Portsmouth Hospitals NHS
NHS Trust

Standard Operating Procedure (SOP) for Data Management

For Completion by SOP Author	
Reference Number	Reference PHTRD/SOP/013
Version	V1.0 26 Apr 2016
Document Author(s)	Juan Campos-Perez, Carole Fogg
Document Reviewer(s)	Helen Munday, Christine Bevan, Laura Marshall

For Completion by Research Dept., SOP Controller	
Name of Responsible Committee	Research Governance Group 26 Apr 2016
Issue Date	23 May 2016
Implementation Date	23 May 2016
Review date	23 May 2017
Electronic location	G-drive-Policies and SOPS-Data Management-Issued

The definitive versions of all Portsmouth Hospitals Trust SOPs, Templates and Forms for Research are online at http://www.porthosp.nhs.uk/research-department

If you are reading this SOP in printed form then you are reading an uncontrolled document. You must therefore verify that the version number and date given below are the most recent, by cross-checking with the Trust research website before proceeding with implementation.

Portsmouth Hospitals NHS Trust is committed to ensuring that, as far as is reasonably practicable, the way we provide services to the public and the way we treat our staff reflects their individual needs and does not discriminate against individuals or groups on any grounds. This SOP has been assessed accordingly

CONTENTS

1. INTRODUCTION

Good data management practices are essential for clinical research and should be discussed fully during the planning stages of the study.

The data management process typically covers the design and production of the data capture tool (paper or electronic, commonly referred to as the Case Report Form or CRF), along with the design, construction, validation, release and subsequent amendments to the database to maintain the data electronically. It also includes the processing of data (entry/uploading, cleaning, quality control checks and query management) and the production of the final dataset ready for analysis. How study data should be managed and validated will vary depending on the design of each individual project. Therefore individual research protocols and/or study specific data management plans should be adhered to in this regard.

The International Conference on Harmonization Guidelines for Good Clinical Practice (ICH GCP) specify that appropriately qualified individuals should supervise the trial data handling, verify the data and conduct the statistical analysis (ICH GCP 5.5).

2. PURPOSE

The overall purpose of this Standard Operating Procedure (SOP) is to provide guidance for managing data and ensuring all data is collected, verified and analysed in the appropriate manner to preserve the scientific integrity of the research.

3. SCOPE

The information contained in this document should be used for all studies Sponsored by Portsmouth Hospitals NHS Trust (PHT). Clinical Trials of Investigational Medical Products (CTIMPs) must also adhere to the guidelines described as per Good Clinical Practice and the Research Governance Framework.

In the event of an infection outbreak, flu pandemic or major incident, the Trust recognises that it may not be possible to adhere to all aspects of this document. In such circumstances, staff should take advice from their manager and all possible action must be taken to maintain ongoing patient and staff safety

4. ABBREVIATIONS & DEFINITIONS

CRF: Case Report Form
CTIMP: Clinical Trial of an Investigational Medicinal Product
DM: Data Manager
DMP: Data Management Plan
eCRF: Electronic CRF
GCP: Good Clinical Practice
PHT: Portsmouth Hospitals NHS Trust
R&I: Research & Innovation
SOP: Standard Operating Procedure

5. DUTIES AND RESPONSIBILITIES

Chief Investigator: Oversight and knowledge of the data management process.

Trial/Study Coordinator: Ensure the data management process is followed; self monitoring; communicate with Data Manager

