Data Management

NJRO-REG-SOP-012

Contents

1. Background/Introduction

All clinical research must be carried out in accordance with International Committee on Harmonisation (ICH) Good Clinical Practice (GCP). GCP is an international ethical and scientific quality standard for the design, conduct, performance, monitoring, auditing, recording, analyses and reporting of trials that involve the participation of human subjects. Compliance with GCP provides assurance that the data and reported results are credible and accurate and that the rights, wellbeing and safety of participants are protected. There are a number of principles that specifically relate to data:

ICH GCP Principle 2.10:
"All clinical trial information should be recorded, handled, and stored in a way that allows accurate recording, interpretation and verification."

ICH GCP Principle 2.11:

"The confidentiality of records that could identify subjects should be protected, respecting the privacy and confidentiality rules in accordance with the applicable regulatory requirement(s)."[1]

Data management encompasses all aspects of data entry, data processing and data validation to ensure that quality control is applied to all stages of data handling. It is essential that all stages of the process are documented to ensure compliance with Regulation 31A (4) (*Trial Master File and Archiving*) of SI 2004/1031.
Within The Newcastle upon Tyne Hospitals NHS Foundation Trust (NUTH FT) the Research & Development (R&D) Team are responsible for ensuring that a clinical trial complies with the legislation and GCP.

2. Purpose

The purpose of this Standard Operating Procedure (SOP) is to describe which data management activities are required for clinical research studies where NUTH FT is research sponsor. It outlines what documentation should be in place and stored within the Trial Master File and it specifies the procedures to be carried out by R&D to ensure sufficient sponsor oversight.
This SOP does not cover any activities that are expected to be carried out by the trial statistician such as randomisation, blinding and analysis.

[1] For the UK, the applicable regulatory requirement is the Data Protection Act, 1998.

3. Scope of Document

This SOP is applicable to all personnel carrying out Clinical Trials of an Investigational Medicinal Product (CTIMPs) within NUTH FT.

For other forms of clinical research, personnel should consider following this procedure as best practice.

Members of the NUTH FT R&D Team with responsibility for performing sponsor activities on behalf of NUTH FT must also abide by this SOP.

4. Procedure

4.1 Data Management Activities

Data Management is not specifically covered in the legislation however the sponsor must adhere to GCP Principle 2.10. The data management process typically covers the determination, design and production of the data collection tool, the processing of the data and the production of the final dataset(s) ready for analysis. Key activities of data management and the principles that should be applied are explained below:

4.1.1 Data Collection Tools

All data to be collected will be detailed in the study protocol. How the data will be collected will need to be determined prior to the conduct of the study. The data collection tool or Case Report Form (CRF) to be used should be determined based on the requirements of the protocol. For small, simple studies this may be a paper CRF (pCRF) and Microsoft Excel spreadsheet and for large, complex studies this may be an electronic CRF (eCRF). It may also be necessary to capture data from study participants directly, for example by a study diary.

4.1.1.1 pCRFs

A pCRF is a paper document of one or more pages, often with at least one 'no carbon required' NCR copy included, used to collect data about trial subjects. Where pCRFs are used, it may also be necessary to have a clinical database to aid with analysis. Where this is the case, it is necessary to ensure that any data transcribed into the database is validated as well as the database itself (see sections 4.1.3 and 4.1.5)

4.1.1.2 eCRFs

An eCRF is an electronic data collection tool that allows investigator sites to enter trial subjects' data directly into the system via a web-portal.

Studies managed by the Newcastle Clinical Trials Unit (NCTU) should use the unit's preferred eCRF.

For studies not managed through NCTU, where an eCRF is the preferred data collection tool, contact R&D for guidance regarding vendor selection.

