| | |
|---|---|
| **Department** | Research & Development |
| **Title of SOP** | **Data Management SOP** |
| **SOP reference no.** | R&D GCP SOP 13 |
| **Author** | G Constable |
| **Reviewed by** | J Illingworth, S Moffat and L Cox |
| **Current version number and date** | Version 6, 17.08.2022 |
| **Next review date** | 17.08.2022 |
| **Target audience** | Research, pharmacy and R&D staff |
| **Distribution** | HUTH R&D internet<br>Click on GCP SOPs for HUTH-sponsored CTIMPs<br>https://www.hey.nhs.uk/research/researchers/gcp-sops-for-hey-sponsored-ctimps/ |

**When this document is viewed as a paper copy, the reader is responsible for checking that it is the most recent version.**



Signed copy kept in R&D Department

| Authorized by | Name | Sign | Date |
|---|---|---|---|
| **R&D Director** | Professor Thozhukat Sathyapalan | DocuSigned by: Prof Thozhukat Sathyapalan 645722D6386948C... | 18/8/2022 |
| **R&D Manager** | James Illingworth | DocuSigned by: James Illingworth 3DDA0F96873F430... | 15/9/2022 |

This page details the version history and the main changes made for each new version.
The new changes are in red font.

| Version Log | | |
|---|---|---|
| **Version number and date** | **Author** | **Details of significant changes** |
| Version 1, 14.04.11 | J Pacynko | Original SOP approved by R&D Committee on 14.04.11. |
| Version 2, 13.08.12 | J Pacynko | Page 9 - 4.10 clarified to state that for blind randomized studies, a copy of the locked dataset and end of trial form must be received by R&D prior to release of the randomization list by pharmacy. |
| Version 3, 29.08.12 | J Pacynko | Appendix 1 – Data management file note, database lock section up-dated on page 13 to state that for randomized trials R&D requires a copy of the locked trial data and end of trial form before release of randomization list. |
| Version 4, 05.10.15 | J Pacynko | Amended information is in red type, in summary:<br>4.2 R&D will supply investigators with either a Microsoft Excel Spreadsheet 2010 or a Microsoft Access Database 2010 already with the audit trail set up.<br>4.3 Up-dated information on MedDRA for medical data coding.<br>4.4 Addition of data entry in the electronic CRF.<br>4.5 Addition of the data query process for paper and electronic CRFs.<br>4.7 For multi-site trials, addition of information about the transfer of CRF data from sites to a co-ordinating centre.<br>Appendix 1 – Data management plan up-dated. |
| Version 5, 18.02.19 | J Pacynko | -Mention of Hull and East Yorkshire Hospitals NHS Trust has been replaced with Hull University Teaching Hospitals NHS Trust.<br>-Addition of Sponsor requirement to review the; Data Management Plan, Computer System Validation reports, User Access Testing reports, Statistical Analysis Plan, Clinical Study Report and publications.<br>-4.1.5 CSV and UAT to be carried out on all electronic data systems.<br>-4.7.8 For multi-site trials run by a CTU with an electronic CRF, at the end of the trial, the CTU should send sites the eCRF data and metadata for sites to review prior to revoking site's access to the eCRF.<br>-4.9 Data should be backed up contemporaneously.<br>-4.11.2 The final checked dataset is protected from deletion and editing by password protection and restricted access to authorised staff only.<br>-4.11.3 Randomisation list to be released to Sponsor (R&D QA) only at the end of the trial. R&D QA to then forward to statistician.<br>-4.12 Addition of statistical analysis section.<br>-4.12.3 The order of data management events from the spreadsheet/database going live to the statistical analysis must be documented in the Data Timelines Log. |
| Version 6, | G Constable | Add amends<br>4.0 Clarification of wording<br>4.1 Addition of GCP principles<br>4.1.1 - Adding of suitable location to the possible locations for the database (to capture databases not located on trust or university servers).<br>-Added wording to ensure database and CRF are developed in parallel to maintain consistency and accurate data storage.<br>4.1.2 Additional wording to ensure archiving is in line with related SOP<br>4.2.3. Clarification on access to spreadsheets (used as databases).<br>4.2.5 Additional wording to re enforce that significant change to the database requires version control.<br>4.7.10 Additional wording emphasising the need for explicit levels of access, such as 'read only'. Additional wording stating all electronic data transfer requires a full audit trail.<br>4.7.11 Rewording of statement for clarity<br>4.12.1 Added line to the statistical analysis section stating that this section (4.12) is not exhaustive and that *SOP 20 Statistical Analysis* should be followed. |