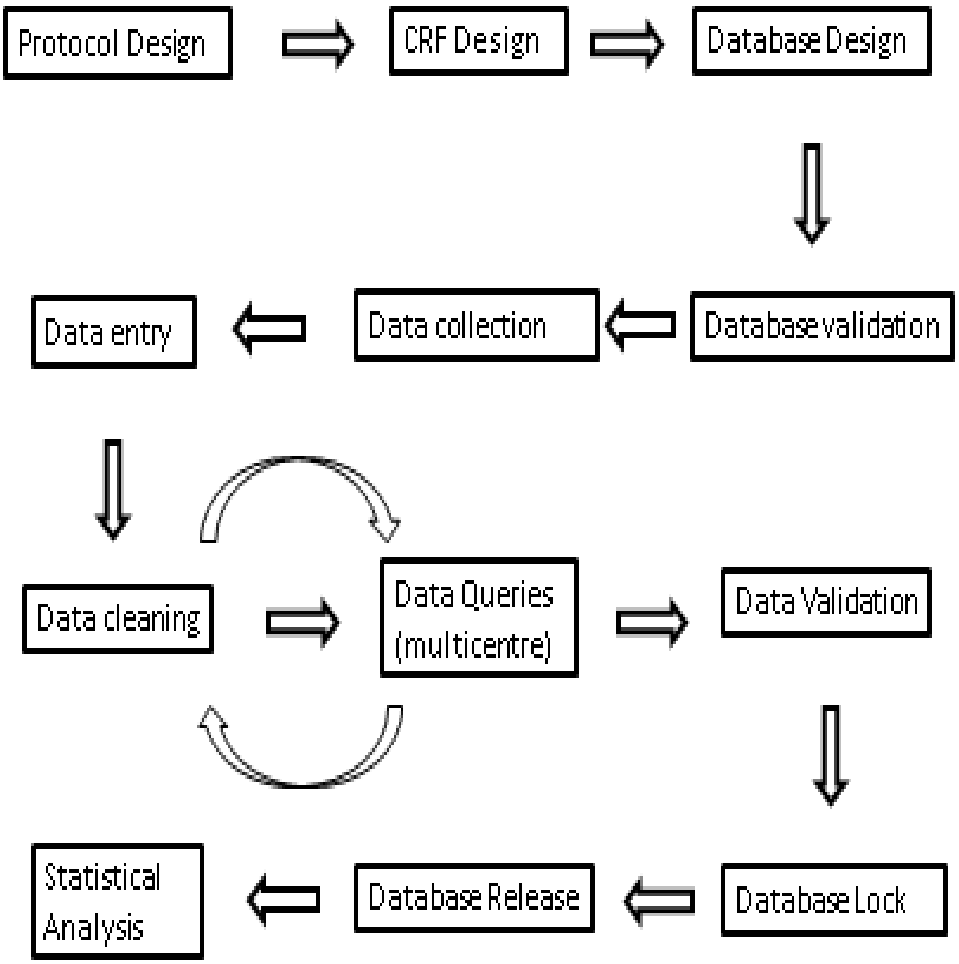
Data Manager: Responsible for implementing and oversight of data management process.

Sponsor: Monitoring that this SOP is followed as described; overall responsibility for implementing systems are in place to ensure data security and data quality.

Statistician: Involved in developing the data management process; assist the Data Manager to implement the process.

6. PROCESS

The diagram below outlines the main steps and requirements to be considered when designing a data management process for clinical research.



6.1 PROTOCOL DESIGN

Data management details should be outlined within the protocol for all Sponsored studies. For CTIMPS, multicentre studies and as decided by the Research Office (for example high recruitment numbers, multiple data sources, large data sets) an additional data management plan should be in place.

The protocol and/or data management plan should include the following information:

- Description of procedures for data collection (electronic or paper Case Report Forms (CRFs) consideration to given as to whether any files will be be posted, faxed, submitted on the internet or transferred electronically).
- A list of source documents (see section 6.1.1 below).
- Database design and validation.
- Description of which data (including safety) are collected and recorded in the CRF.
- Details of methodology implemented to ensure data validity (see section 6.4 below) including quality assurance for completion of CRFs and data entry, and integral database mechanisms (field checks, data cleaning and queries etc.), quality control checks of a sample of data on the database against the source data and at each stage of data transfer to separate file types.
- For Data Management Plan Adherence to the Data Protection Act 1998.
- Outline the duration and location of record/database retention.
- Description of how data will be stored and whether in electronic or paper form, how security would be ensured and whether data will be transferred
    - describe the system to prevent access to the database to unauthorised users
    - describe the data backup process and data lock;
    - post data lock changes for example from queries raised by the statistician;
    - describe the "audit trail" to be used in order to track any changes;
    - describe how you will keep a record of individuals authorised to make changes to the data).
    - database release.
- Plans for archiving of trial data (please see PHT/RDSOP/011 - Preparation and Procedure to Archive).

6.1.1 Source documentation

Source data is defined as the first place where data which will be used for the study is written. Complete and accurate source documentation is critical for all clinical research.

- The most common source documentation is the patient's medical or clinical record. If some parts of a protocol treatment are given at another hospital/clinic, make sure that copies of the relevant information be collected and maintained at the site that entered the patient on the trial.

- It is strongly recommended that any external data be requested on an on going basis to ensure that the participant's CRF is complete. Where possible, information about the fact that a patient is on a clinical trial and about the data required for the trial should be given prospectively to all health care professionals e.g. GPs, involved in the patients care.

- Other types of source materials include films such as X-rays and CT scans. Pathology slides may also be considered source materials for specific trials. A Case Report Form may also be a source data form if data is entered directly onto the CRF.

