Portsmouth Hospitals University NHS Trust

Standard Operating Procedure (SOP)

Data Management of Research Studies at PHU

For Completion by SOP Author	
Reference Number	PHTRD/SOP/013
Version	V2.0, 4th November 2020
Document Author(s)	Milan Chauhan, Research Data Coordinator
Document Reviewer(s)	Paul Meredith, Information Analyst

For Completion by Research Dept., SOP Controller	
Name of Responsible Committee	Research Delivery Meeting
Issue Date	4th November 2020
Implementation Date	4th December 2020
Review date	4th November 2023
Electronic location	G:\Research and Development - Research Office\Policies and SOPS\# Active SOPS\SOP 013 - Data Management

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

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 University 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

Data management practices are required to ensure that the outcomes of the trial specified in the clinical trial report or publication are accurate and were captured in accordance with the approved trial protocol.

The data management process involves the design and production of the data collection tool (CRF – Case Report Form, electronic or paper), along with the design, construction, validation and 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. Quality control must be implemented at each stage of the data management process to ensure that all data are reliable and have been processed correctly. 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.

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 of all studies sponsored by Portsmouth Hospitals University NHS Trust.

3. SCOPE

The information contained in this document should be used for all studies Sponsored by Portsmouth Hospitals University NHS Trust (PHU). 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 AND DEFINITIONS

Abbreviation	Meaning
CRF	Case Report Form
CTIMP	Clinical Trial of an Investigational Medicinal Product
DM	Data Manager
DMP	Data Management Plan
CTDMS	Clinical Trial Data Management System
eCRF	Electronic Case Report Form
GCP	Good Clinical Practice
PHU	Portsmouth Hospitals University NHS Trust
SDV	Source Data Verification
SOP	Standard Operating Procedure
TMF	Trial Master File

5. DUTIES AND RESPONSIBILITIES

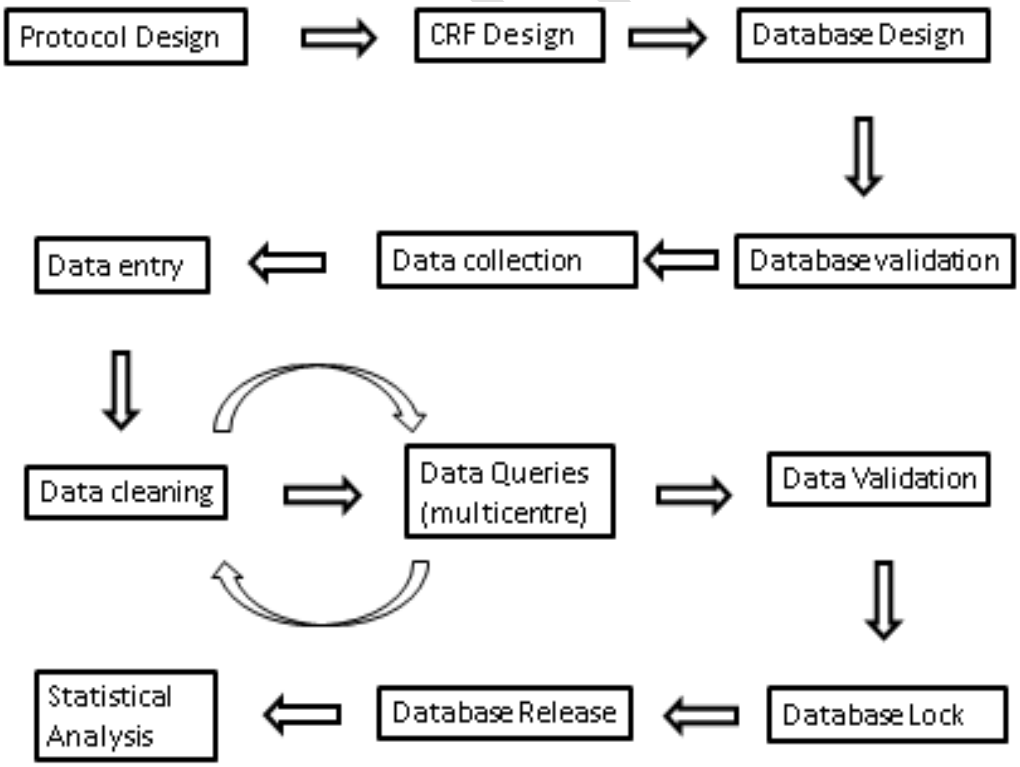
In accordance with GCP Guidelines, data handling, verification and conduct of statistical analysis must be conducted by appropriately qualified and trained individuals. The Chief Investigator of a PHU sponsored study is

responsible for the data generated from the study. Where this is further delegated to another member of the research team or an external organization, this must be documented in the Delegation Log.

Role	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 (DMP) 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 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 2018.
- 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.