4.1.2 CRF Design

Once an appropriate data collection tool has been selected it is important to design the CRF appropriately. Only collect data that is specified in the protocol. Do not collect data that will not be analysed. Consider the following when designing either a pCRF or eCRF:

- State the study name or reference, site name or reference and patient ID on every CRF page.
- For pCRFs, use page numbers so that pages cannot be missed.
- Use a consistent approach throughout, e.g.:
  - When collecting dates, use the same date format throughout
  - Place checkboxes after the question, or before but not both
- Use indicator questions to avoid having to make assumptions, e.g.:
  - Did the patient experience any AEs? Yes ☐ No ☐
  - NOT: AEs? None ☐
- Use Yes/No answers instead of single checkboxes, e.g.:
  - Is the patient continuing? Yes ☐ No ☐ … NOT Continuing? ☐
- Avoid free-text answers where possible as they're very difficult to analyse. Provide suggested answers e.g.:
  - Use drop down lists
- Don't collect data that can be derived, e.g.:
  - Collect date of birth, not age
  - Collect weight and height, not BMI
- Provide clear and specific instructions, e.g.:
  - State if/when fields should be completed: If other, then specify ……
  - Specifically state: check ONE box only
  - Provide guidance on the answer format: Date _ _ / _ _ _ / _ _ _ _
- Avoid collecting the same piece of data in two places otherwise checks will need to be carried out to ensure consistency

4.1.3 Database Validation

Database validation is the documented process of checking the functionality of the database to ensure that it performs as expected. It is important to ensure that the chosen system is appropriately validated before conducting the study. Off-the-shelf databases should be validated by the developer with documented proof of validation. The study team will carry out User Acceptance Testing (UAT), on the database to ensure that it is fit for purpose.

Consider the following when carrying out UAT:

- Ensure the database accepts data you expect to be entered
- Ensure that data validations (see section 4.1.5) work correctly and that the database rejects data you don't want to be entered
- Ensure that any user accounts are set up correctly and can only access the areas of the database assigned to that account
- Ensure that data can be extracted for analysis appropriately
- Ensure that any data being transferred from other databases is accepted
- Ensure that an audit trail is provided when any changes are made.

Once all system testing has been carried out and documented, the system can be released for use in the study. Confirmation of this release notice should be documented in the trial master file (TMF)

4.1.4 Data Entry

In order to ensure the accuracy, completeness, legibility and timeliness of the data reported in the CRFs, the study team should ensure that data entry guidelines and/or training is provided to site staff. This will limit the number of data errors and reduce the number of queries required (see section 4.1.6).

4.1.4.1 pCRFs

pCRFs are completed by the Investigator or designee and are then faxed or posted to the study team for data entry into the clinical database. A copy of each pCRF will be retained at the investigator site. Where pCRFs are faxed to the study team, it is not necessary for the pCRF to be on NCR paper.

When using pCRFs the study team should consider using wallets to ensure all participant data is kept together in one place.

Where pCRFs are used, it will be necessary to use a clinical database to facilitate analysis of the dataset(s). The clinical database may be in the form of a MS Excel spreadsheet. Data is to be transcribed from the pCRF into the database by the study team. It may be appropriate to contract out this activity. Where this is considered, please contact R&D for guidance on vendor selection.

4.1.4.2 eCRFs and Electronic Transfer of Data

Data is transcribed from the source data[2] into the database by the Investigator or designee.

Where data is transferred electronically into the database from a service provider, such as a central laboratory, there must be a record of the data held at the investigator site or with the service provider as this is considered to be the source data.

4.1.5 Data Validation

Data validation is the process of checking data to ensure that it is sensible and reasonable. Methods of data validation should be determined before conducting the study. This may include checks such as consistency, accuracy and completeness. Validation may occur at the time of data entry within an eCRF, or in a timely manner for pCRFs. All checks, whether computerised or manual should be documented in a validation specification or data validation plan.