- A list of source documents should be recorded in the protocol to allow source data verification at the time of monitoring. Where possible this should indicate where each data point will be recorded for the first time.

6.2 CASE REPORT FORMS (CRF) DESIGN

To ensure that the trial data are recorded in a consistent way for all cases entered, CRFs need to be designed and tested prior to use. For a detailed description of the CRF design process please refer to the CRF Design Work Instruction (PHT/WI/038).

1. Review the CRF to check it has fields to collect all the data required by the protocol.

2. Once CRFs have been designed, it is recommended that you pilot the CRF. Where a research nurse will be involved in the study, it is recommended that they review the CRF. This can provide valuable feedback on potential problems prior to the activation of the trial.

3. When completing Case Report Forms for a clinical trial, there are quality controls that should be followed to comply with GCP:

- Eligibility checking
- Logging receipt of data
- Checking for correct identifiers
- Checking for data completeness
- Logical and consistency checks
- Manual or computerised checks
- Assessment of study endpoints

4. These quality controls should be checked during monitoring and consideration should be given to this when designing the CRF. It may be useful to create CRF completion guidelines, especially where there will be numerous personnel entering data on to the CRF.

5. Any changes to the CRF need to be version controlled. Completion of CRFs should be signed off by the Chief Investigator or delegated individual.

6.3 DATABASE DESIGN

1. Always ensure that the database used for the trial is designed in parallel with the definition of data items to be collected (as per the protocol and the CRF) so that the data captured are complete, accurate, reliable and consistent.

2. A commonly used database structure will have records that mirror the CRF. For each type of form used in the trial, there will be a database record. With this structure it is essential that all records for one patient can be linked together. The unique patient identifier for the trial should be on all records so that this linkage is possible.

3. If the database used is large, with many different sources of data and many different record types, then the database should be set up and designed by an experienced database analyst/manager or delegated to a Clinical Trials Unit (CTU).

4. The Database specification should be written after the CRF has been designed and should describe variable types, limits, ranges and inter-variable consistency checks. For example, the

value ranges check (the data item in a field is within an expected range of values); the field type check (verifies that the data entered in a field are of the correct type) and logical checks (ensure that the data "makes sense").

5. Carefully review the software features and to select the product that will most closely meet the requirements of the project. Some features that may be important in selecting a data management system are shown in appendix I.

6. In evaluating possible systems, first decide which of these features are important to the project, and then assess which ones are available in each of the systems reviewed.

7. In making a software selection, it needs to be decided whether any of the computing will be done at the sites, or whether all will be done centrally. If a centralised system is to be used, consider electronic access, whether electronic CRFs will be used.

6.4 DATABASE VALIDATION

Data validation is the process of checking the data for elements such as logical consistency, protocol deviations and missing, incorrect or implausible data. This is achieved by setting up formalised validation or edit checks on the data. The process can be manual, electronic or a mixture of both. The aim of 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 such as the Data Management Plan (DMP) or protocol.

1. Prepare a plan for how the data will be validated. The plan may cover roles and responsibilities, the types of check and how they will be chosen and documented, the processes used for implementing the checks and how any problems will be resolved.

2. Define, review and agree the description of checks to be performed by the data manager or assigned CTU. It is recommended that all the proposed checks are listed in the specification together with the methodology by which they will be undertaken. These include checks that are:

- Performed manually by reviewing print outs with possible reference back to actual CRFs
- Programmed into a eCRF or data entry system (on entry)
- Done on a batch or continual validation basis

6.5 DATA COLLECTION

Before starting data collection, in an ideal situation the following procedures should be carried out in order to minimise the risks of mistakes or breaching privacy guidelines:

1. After set up, test or pilot the system before you use it and maintain an adequate record of this procedure. It would be a good idea to write a Study Specific Procedure or a working practice document detailing how you set up your electronic data capture systems. The appropriate persons need to be trained. It would also be helpful to write a Privacy Risk Assessment.

2. After testing, when the data collection process is working as it should, train all users of the system. A record should be kept of the training and a detailed diagram and description of how data will be collected should be provided at training.

3. Ensure the validity of data. This can be done by auditing the system during data collection. This is necessary to make sure the source data is identified and data transcribed correctly onto data collection system.

4. For electronic data collection systems, other issues to consider include whether an electronic CRF will be used, or whether data will be entered straight onto a website. When designing forms to collect data electronically you should consider the use of 'validation rules'.

5. Do not forget to conduct regular backups of your data, if outsourcing data collection or storage ensure that the company have backup systems in place.

6. When ready to archive your data, include both hard copies and electronic data. Documents not archived need to be disposed of securely.