- Source documents are considered "Essential Documents" that allow evaluation of the study and ensure quality of the data and serve to certify Sponsor and CI compliance with the relevant regulatory requirements.
- 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 ongoing 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 and DMP 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.
- The process of Source Data Verification (SDV) is an evaluation of the data recorded in the data collection tools (paper or electronic CRFs) against the source documents.

6.1.2 MONITORING PLAN

The Monitoring Plan is a document that describes the strategy, methods, responsibilities, and requirements for monitoring the trial. Source Data Verification for all required data should be carried out during monitoring visits in accordance with the trial monitoring plan. Trial personnel on site can also perform SDV during self-monitoring. Central data monitoring should be detailed in the DMP.

6.1.3 DATA MANAGEMENT PLAN

Trial-specific DMPs should be developed for every Sponsored trial. This document should serve as a Service Level Agreement between the Data Manager (DM) and the Investigator to define all data management activities for the trial. It should be in place after the trial protocol has been finalized and prior to recruitment. The extent of data management activities described in the DMP will be dependent on the complexity of the trial and associated risks.

The Research Office should maintain and version control a template document for the DMP. The DM or a study team member with appropriate experience and qualifications should complete the DMP template for the specific study. An exception can be made for studies led by students who may be required to complete the DMP themselves as part of their project – on such occasions, the Research Office should provide the latest version of the DMP template to the student and the DM should provide guidance and review it before the final version of the DMP is implemented. The DMP should be reviewed and signed off by the R&D Manager on behalf of the Sponsor, Trial Manager, and Trial Investigator. Amended versions of the DMP should be reviewed and stored in the TMF. The DMP should include:

- All data collection tools to be used
- The data management system to be used – description of systems and validation, documentation
- Description of device, software and data if applicable
- Central data monitoring plan including:
    - Activity to be checked by DM
    - Estimated frequency of check
    - Who will perform the activity
    - Output
- Medical coding dictionaries if applicable
- Data verification requirements (Source Data Verification, Medical Review, Data Review)
- The process for opening and resolving queries
- Anticipated database locks
- Datasets to be generated including export formats
- Data backing up, disaster recovery plans and archiving

6.2 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). Using a template CRF for all studies to provide consistency is considered good practice. If REDCap is used, paper CRFs can be exported as PDFs and printed to be used as paper CRFs however paper CRFs produced by the DM's templates are preferred.

1. Review the CRF to ensure that it collects all of the data required by the protocol, especially the primary and secondary endpoints specified in the DMP. Data that isn't specified in the protocol and will not

affect the analysis or the conduct of the trial should be considered irrelevant and should not be included on the CRF.

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. Using a CRF template with consistent formatting can provide familiarity to the research team and aid the acceptance of the CRF.
3. When completing Case Report Forms for a clinical trial, there are quality controls that should be followed to comply with GCP:
    a. Eligibility checking
    b. Logging receipt of data
    c. Checking for correct identifiers
    d. Checking for data completeness
    e. Logical and consistency checks
    f. Manual or computerised checks
    g. 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. Each completed version of the CRF should be approved and signed off by the DM, Statistician, Trial Methodologist, CI, and the Trial Manager. A copy of each version of the CRF used during the study should be stored in the TMF.

6.2.1 PARTICIPANT IDENTIFICATION

The data entered into the study database must be de-identified, that is all direct identifies are to be removed and replaced with a study specific participant identifier. (In the USA there is a legal definition of protected health information, PHI, which includes personal identifiers so you may see references to PHI in software with an American origin such as REDCap.) Email addresses can be recorded only for the purpose of sending surveys using a built-in survey module. The field which the email address is collected should be marked as an identifier to reduce unauthorized access to the information.

It is common practice to record initials to reduce the risk of entering data against the wrong participant. This or the use of some other non-specific data should be risk assessed, and where used it should be removed from the study data set when producing the data set for analysis.

6.3 DATABASE DESIGN

The type of database used during the trial may vary depending on the type of trial e.g. simple qualitative studies may not require a system with the same level of sophistication as highly regulated CTIMPs. Where possible, the PHU instance of REDCap should be used.

While the complexity and sophistication of the database will increase proportionately with the risk of the study, the sponsor must ensure that the principles of ICH GCP are adhered to:

"A standard for the design, conduct, performance, monitoring, auditing, recording, analyses, and reporting of clinical trials that provides assurance that the data and reported results are credible, and accurate, and that the rights, integrity, and confidentiality of trial subjects are protected"

For simple, low risk trials, a Microsoft Access database (minimum) may be considered as appropriate if a specialised clinical data management system such as REDCap (recommended) is not available. In order to comply with the GCP guidelines, an audit trail that is attributable to the users must be implemented.