4.1.5.1. Self-Evident Corrections

Self-evident corrections, such as errors to dates written in early January stating the previous year, can be made to data without specific referral to the investigator with a data query however these should be limited and not used as a substitute for a formal data query process. Where self-evident corrections or assumptions are made, these should be documented in a Self-evident Corrections document and should be approved by the investigator and statistician prior to implementation.

A list of all self-evident corrections made to the data must be provided to the investigator at the end of the study.

4.1.6 Data Cleaning

Data cleaning is the process for identifying and correcting inaccuracies in the dataset before analysis. Data queries are raised and sent to site for all discrepancies identified from Quality Control (QC) and data validation processes. All inaccuracies identified, should be corrected by the investigator or designee only, unless it is a self-evident correction agreed with the investigator ahead of the conduct of the study. All queries raised and subsequent changes made to the data must be documented via an electronic or paper audit trail.

4.1.6.1. Writing Queries

[2] Source data is defined as all information in original records and certified copies of original records of clinical findings, observations, or other activities in a clinical trial necessary for the reconstruction and evaluation of the trial

Queries should not be leading. Data within the eCRF or pCRF should reflect what's written in the source documentation. It is not appropriate for the study team to influence what data is entered when querying anomalies. Consider the wording of queries to avoid leading. For example: Where an AE onset date is 01 Mar 2016 and the end date is 02 Feb 2016. This suggests that one or both of the dates are incorrect. Consider writing a query as follows.

"Please review the onset date and end date of this AE"

NOT

"Please change the end date of this AE to 02/03/2016."

Unless you have reviewed the source documentation and can advise site where to locate this data within the source documentation it is inappropriate to make assumptions about the data.

4.1.6.2. pCRFs

Data clarification forms (DCFs) should be used to document any data queries raised. Any subsequent change or correction to data should be initialled, dated and explained if necessary and should not obscure the original entry in the pCRF.

4.1.6.3 eCRFs

Where eCRFs are used, data queries will normally be generated in the clinical database and any subsequent changes to data made via eCRFs will be documented via an electronic audit trail. If an electronic audit trail is not available then a paper audit trail of any changes must be maintained via the use of DCFs to request changes.

4.1.6.4. Data Coding

Data Coding is the process of applying a standard code to verbatim terms added to free-text fields. There are a number of standard codes available for use, such as MedDRA or WHO Adverse Reaction Terminology. Coding can be applied to a number of datasets but particularly for large trials, coding of Adverse Events (AEs) should be considered. This process can be carried out within the system automatically or it can be done manually. Often an element of manual coding will be required even for automatic coding.

4.1.6.5. Pharmacovigilance Reconciliation

Pharmacovigilance reconciliation is the process of ensuring that serious adverse events (SAEs) and serious adverse reactions (SARs) reported to the study team are reconciled with the data held in the clinical database. The reconciliation process should be formally documented including who is responsible for carrying

out this task and when this process will be carried out. For small studies where few SAEs are reported, it may be appropriate to reconcile at the end of the study only. Often this process will be required as part of the preparation for the development safety update report (DSUR) (see SOP-JRO-07 and SOP-JRO-08)

4.1.7 Data Quality

The procedure for assessing quality before the data are analysed should be formally documented before conducting the study. An escalation process for dealing with root causes should also be specified. This could be documented in the data management plan or a specific QC plan.

4.1.7.1. Error Rate

Error rates are often used to determine the quality of the dataset(s). A proportion of the database is selected, not necessarily at random but often a risk based approach is used, where QC is performed on the data and then a decision is made as to whether the data are accurate enough to ensure the reliability of the results.

4.1.7.2. Source Data Verification (SDV)

SDV is the process of checking the source data against the data in the clinical database and it is a key method of assessing data quality. This is often carried out by Monitors/Trial Managers within the study team. The process for carrying out SDV should be detailed in the monitoring plan (see SOP-JRO-19).