6.6 DATA ENTRY

For any clinical trial the transfer of data from the source data to the paper case report forms and then to electronic format are critical steps, and accuracy of data entry is essential. It is therefore extremely important that the data entry system be set up with adequate quality control checks.

1. Choose a quality control check for data entry. There are several ways of doing this, and the method chosen will depend on the training of the person doing the data entry, the software that is being used, and the programming support available and should be done according to a study specific data entry SOP:

- ***Double data entry***: In many environments, data are entered twice to ensure a high degree of accuracy. This technique may only be done in a proportion of forms, then additional forms checked according to error rate.

- ***Single data entry***: The alternative to double data entry is to enter data only once and then to introduce some supplementary quality control checks. The secondary checks can be by visual review of the data forms against the data entered or by developing computer checks of value ranges, field data types, and logical relationships between data items.

6.7 DATA CLEANING

Data cleaning is the process of identifying errors / inconsistencies / missing data spotted at different time points depending on the study and methods used. Errors should be corrected where possible, but no changes should be made without proper justification. Appropriate audit trails should be kept to document changes in the data.

The process of data cleaning is as follows:

1. The DM cleans and validates data entered into the database. If no problems are found, then the dataset can be validated.

2. If issues are identified such as missing values or inconsistencies then a data query is raised. If the study is multicentre the query is addressed to the site where the data originates from.

3. The site solves the issue and sends back the query. The query is checked and accepted by the DM who issues a queries resolution.

4. Corrections are entered onto the database. Once all data queries are resolved and the DM is satisfied the dataset is clean the DM proceeds to lock the data.

6.8 DATABASE LOCK

How the database is physically locked is hugely dependent upon how large the data management function is, the system being used and whether it is a one or two step process.

1. Decide upon the process to be used: whether the database is taken away from the manager and only a limited number of people can add, modify or delete data for the rest of the process (soft lock) or whether the right to make changes to the database is removed form all personnel and no changes should be made after this point (hard lock).

2. If the database has an in-built functionality allowing them to be locked in terms of access rights which can then be controlled by a senior data manager, head of department or IT department, it is recommended that staff controlling the lock are independent from the day to day data manager.

3. The decision on the level of security needs to be made on a risk-based approach, and the lock procedure must be appropriately robust to protect the final data with documentation available to show how and when the lock was done.

4. If it is necessary to correct previously missed errors or inconsistencies after the data has been released for analysis, there should be a process in place to unlock the database, correct the data and provide new extracted datasets after the query has been resolved.

5. Repeated locking and unlocking will be viewed with concern by inspectors because this will have a serious impact on the credibility of the trial. The justification for requesting the unlock, the written approval and the effect on the statistical outcome should be documented and must be filed in the TMF prior to unlocking. When re-locked, the new final database should not overwrite any analysis datasets that were created at the original database lock.

6.9 DATA RELEASE

Data release happens after the database is locked and ready to be provided for statistical analysis. The Data Manager (DM) should confirm that the data is ready to be released for analysis once:

- All data queries have been resolved and the database updated.
- Any issues identified from Quality Control (QC checks) have been addressed.
- The data has passed an error-rate audit, if applicable.

1. Where different datasets are to be provided to different staff members (e.g. trial statistician, health economist), a dataset specification detailing which variables or data forms are required for analysis should be prepared for the data manager.

2. The DM should document the release of the data, usually via email to the statistician and other members of the study team. A copy of the data release documentation (email) should be filed in the Study Master File (SMF).

3. If a data lock facility is present in the database then the DM should lock the database. If this facility does not exist then the DM should ensure that data cannot be altered by restricting access to the data by securing its location or setting up a password.

4. For trials where the data has been released for the purposes of an interim analysis it is acceptable for the database to be unlocked to continue data entry for the rest of the trial. A copy of the datasets used to conduct the interim analysis must be maintained.

6.10 STATISTICAL ANALYSIS

1. Make sure there is a clear boundary between data management and statistical analysis, with a final dataset locked and subsequently released even if the statistician also doubles as data manager. This is particularly important in blinded trials to avoid accusations of bias.

2. If the data management system does not come with software to interface with statistical software, one needs to be written so that the data can be retrieved from the database and put into statistical software.

3. Make sure the program is thoroughly and rigorously tested to ensure that accurate data values are inputted into the statistical package. A sample of the data should be checked at each transfer of data to ensure it has not been corrupted.

4. Statistical analysis should be carried out according to the current version of the Statistical Analysis Plan.

5. Once the statistical analysis is completed, the statistician should prepare a report which is then handed over to the research team for publication.