**Please note for definitions of acronyms refer to Appendix 1 of Management of SOPs. Refer to Appendix 2 of Management of SOPs for the standards to which clinical trials that investigate the safety and/or efficacy of a medicinal product are conducted.**

**All the R&D GCP SOPs are available at:**
**https://www.HUTH.nhs.uk/research/researchers/gcp-sops-for-HUTH-sponsored-ctimps/**

# 1 Purpose

**1.1** This SOP describes the procedures for the management of data produced from Hull University Teaching Hospitals NHS Trust sponsored clinical trials with investigational medicinal products (HUTH-sponsored CTIMPs). This includes the procedures involved with collecting, validating, entering data into a spreadsheet/database and storing the data from HUTH-sponsored CTIMPs.

**1.2** It should be noted that within this SOP the Sponsor tasks are carried out by the QA staff in the R&D Department at Hull University Teaching Hospitals NHS Trust.

# 2 Introduction

**2.1** An essential element of conducting a clinical trial is ensuring that the data collected is of the highest quality. Equally important is that the data management process maintains the quality of the data collected.

**2.2** UK CT Regulations Schedule 1, Part 2 (9) stipulates that all clinical information shall be recorded, handled and stored in such a way that it can be accurately reported, interpreted and verified, while the confidentiality of records of the trial subjects remains protected.

**2.3** With the above aim, the SOP describes the data management process including: data collection, data entry, data verification, data storage and protection and database lock.

**2.4** The SOP also describes the use of an Independent Data Monitoring Committee for assessing data during interim analyses, and how such a committee should operate.

# 3 Who should use this SOP

**3.1** This SOP should be used by:

- All research staff involved with HUTH-sponsored CTIMPs – Chief/Principal Investigator, co-investigators, research nurses, clinical trial assistants, trial managers, clinical trial co-ordinators, data managers, administrators etc.
- Clinical trials pharmacy staff – technicians and pharmacists.
- All HUTH R&D QA staff who manage the sponsorship of HUTH-sponsored CTIMPs.
- Research staff involved with clinical trials sponsored by an external organisation where the sponsor has no SOP for data management. HUTH R&D SOPs are defaulted to in this case.

**3.2** ICH GCP 5.5 specifies that the Sponsor should utilise appropriately qualified individuals to supervise the overall conduct of the trial, to handle the data, to verify the data, to conduct the statistical analyses and to prepare the trial reports.

**3.3** For single-site trials where there is just the one investigator site at HUTH, the Sponsor normally delegates data management to the Principal Investigator (PI). If the PI delegates data management to another member of the research team this should be clearly specified on the trial Delegation Log.

**3.4** For multi-site trials sponsored by HUTH, data management may be undertaken by a Clinical Trials Unit.

# 4 Data management process

The process of data management includes converting the data collected using data collection tools, known as Case Report Forms (CRFs), into electronic data that can then be statistically analysed.

The procedures described below will need to be adapted according to the size and complexity of the trial, as smaller trials may not require all the processes described.

### 4.1 Trial database

Once the CRF has been designed in accordance with the protocol, the database should be designed to store the information collected. The type and size of the database will depend on the size and complexity of the study and could vary between a standard Excel spreadsheet to a more technical Data Management System.

The specific database designed for the trial must conform to the ICH GCP principles as follows: "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." *E6(R2) 1.24*.

4.1.1 When developing a database the following points should be addressed, although they may need to be amended according to the size and complexity of the trial.