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 metadata (data about the data to be collected) should be stored and versioned after each revision of the database design. The metadata should be generated using the same columns as a REDCap Data Dictionary e.g. field name, field type (checkbox, dropdown), validation (minimum and maximum values accepted), whether any branching logic is applied (show field only if condition is met) etc. Please see Appendix for sample metadata.

When a commercially produced and validated electronic data capture system such as REDCap is to be used (recommended), the following requirements from the Sponsor must be fulfilled:

- There must be documented evidence of the validation of the system.
- SOPs should be in place for using the system including training materials and training records
- An audit trail is in place that maintains records of changes to the trial data and can attribute those changes to each database user.
- Data integrity should be ensured by a data query resolution system to trace and attribute data corrections.
- Security system must prevent unauthorised access to the trial data. Two-factor authentication should be implemented if available.
- Appropriate software updates relating to the database security should be actioned and documented while minimizing downtime.
- An audit of the users with access to the trial data must be maintained. This includes their role in the trial, their trial-specific permissions, account creation and expiration dates, access to modules (API tokens, logging, exports etc).
- The trial data must be adequately backed up. The procedure and frequency of backing up the database should be detailed in the DMP. The system should also allow exporting of the study structure and data into XML format to allow snapshots of the trial to be backed up locally by the DM.
- If the trial requires blinding, it must be maintained during data entry and in processing.
- If data are transformed during processing it should always be possible to compare the original data and observations with the processed data.
- An unambiguous subject identification code should be used to prevent any patient identifiable data being uploaded.

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 DMP or protocol. Quality control should be applied at each stage

of the data handling process to ensure the reliability of the data. If data are transformed during processing, it should always be possible compare to the original data.

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

6.4.1 USER PERMISSIONS AND TRAINING

Study members documented on the Delegation Log as requiring access to the CTDMS should be given user-specific password-protected accounts. The DM should then configure their permissions based on the users role within the study (limit data export access, restrict identifier fields). For Data Entry roles, the users should only be able to enter and edit data on relevant forms e.g. If the user is based in the laboratory and only inputs results from laboratory tests, they should only be given permissions to the data collection form containing laboratory results. The user's permissions should be in accordance with the Delegation Log and their role within the study i.e if they can access identifier fields (Section **Error! Reference source not found.Error! Reference source not found.**)

A trial-specific database training checklist should be generated and each prospective user on the Delegation Log should be trained by the DM or a qualified delegate. When the Sponsors instance of REDCap is used, each new user must be added to the User Authorisation project and the results of their training should be uploaded. When using REDCap, the username should be the user's Windows alias to avoid confusion of user accounts.

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. Each version of the database should undergo User Acceptance Training which should be documented in the TMF.
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 directly into the CTDMS. 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

The trial data should be accurate, complete, contemporaneous, attributable, and original and the CI is responsible for ensuring these principles are adhered to when data is reported on the CRFs. When a paper CRF is used, prior to data entry the CRF must be reviewed for any missing data, incomplete fields or data that could be invalid. When these discrepancies on the CRF are noticed, the CI must be notified and any changes recorded, initialled and dated on the CRF without obscuring the original data entry.

Completed CRFs should be scanned and stored securely. This not required but is considered good practice.

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.

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, and then additional forms checked according to error rate. Double data entry is not required but should be considered good practice.
- ***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. This is more applicable to smaller, single centre studies without the staffing required for double data entry. After data entry, a visual check can be performed between the recorded paper CRF and the database record.

6.7 DATA CLEANING AND VALIDATION

The aim of data cleaning is to generate a dataset that is of appropriate quality as decided as part of the risk assessment and defined in the DMP and protocol. The process aims to validate the data by checking for logical inconsistency, protocol deviations, and missing/invalid data entry using both manual and electronic methods at different time points during the trial. Errors should be corrected where possible, but no changes to the data should be made without proper justification. When an error/inconsistency in the dataset is identified, a query should be raised. The query should highlight details of the discrepancy so it can be resolved by clarification with an assigned user(s)/user group(s). The process of opening, assigning and resolving of a query should be documented in the trial audit trail. Rules should be programmed into the database to identify records with logical inconsistencies (e.g. [pregnant] = “Yes” and [gender] = “Male”).
Data cleaning should take place throughout the study period with external data validation checks performed on a periodic basis to facilitate central monitoring and to provide reports for Data Monitoring Committee and Trial Steering Committee meetings. Data entry should be reviewed for potential discrepancies in primary and secondary outcomes across all sites.

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. Discrepancies within the database can be identified by executing rules to identify

logical inconsistencies or outlier data points. These rules can be executed at any time point, including during data entry.

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 resolves the issue and responds to the query. The query is checked and accepted by the DM who issues a queries resolution.
4. The corrections are entered onto the database and users with the requisite permissions can lock each complete record. Once all data queries are resolved and the DM is satisfied the dataset is clean, the DM proceeds to locking the database.