7 DATA MANAGEMENT FOR SMALL, LOW RISK STUDIES

For small, non-commercial, low-risk studies, the above requirements might be excessively time consuming. A pragmatic approach therefore needs to be used in a case by case basis. In such cases the following guidelines might be suitable:

1. MS Excel may be considered appropriate to be used for the data management and analysis, as long as it is used by someone with expertise and data validation tools are utilised.

2. The data validation process may be fairly minimal and conducted manually (checks of print-outs and CRFs). Where any checks on the data's validity are being done, however, these must be documented and retained, so it is evident from the TMF what was checked, when and by whom.

3. For smaller databases, for example an MS Excel spread sheet file, the locking process may mean resetting the protection rights of the file or the folder containing the files by someone independent from the operational management, such as the IT department, or study Sponsor. Usually, system-level access properties need to be changed rather than password-protecting the file although this level of protection may be justified in small, open trials managed by an investigator.

8 TRAINING REQUIREMENTS

- The Research Dept. will endeavour to notify individuals of SOP developments that may be relevant to them. Updates on SOPs will feature in research newsletters and communications. It is the responsibility of all research active staff to ensure that they read the issued updates that may be relevant to them.

- When a new SOP is authorised, or when an existing SOP is revised, self directed training must be carried out by all staff to which the SOP is relevant and this training documented in their training record. A template is provided to support this process. A

study specific SOP training plan will be developed for investigators on high risk PHT Sponsored studies.

- Staff should take time to read and fully understand the SOP and relevant documents, ensuring that they are able to implement the SOP when required. If clarification is needed then the trainee should approach their line manager and the SOP Controller who will arrange additional training. All staff should complete their training prior to the published implementation date which will normally be between 2-6 weeks after publication.

- All staff are responsible for maintaining their own SOP training Records and copies must be made available to line managers, the SOP Controller or study monitors and sponsors on request.

9 REFERENCES AND ASSOCIATED DOCUMENTATION

MHRA Good Clinical practice Guide 2012.

McFadden, Eleanor. 2007. Management of data in clinical trials – Second edition. Wiley Inter-Science.

North Bristol NHS Trust R&I. Research Management SOP: managing and validating research data.

Royal free Hampstead NHS Trust. Guide on good practice for data management for chief Investigators of research sponsored by UCL/Royal Free.

Hulley et al. 2007. Designing clinical research. Wolters Kluwer.

10 VERSION HISTORY LOG

This area should detail the version history for this document. It should detail the key elements of the changes to the versions.

Version	Date Implemented	Details of Significant Changes

11 APPENDICES

Appendix I: Features to consider when selecting a data management system:

Data management system features	Required for project	Available in system under review
Support for direct data entry and, when applicable, data verification		
Ability to do ad hoc queries against the database without extensive programming skills		
Built-in audit trial capability		
Utilities available for data back up and recovery		
Interface with statistical analysis software		
Appropriate representation of missing values		
Support for a variety of data collection instruments and methods		
Training and technical support		
Programming and report writing tools to simplify usage		
Cost effective		
Efficient storage and retrieval of the volume of data required for the trial		
Widely used by other sites		
Prevention of duplicate entries		
User-friendly data entry interface		
Controlled access		
Capacity for data lock		

Portsmouth Hospitals NHS
NHS Trust

CONFIRMATION OF SOP TRAINING RECORD

A copy of this record may be kept in your personal training file to confirm your training in a specific SOP. The research department or your line manager may request copies to verify your training. If required by a study Sponsor a record may also need to be kept in the Trial Master Files (TMF) or Investigator Site Files (ISF).

SOP Details: To be completed by the SOP Controller	
Title of SOP	Data Management
Reference Number	Insert SOP Reference Number. PHT/RDSOP/013
Version	V1.0 26 Apr 2016
Issue Date	23 May 2016
Implementation Date	23 May 2016

Personnel Details	
Name	
Job Title & Research Role	
Date of Training	
Nature of Training	Self Directed/Delivered by etc
Records of any meetings to clarify details in SOP	

Signatures
I confirm that I have read and consider myself to be sufficiently trained in the above Standard Operating Procedure with regards to my individual roles and responsibilities Signature of Trainee …… Date …….
I confirm training in the above SOP was delivered as recorded above and that the trainee may be considered sufficiently trained in their roles and responsibilities Signature of Trainer …… Date ……. .

Additional Notes & Signatures
Signature of Trainer (where appropriate) I confirm training in the above SOP was delivered as recorded above and that the trainee may be considered sufficiently trained in their roles and responsibilities Signature of Trainer …… Date ……