The database should:

- Be held on an appropriate NHS, University server or a suitable online based platform.
- Have security access requirements (unique username and password) with no public access
- Have appropriate levels of access (e.g. pharmacy to drug accountability section only)
- Have the ability for more than one user to use the system at the same time
- Have regular adequate back up of the data
- Have disaster recovery plans (including preventative, detective and corrective measures)
- Enable data to be exported to a data analysis package (e.g. SPSS, STATA) with minimal effort
- Ensure data queries can be generated
- Ensure an audit trail of data corrections can be generated.
- Be subjected to computer system validation (CSV) checks to demonstrate that the database is fit for purpose prior to use. Results of validation testing should be documented in CSV reports and be reviewed by the Sponsor. The review should be documented using the CSV review checklist (Working Instruction 24) and in the Data Timelines Log (Working Instruction 19) and saved in the TMF.
- Be designed in parallel with the approved protocol and CRF to ensure data collected at all time-points are accounted for and comply fully with DI principles.

4.1.2 Those that set-up and manage the database should:

- Provide training for relevant staff on the use of the database (record on Training log).
- Maintain a security system that prevents unauthorised access to data (ICH GCP 5.5.3)
- Allow direct access to the database by monitors, auditors and inspectors on a read-only basis and under the supervision of the Data Manager.
- Ensure that data changes are documented and that there is no deletion of entered data (i.e. maintain an audit trail, ICH GCP 4.9.3 and 5.5.3).
- Keep a list of individuals (use the trial Delegation Log) who are authorised to make data changes (ICH GCP 5.5.3)
- Safeguard the blinding, if applicable (e.g. maintain the blinding during data entry and processing, ICH GCP 5.5.3).
- Make sure that if data are transformed during processing, it is always possible to compare the original data with the processed data (ICH GCP 5.5.4).
- Ensure the database systems are checked and maintained on a regular basis.
- Be aware of appropriate software upgrades relating to the database.
- Define database lock-down procedures to ensure access to the final dataset is permanently restricted for final analysis and report, prior to archiving – see 4.11 below.
- Ensure archived data is maintained correctly and in-line with R&D GCP SOP 14 Archiving.

- Ensure that the database will be complete and accessible throughout the period of archiving by looking at digital preservation issues.
- Ensure stored trial data is archived on an appropriate NHS Trust server with security access restrictions.
- Produce a Data Management Plan or trial-specific SOP for managing the trial database (ICH GCP 5.5.3) – see 4.8 below.

4.1.3 It should be noted that where the database is to be held/archived by an external organisation (e.g. CTU) on behalf of the Sponsor that appropriate security access restrictions are in place. This should be clearly demonstrated in a written Agreement between the organisation and Sponsor.

4.1.4 Researchers are advised to use commercially produced and validated software wherever possible. Specialist software that has been produced 'in-house' or as a one off application by a commercial company must be subjected to computer system validation processes (CSV). Results of validation testing (CSV reports) will be reviewed by the Sponsor and must be filed in the Trial Master File (TMF) – see 4.1.2 above.

4.1.5 Investigators should ensure that accreditation checks, CSV and user access testing (UAT) have been carried out for all other electronic data systems e.g. for electronic randomisation systems or electronic CRFs. This is to demonstrate that systems are fit for purpose before use.

4.1.6 CSV and UAT reports for electronic randomisation systems and CRFs will need to be reviewed by the Sponsor and the review should be documented in the CSV review checklist (WI 24) and Data Timelines Log (WI 19) and saved in the TMF.

### 4.2 Data confidentiality and audit trail

The following points should be taken into consideration when collecting clinical trial data:

4.2.1 Data should only be collected that is relevant for the purpose of the clinical trial according to the trial protocol. It is recommended to seek advice from a trial statistician as early as possible in the trial design process to facilitate this.

4.2.2 All staff are responsible for keeping data secure and confidential at all times and must comply with the UK Clinical Trial Regulations, ICH GCP, General Data Protection Regulation (GDPR) and associated Trust policies regarding confidentiality and security.

