponsor requirements must be complied with:

a) The system must be validated and there must be documentary evidence of the validation
b) SOPs must be in place for use of the system
c) The system must be designed in such a way that there is no deletion of entered data (i.e., an audit trail is maintained)
d) A security system must be in place which prevents unauthorised access to the data
e) A list must be maintained of the individuals authorised to make data changes
f) The data must be adequately backed up
g) The blinding (if applicable) must be maintained during data entry and processing
h) If data are transformed during processing it should always be possible to compare the original data and observations with the processed data
i) An unambiguous subject identification code should be used which allows identification of all the data reported for each subject.

The data entry process should be defined for each specific study.

5.0 Data Capture and Entry

The Principal Investigator (PI) is responsible for ensuring the accuracy, completeness, legibility and timeliness of the data reported at their research site. It is typical for research data to be initially captured on paper CRFs and then subsequently transposed/entered into the electronic database. Where this is the case, the following process should be followed:

1. Data captured on the paper CRF must be accurate, complete and legible
2. Any data derived from source documents should be consistent with the source documents (or the discrepancy explained)
3. CRFs should be reviewed for any missing data, incomplete fields or data outside normal ranges
4. If discrepancies are picked up they should be clarified with the PI (or their delegate) and any changes to the paper CRF must be recorded, initialled and dated in a GCP compliant manner
5. A single line striking through the incorrect entry must be used so that the original data entry is not obscured
6. The data should then be transposed/entered into the electronic database by a delegated member of the research team

6.0 Data Validation

The aim of data validation is to generate a database/dataset that is of appropriate quality as decided as part of the risk assessment and defined in written procedures, data management plan and/or protocol. It is the process of checking the data for such elements as logical

consistency, protocol deviations and missing/incorrect/implausible data. To achieve this, formalised validation or edit checks must be put in place. These can be manual, electronic or a mixture of both. These processes identify data issues that can then be resolved by clarification with the investigator and/or vendors/holders of related databases (for example laboratory analysis vendors). When all data cleaning and data validation has been completed, the data should be formally declared 'clean' and the database locked for analysis.

7.0 Database/Dataset Lock

The aim of the data management process is to provide a high quality and appropriately clean final database/dataset suitable for statistical analysis. There should be a process for controlling this and it should be clear when the database/dataset is declared as final, what has been checked in order to make the decision, how the final database is made available, where the final database is stored and how it is accessed and protected.

A checklist to confirm what activities have been completed is recommended. Examples of activities that may be included on the list are confirmation that:

- All the data queries have been closed
- Pharmacovigilance reconciliation has been completed, i.e. all adverse events (AEs) and/or serious adverse events (SAEs) have been followed up and closed out
- All electronic data have been received, uploaded and validated
  - All clinical coding completed (where MedDRA or CTCAE have been used
  - All external data received and imported (where applicable)
- The statistical analysis plan has been approved.

When a clean dataset is required at time points other than the end of the study/trial (e.g., interim analysis or for data monitoring committee meetings), the same principles apply as to the final database. However, it might also be appropriate, dependent on the purpose of the analysis, to provide 'un-cleaned' databases/datasets in certain circumstances.

How the database is "physically" locked will depend on the individual study/trial but the outcome is that the data are protected from editing or deleting. Whatever the situation, the lock procedure must be appropriately robust to protect the final data and there should be documentation and an ability to demonstrate how and when the physical lock was done.

8.0 Unlocking the Final Database/Dataset

Occasionally it may be necessary to correct previously missed data errors or inconsistencies after the database has been locked and released for analysis. There should therefore be a process in place to resolve this by unlocking the database, correcting the data and providing new extracted datasets after the query has been resolved. However, unlocking a final locked database or dataset should be limited to important corrections (that is if the data to be changed will have a significant impact on the reliability of the results) and the final study/trial report should include details of all changes made to the database while it was unlocked.

9.0 Data Back Up and Disaster Recovery

It is essential that research data is backed up. UoL IT Policies (or those of the Sponsor's delegate (i.e., a third party) should be followed.

Further information on data management help can be found through the following links:

- https://uniofleicester.sharepoint.com/sites/staff/research-enterprise-support/staff-listing/SitePages/Research-Data-Management.aspx

- https://uniofleicester.sharepoint.com/sites/get-it-help/SitePages/where-store-files.aspx
- https://uniofleicester.sharepoint.com/sites/Research-Governance-Ethics-Integrity/Shared%20Documents/Forms/AllItems.aspx?id=%2Fsites%2FResearch%2DGovernance%2DEthics%2DIntegrity%2FShared%20Documents%2FREGI%20Office%2FNon%2DREGI%20documents%2FNHSDataMgmtGuidance%5Fv1%2D1%5F29062022%2Epdf&parent=%2Fsites%2FResearch%2DGovernance%2DEthics%2DIntegrity%2FShared%20Documents%2FREGI%20Office%2FNon%2DREGI%20documents

10.0 Responsibilities

Responsibility	Undertaken by	Activity
Sponsor and CI	Head of Research Governance or delegate and CI	Determine type of database required at Sponsor Risk Assessment/Sponsor Review.
CI and PI	PI or delegate	Ensure the accuracy, completeness, legibility and timeliness of data reported in the CRFs.
Sponsor, CI, PI or delegate(s)	Sponsor, CI, PI or delegate(s)	Complete quality controls throughout the data management process
Sponsor and CI	Head of Research Governance or their Delegate and CI	Ensure that any external providers have been identified as part of the Sponsor Risk Assessment Process and appropriate agreements in place before data capture begins.

11.0 Development and approval Record for this document

This table is used to track the development and approval of the document

Author	Job title	Reviewed by	Approved by	Date approved
Cat Taylor	Head of Research Governance	UoL Research Management and Operations Group (RSMOG)	Professor Nigel Brunskill 	19/01/2024

12.0 Review Record

This table is used to track any changes made on revised/reviewed versions:

Date	Issue Number	Reviewed By	Description Of Changes (If Any)
August 2015	2	Wendy Gamble	Reviewed and amended to include University of Loughborough on front page and amendments to version number and dates
Nov 2016	3	Diane Delahooke	Logo and RGO address change.
August 2019	3.1	Cat Taylor	RGO address change and update to ICH-GCP R2 Addendum.
Sept 2021	3.2	Cat Taylor	Administrative changes
January 2024	3.3	Cat Taylor	Administrative Changes Detailed data capture and entry details added New data management website links added Removal of distribution record