4.1.8 Database Lock

Database lock is the process of locking the data and restricting access to it once it is deemed final and ready for analysis. The database lock process should be formally documented before conducting the study and may be detailed in the Data Management Plan (DMP). It may be necessary to carry out a similar process for interim analysis and this is often called a snapshot, where the data is extracted at a pre-defined time, or a temporary database freeze, where access to the database is restricted for a limited time in order to carry out data extraction.

All outstanding queries should be resolved or explained before the database lock is carried out. After the data has been locked, it may be necessary on occasion to unlock the data to resolve any outstanding queries, however, this should be avoided where possible. If this process is required, justification and approval should be clearly documented and provided to the statistician before analysis begins.

4.1.9 Data Release

Data release is the process of providing the dataset(s) to the statistician for analysis. The method of how and when this will be carried out should be formally documented

before conducting the study. This may be detailed in the DMP or in the Statistical Analysis Plan (SAP).

For studies where requests for access to the data are expected, a formal process for approving/denying requests and providing data must be in place and decisions should be documented on each occasion. In some cases it may be appropriate to use an independent access committee to assess all requests and their decisions should be communicated in writing to the Statistician and Trial Steering Committee. Checks should also be carried out to ensure that data can be shared in accordance with the protocol and the consent process.

The method for transfer of data to the requester should be considered in order to maintain participant and data confidentiality. Security measures should be implemented where possible such as:

- Sending emails using nhs.net accounts and sending to nhs.net accounts to ensure it remains behind the firewall
- Sending confidential information in zipped files using 7 zip or other encryption software and providing the password by another method.

All data releases should be recorded in a document such as a data release log.

4.2 Documentation

It is essential that all aspects of the data management process are documented appropriately. Some of the key documents that should be considered are detailed below:

4.2.1 Essential Documents

Essential documents are documents which individually or collectively permit evaluation of the conduct of the trial and the quality of the data produced. The essential documents that specifically relate to data management are as follows:

- Sample Case Report Form and any revisions (Investigator Site File (ISF) and TMF)
- Signed, dated and completed CRFs (ISF and TMF)
- Documentation of CRF corrections and a signature sheet showing the signatures of those that are authorised to make CRF corrections (ISF and TMF)

4.2.2 Data Management Plan

The DMP is a key document for any study but is essential for all large and/or complex studies. The entire data management process from database design to data release should be detailed and should be specific to the study. It should cover areas such as, but not limited to:

- Systems to be used
- Validation
- The query process
- Pharmacovigilance reconciliation
- The database lock process
- The data release process

The DMP should be reviewed and signed by the Statistician and by Sponsor or designee and filed in the TMF. This is a live document and should be updated when any changes to any data management activity is required.

4.2.3 eCRF Specification and Annotated CRFs

For all data collection tools a specification should be created to detail all data collection fields within the database. This should detail the following:

- The questions being asked
- The response type, e.g. date, numerical, text
- The format of the response e.g. Date = DDMMMYYYY, Numerical = 4 digits
- Whether the question is mandatory or conditional

This specification may be in the form of an annotated CRF. For pCRFs, the annotated CRF is completed by making notes against each field on the paper form. For eCRFs, this can be carried out by using screen shots of the database. This document creates a link between the questions in the CRF and the dataset(s) used for analysis.

This document should be signed by the Chief Investigator and filed in the TMF.

4.2.4 Data Validation Plan (DVP) or Specification

The data validation plan or specification may not necessarily be a stand-alone document but could be part of the DMP. It should list all computerised and manual data checks that will be carried out for the study and at what point.

The DVP should be reviewed and signed by the chief investigator and Sponsor or designee and filed in the TMF.

4.3 Oversight of Data Management Activities

The data management activities for the study should be determined, communicated to and agreed with R&D before the study begins within a delegation agreement. A timeframe for when the necessary documentation will be completed should also be agreed before the conduct of the study.

All data management documentation held by the study team will be reviewed by R&D during the annual audit cycle.