4.2.3 Data held electronically should only be accessible to authorised personnel with their own log-in username and password. Where a password protected Excel spreadsheet is deemed appropriate, access to the Y drive that holds the spreadsheet is via individual usernames and passwords but direct access to the spreadsheet is via a password that will be shared by delegated individuals within the research team i.e., PI and RN etc.

4.2.4 All changes to data held electronically must have a clear and complete audit trail (activity history). In other words, it must be clear who has entered or deleted any data and when. If any changes have been made to the data it would also be useful to record the reason why, if this is possible.

4.2.5 R&D QA staff will supply investigators with either a Microsoft Excel Spreadsheet or a Microsoft Access Database already with the audit trail set up. The spreadsheet or database will include the trial title and the first 3 columns; patient initials, patient study number and date of birth. The spreadsheet or database will be password protected. The audit trail history will be checked by the R&D monitor at monitoring visits. Any significant changes to the database will be tracked via version control.

## 4.3 Data coding

4.3.1 It is recommended to seek advice from the trial statistician on how to code for the CRF data.

4.3.2 CRF data which is in text format should be coded, using either a numerical or alphabetical code that can then be used for analysis. These codes should be decided before data entry begins e.g. 1 = Yes, 0 = No. Codes should also be in place for answers such as 'not known' or 'not applicable' or for missing data. It is important to make sure that the value chosen to represent missing data would not be feasible as an actual response.

4.3.3 Clinical data also needs to be coded for analysing adverse events. The Medical Dictionary for Regulatory Activities (MedDRA) is a widely used database of coded medical terminology. A code is assigned for each medical condition and adverse reaction.

4.3.4 MedDRA can be accessed through http://www.meddra.org/. Subscription is free for non-profit / non-commercial organisations.

4.3.5 Coding may be done either during the initial data collection from the participant, or after the data collection but prior to entering the data on the database, or when the data is entered into the database.

4.3.6 Where auto-coding is not possible at data entry then manual coding should be performed.

## 4.4 Data Entry

### Paper CRF (pCRF)

4.4.1 This applies to HUTH-sponsored single or multi-site CTIMPs where the CRF is in paper format.

4.4.2 For multi-site CTIMPs, where pCRFs are sent by sites to a CTU, on initial receipt of pCRFs, the forms should be date stamped.

4.4.3 The data should be entered into the database by trained data entry staff.

4.4.4 The data entry process should be defined either in the Data Management Plan or in the trial-specific Data Management SOP – see 4.8 below.

4.4.5 During data entry by trained staff, human errors may occur. Two methods can be used to reduce the incidence of errors: single data entry with control checks or double data entry.

### Single data entry with quality checks

4.4.6 This method is more suitable for smaller single or multi-site studies with less staff available for data entry and/or less sophisticated database software. Once the data has been entered for a CRF page or visit, a visual check is done between what is recorded on the paper CRF and what was entered into the spreadsheet or database. A reminder and tick box for this check should be present in the footer of each CRF page with wording such as '□ Data entered checked and consistent with pCRF and medical notes'. Either the same person may check the data entered against the source data or a second member of staff may do all or some quality checks.

### Double data entry

4.4.7 This method involves two people entering the same CRF data onto the database independently of each other. Depending on the software used, the data is entered twice into the database on two separate files, which are then compared by the system for accuracy. If the two entries do not match this would be flagged up by the database. Alternatively when the second data entry person enters the data, if it differs from that entered by the first person, a message immediately appears on screen and the original data can be checked. This method depends on the availability of a technically capable database and sufficient staff time.

### Electronic CRF (eCRF)

4.4.8 For large multi-site studies, the use of an eCRF where data is entered at site, has many advantages. Data checks can occur by site staff at the point of data entry at site which reduces the amount of data queries generated in the eCRF system. The co-ordinating centre can remotely monitor recruitment into the trial and the quality of data, and there is a low risk of losing data.

4.4.9 Sites will need to be provided with clear instructions on how and when to enter data into the eCRF and the level of quality checks to be done against the source data. Sites may be asked to do single data entry with either the same person checking the data entered against the source data or sites may be asked to involve a second member of staff to do all or some quality control checks. The method of quality checks needs to be specified clearly in the eCRF instructions to site and in the Data Management Plan or SOP (see 4.8 below).

