

LOUGHBOROUGH UNIVERSITY RESEARCH OFFICE STANDARD OPERATING PROCEDURE

Loughborough University (LU) Research Office SOP-1036 LU

Data Management Process for NHS Research Sponsored by Loughborough University

Effective Date: January 2016

1.0 Introduction

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

Data management processes are required to ensure that the data included in the clinical trial report or publication of the results are accurate and were captured in accordance with the approved protocol.

2.0 Scope

This SOP applies to all research studies sponsored by Loughborough University (LU).

3.0 Definition

The data management process covers the design and production of the data capture tool (see SOP -1039, 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 should be applied at each stage of data handling to ensure that all data are reliable and have been processed correctly.

4.0 Delegation

In accordance with GCP, data handling must be conducted by appropriately qualified and trained individuals. LU as Sponsor delegates the responsibility for data management to the Chief Investigator for each individual study. 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. When delegated to an external organisation the roles and responsibilities will be clearly detailed in an appropriate agreement.

5.0 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 trial-specific database and data entry system. This should also be validated to confirm that the database system is fit for purpose specifically for the trial.

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.

5.1 Choice of Data Management Database

The type of database used needs to be proportionate to the risk associated with the study and this will be determined during the Sponsor Risk Assessment process.

Whilst the type of database used 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.

For small simple trials MS Excel may be considered appropriate to be used for the data management and analysis rather than a formal clinical trial database. Whatever system is chosen, appropriate validation must be conducted and all transactions must have a clear and complete audit trail. For some software, e.g. MS Excel, this may necessitate the printing of data and the certification and dating of the data as an accurate record.

5.2 Electronic Trial Data Handling Systems

When using electronic trial data handling or remote electronic trial 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.

6.0 Data Entry/Capture

The investigator should ensure the accuracy, completeness, legibility and timeliness of the data reported in the case report forms (CRFs). Prior to data entry when using a paper CRF, the form should be reviewed for any missing data, incomplete fields or data outside normal ranges. If discrepancies are picked up at this stage they should be clarified with the chief investigator (CI) or delegated individual and any changes recorded, initialled and dated. The original data entry must not be obscured. In addition, any CRF data derived from source documents should be consistent with the source documents or the discrepancy explained. On completion of the above process, the data must be entered into the database by a delegated member of the research team. The data entry process should be defined for each specific study.

7.0 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.

8.0 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:

- All the data queries have been closed
- All electronic data have been received, uploaded and validated
- The statistical analysis plan has been approved.

For a more simple situation, for example involving use of a spreadsheet, a file note signed by the person responsible, together with the properties of the MS Excel file (for example, date/time stamped, file location), would be sufficient to confirm that the data are final.

When a clean dataset is required at time points other than the end of the trial, for example, at the time of an 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 very much on the individual 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.

9.0 Unlocking the final Database/Dataset

Occasionally it may be necessary to correct previously missed data errors or inconsistencies after the data have been 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 trial report should include details of all changes made to the database while it was unlocked.

10.0 Data Back Up and Disaster Recovery

It is essential that research data is backed up. Further information and help can be found through the following link: <http://www.lboro.ac.uk/services/it/>

11.0 Responsibilities

	Responsibility	Undertaken by	Activity
1	LU Research Governance Officer & CI	Research Governance Officer or delegate & CI	Determine type of database required at Sponsor Risk Assessment
2	CI	CI or delegate	Ensure the accuracy, completeness, legibility and timeliness of data reported in the CRFs.
3	CI, Research Team Data Management Responsible, Clinical Trial Monitor	CI, Research Team Data Management Responsible, Clinical Trial Monitor	QC throughout the data management process
4	UoL Research Governance Officer & CI	Research Governance Officer or their Delegate & 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.

This table is used to track the development and approval of the document and any changes made on revised / reviewed versions:

DEVELOPMENT AND APPROVAL RECORD FOR THIS DOCUMENT			
Author / Lead Officer:	Jackie Green		Job Title: Research Governance Officer
Approved by:	Ethics Committee		Date Approved: 5/2/16
REVIEW RECORD			
Date	Issue No.	Reviewed By	Description Of Changes (If Any)
DISTRIBUTION RECORD:			
Date	Name	Dept	Received