5. Review and Monitoring

This SOP will be reviewed every two years or more frequently in response to changes in legislation or guidelines. The use of this SOP will be monitored on an ongoing basis and during the annual audit cycle.

6. References

6.1 ICH Harmonised Tripartite Guideline, Guideline for Good Clinical Practice (1996) – http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf (accessed Nov 2016)

6.2 Regulation 31A (4) of SI 2004/1031 – http://www.legislation.gov.uk/uksi/2006/1928/pdfs/uksi_20061928_en.pdf (accessed Nov 2016)

6.3 MHRA Good Clinical Practice Guide, 2012

6.4 SOP-JRO-07 Adverse Event Reporting for CTIMPs and ATMPs

6.5 SOP-JRO-08 Adverse Event Reporting for non-CTIMPs

6.6 SOP-JRO-19 Monitoring of Research

7. Appendices

Appendix 1 – Glossary of Terms

7.1 Appendix 1 – Glossary of Terms

Term	Definition
Adverse Event (AE)	Any untoward medical occurrence in a subject to whom the medicinal product has been administered, including occurrences that are not necessarily caused by or related to that product.
Annotated CRF	A CRF which states variable descriptions next to data entry fields.
Audit Trail	Documentation that allows reconstruction of the course of events.
Case Report Form (CRF)	A questionnaire used to collect data from each participating patient.
Data Management Plan (DMP)	A document detailing all specific data management activities to be carried out for a clinical trial.
Data Clarification Form	A form used for documenting queries, the response to queries and the subsequent changes to the data.
Data Query	A request for data to be changed or corrected.
Data Release	The process of making the data available for analysis.
Data Validation	The process of checking the data for consistency, accuracy, completeness and reasonability.
Data Validation Plan (DVP)	A document detailing all checks to be carried out on the data, both manual and electronic.
Database	A computerised comprehensive collection of related data organised for convenient access.
Database Freeze	A temporary process of preventing changes being made to the data within the database.
Database Lock	A permanent process of preventing changes being made to the data within the database.
Database Validation	The process of testing the database to ensure that it performs as expected.
Dataset	A collection of data records.
Development Safety Update Report (DSUR)	A common standard for periodic reporting of drugs under development among the ICH regions.
Essential Documents	Essential documents are documents which individually or collectively permit evaluation of the conduct of the trial and the quality of the data produced.
Electronic CRF (eCRF)	*(see CRF)* An electronic version of the CRF. Often this is referred to as the clinical database where data can be entered directly by site staff.
Investigator Site File (ISF)	A file of essential documents held by the Principal Investigator at a trial site.
Paper CRF (pCRF)	*(see CRF)* A paper version of the CRF. The CRF is completed in hard-copy and the answers are then transcribed into a clinical database such as an excel spreadsheet.

Pharmacovigilance	The practice of monitoring the effects of medical drugs after they have been licensed for use.
Quality Control (QC)	Operational techniques and activities undertaken within the quality assurance system to verify that the requirements for quality of the trial-related activities have been fulfilled.
Serious Adverse Event (SAE)	Any AE that results in death, is life-threatening, requires in-patient hospitalisation or prolongation of existing hospitalisation, results in persistent or significant disability or incapacity, or consists of a congenital abnormality or birth defect.
Source Data	All information in original records and certified copies of original records of clinical findings, observations, or other activities in a clinical trial necessary for the reconstruction and evaluation of the trial.
Source Data Verification (SDV)	The act of checking the accuracy and completeness of the CRF entries against the source documents in accordance with the protocol requirements.
Statistical Analysis Plan (SAP)	A document detailing the statistical methodology and planned analysis for the trial.
Trial Master File (TMF)	A file containing essential documents held by the sponsor or Chief Investigator.
User Acceptance Testing (UAT)	The last phase of the database testing process. During UAT, actual database users test the database to make sure it can handle required tasks in real-world scenarios, according to specifications.