## 4.5 Data query (DQ) process

### Paper CRF

4.5.1 Prior to and during data entry, CRFs must be checked for missing data or incomplete responses or data outside of normal ranges. If any inconsistencies are found these should be queried with the investigator.

**Single-site trial at HUTH**
4.5.2 Any change to the data in the pCRF should be initialled, dated and explained (if necessary) and should not obscure the original entry. If any change is required to the spreadsheet it should be clear who has made the change and when and the reason why (if possible).

**Multi-site trial managed by a Clinical Trials Unit (CTU) - MHRA GCP Guide 8.5.7**
4.5.3 If the **original** pCRFs are **kept at site** and a copy of the CRF data has been sent by secure fax, scan or post to the CTU. DQs can be emailed to sites and then the original pCRF can be amended at site if required (rather than the DQ form) and then a copy of the amended CRF re-sent to the CTU. Any required change to the data in the pCRF should be initialled, dated and explained (if necessary) and should not obscure the original entry.

4.5.4 It is recommended that the original pCRFs are kept at site but if the **original** pCRFs are **sent to the CTU** then data query (clarification) forms should be sent to site either by secure post or email (to be printed out at site). The paper DQ form should be completed and signed by the PI or delegated person and then sent to the CTU with a photocopy kept with the pCRF copies at site.

4.5.5 A record should be kept by the CTU of all data queries sent out and replies received.

4.5.6 Sites should be given clear instructions on how to respond to data queries.

4.5.7 The CTU should check that the DQ form has been signed off by an authorised person on the site's Delegation Log.

**Electronic CRF**

4.5.8 If the eCRF is set up with automatic data entry checks then values that are out of expected range, implausible, incorrect or missing, can be flagged up to site staff and resolved at the time of data entry.

4.5.9 An electronic audit trail must be maintained in the eCRF so that the CTU can check who has made any changes to the data, when and why and if those changes have been made by an authorised person on the site Delegation Log.

### 4.6 Data validation

4.6.1 An essential part of the data management process is validation to ensure the most accurate and complete set of data is provided for the statistical analysis. Data validation can be carried out at various stages during the trial:

4.6.2 **When CRFs are completed by the investigator:** To improve accuracy at this stage all staff completing CRFs should be sufficiently trained in their completion. There should be a check done by research staff against the source data in the medical records as they enter data in the eCRF. For large multi-site studies, a CRF completion manual or instructions will be necessary.

4.6.3 **As part of monitoring:** Validation is also carried out as part of the on-going GCP monitoring of the trial by either the R&D monitor or CTU. Validation via monitoring is done through Source Data Verification (SDV). SDV involves checking the data in the CRFs against that in the original source documents (usually patient's medical notes) for accuracy.

4.6.4 **When data is entered into the database by data entry staff:** During data entry the two methods described above (single with checks or double) can be used to make sure the data is accurate and complete. In addition, if the trial database has software that enables automatic data entry checks then this useful system of validation can be used. Alerts can be set up in the eCRF/database for; out-of-range values, incorrect values

e.g. a numeric value entered rather than text, implausible values or for missing values. An Edit Check Specification (ECS) document should be put together by the trial investigators, statistician and data management staff to provide full details of the data entry checks that have been set up, and all checks should be tested before the trial begins.

4.6.5 **Post data-entry:** After the data has been entered and if the database software is able to, it is advisable to carry out systematic post-entry computer tests. Lists should then be created of the following:

- All missing values should be listed
- All values outside of pre-defined range

Logical checks should also be performed to ensure consistent reporting between relevant fields and that there are no implausible difference between fields e.g. male and pregnant.

4.6.6 All checks should be defined before the trial starts, and should be described in the Edit Check Specification document described previously. Data validation should continue until all missing values and inconsistencies are corrected or clarified.

## 4.7 Data storage, protection and transfer

4.7.1 During the entire data management and validation process it is essential that all trial data is processed according to the agreed and current protocol which ensures are kept in a secure location and in accordance with the terms of the GDPR.