With amendments made to data recorded on paper CRFs, the corrections should be made directly by site personnel by drawing a single line through the incorrect item and dating and initialling the correction in pen.

6.8 DATABASE LOCK

The data management process exists to provide a high quality, valid, and appropriately clean dataset that is suitable for statistical analysis. The database locking process is in place to declare the dataset as final and to document what has been checked in order to arrive at the decision, how the final database is made available, where the final database is stored, and how it is accessed ad protected. The process of locking or inactivating the database must be documented and stored in the TMF.

The trial database can be locked in two stages:

- **Hard lock** The data has been cleaned, validated, and is as complete as possible. The dataset is ready for analysis and no further data amendments are anticipated. User edit permissions are revoked and the database is archived and taken offline.
- **Soft lock** The same principles as a hard lock but it is expected that further activity may be undertaken which could make changes to the data. User permissions may or may not be retained. Usually precedes a hard lock.

The protocol may state that an interim analysis is required. When the dataset is deemed to be complete enough for interim analysis by the DM, the database is inactivated/frozen to prevent new data from being added. This is equivalent to a soft lock.

The process of locking/inactivating should be documented with a checklist of completed activities including:

- All data queries have been resolved and closed
- Pharmacovigilance reconciliation has been completed
- Adverse Events have been resolved
- The Statistical Analysis Plan has been approved

1. Decide upon the process to be used: whether access to the database is revoked by the manager with only a limited number of people able to 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 from all personnel and no changes should be made after this point (hard lock).

2. 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.

3. 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.

4. 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 DATABASE UNLOCKING

In order to correct previously missed discrepancies within a locked dataset, the database must be unlocked by the DM. The queries must be opened, resolved and documented within the trial audit trail. Unlocking a database should be restricted to only critically important corrections unless the data to be changed will have a significant impact on the reliability of the results. The final trial report must be amended to document the unlocking of the database.

6.10 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 Trial Master File (TMF).
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.

6.10.1 DATA TRANSFER

6.10.1.1 PAPER CRFs

Trial Management Personnel are responsible for ensuring that paper CRFs are stored securely and are only accessible by authorized personnel who are documented in the Delegation Log. If paper CRFs are transferred for data entry, then they should be sent via courier or registered post to ensure safe delivery. Logs should be maintained to track documents sent and received for each site. After the database has been locked, it is recommended that a scanned copy of the CRF should be stored securely. Casebooks in pdf format for each participant should be downloaded and stored securely.

6.10.1.1 ELECTRONIC DATA TRANSFER

Electronic data should be risk assessed before it is transferred, with the assessment being documented in the DMP.

If an API token is requested by a member of the study team, then the generation and issuing of the token should be documented as well as the member's permissions. API import record calls should be restricted with the exception of mobile/tablet applications configured by the DM.

Study members permissions should be configured by the DM to correspond with the Delegation Log and could be further restricted for data transfer if required by removing free-form text and tagged identifier fields.

If a dataset is to be sent to a member of the study team, a secure-data transfer application should be used, such as the Send-It module built into REDCap where a zip file of up to 32MB in size can be encrypted and stored on the web server and downloaded within a specified time period with a password. This is considered good practice but if it is not available then the data should be sent by NHS email.

Data that is too large to be sent by attachment or secure data transfer applications should be transferred by removable media. The removable media should be encrypted and password protected.

6.11 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 on a case by case basis. In such cases the following guidelines might be suitable:

1. MS Access may be considered appropriate to be used for the data management, 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.

8. TRAINING REQUIREMENTS

- Members of the research department involved in handling data should be trained in this SOP.

"The Research Dept., will endeavour to notify staff of SOP developments that may be relevant to them. SOPs are available on the Research department website. Updates on SOPs will feature in Research

newsletters and communications and disseminate at local research meetings. 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 PHU 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 on request."

9. REFERENCES AND ASSOCIATION DOCUMENTATION

Associated Documents

- CRF Design Work Instruction (PHT/WI/038)

Reference

- 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.
- University Hospitals of Leicester NHS Trust. 2014. Standard Operating Procedure for Data Management Process for Research sponsored by University Hospitals of Leicester NHS Trust
- Hulley et al. 2007. Designing Clinical Research. Wolters Kluwer.

10. VERSION HISTORY LOG

Version	Date Implemented	Details of Significant Changes
1.0	26th April 2016	New document
2.0	4th December 2020	Sections on Monitoring Plan, Data Management Plan, Protected Health Information, and Data Transfer inserted Database Design and User Permissions sections re-written to incorporate REDCap/commercial EDC software Data Cleaning and Database Lock sections updated

11.APPENDICIES

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	PHT/RDSOP/013
Version	2.0
Issue Date	4th November 2020
Implementation Date	4th December 2020

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 …….