4.7.2 Participant confidentiality must be maintained at all times and trial records should be kept in a pseudonymised form identifying participants by their study number and initials rather than name or hospital number.

4.7.3 All stored paper CRFs should be kept in a secure environment such as a locked filing cabinet or cupboard in a room kept locked when not in use. Secure also means protection against environmental damage such as damp, fire or pests. CRFs should not be kept where there are any water sprinklers.

4.7.4 Frequently for HUTH-sponsored CTIMPs, at least part of the CRF is a source document (e.g. VAS, questionnaires or diaries completed by patients). 'The Sponsor should never have sole control of the CRF when it has become a source document at any time' (MHRA GCP Guide 8.2.5).

### Multi-site trial managed by a CTU

### Paper CRF

4.7.5 The investigator site should at all times maintain an independent copy of the data provided to the CTU/Sponsor. Where pCRFs are part source data and are being sent to a CTU for data entry, the **original** CRFs must be retained by the principal investigator **at site**. Copies of CRFs must be sent from site to the CTU via secure email/fax/post. For email transfer, both sender and recipient should have an NHS Mail account (see HUTH Policy CP134 Information transfer & storage procedure). The CTU must keep a record of all CRFs received.

4.7.6 Although very unlikely, if no data in the paper CRF is source, then the original CRFs can be sent to the CTU for processing and a photocopy left at site. However, it must be noted that there is a risk of losing the original CRF in transit between site and the CTU.

### Electronic CRF

4.7.7 Where CRFs are **electronic**, the investigator site should at all times maintain an independent copy of the data provided to the CTU/Sponsor. So that the CTU/Sponsor do not have exclusive control of the data, either paper worksheets should be used by sites or a copy of the eCRF data should be saved during the trial on the site's server. If neither of these are possible then a printout of the CRF should be signed and dated by the Principal Investigator (certified copy) and retained at site **prior to** submitting the eCRF data to the CTU.

4.7.8 At the end of the trial, the CTU should send sites the eCRF data and metadata for sites to review prior to revoking site's access to the eCRF.

4.7.9 In summary, for HUTH-sponsored multi-site trials, whether the pCRFs or eCRFs contain source data or not, we require that:

- Original pCRFs stay at site during the trial and at the PI's archive organisation at the end of the trial.

- Copies of eCRFs stay at site on the site's server during the trial and during the archive period. If for some reason it is not possible for a site to retain eCRFs on the site's server then a certified copy (signed by PI) of the CRF printout should stay at site or in the PI's archive.

**In summary**

| HUTH-sponsored single-site CTIMP | HUTH-sponsored multi-site CTIMP managed by CTU<br>CRFs may or may not contain source data | | |
|---|---|---|---|
| pCRFs only | pCRFs only at sites | eCRFs at sites with paper worksheets | eCRFs at sites with no worksheets |
| Original pCRFs kept at site during trial and archived by R&D archivist at end of trial. | Original pCRFs to stay at site or at PI's archive. Copies to be sent to CTU for data entry. | Original worksheets to stay at site or PI's archive. | Copy of eCRFs stay at site on site's server or if not possible, then certified copy of CRF printout to stay at site or PI's archive. |

4.7.10 If electronic data transfer is used for multi-site trials, this should be via a secure system, password protected and encrypted where possible. The database itself should be password protected, with each data entry staff member having their own password. There should be explicit levels of access (such as read-only access and data entry permissions) assigned appropriately and recorded on the appropriate logs. All electronic data transfer requires a full audit trail.

4.7.11 The Principal Investigator at the site is responsible for maintaining data integrity throughout the trial including maintaining accuracy and completeness of data when uploaded to the database. All necessary precautions to protect an investigator from viewing the entire database should be considered, where possible, thereby preventing investigator decisions based on interim analysis.

4.7.12 All data stored on HUTH network computers must adhere to:

- CP134 HUTH Confidentiality and Information Security Policy incorporating:
  - Procedure for protecting patient information in public areas
  - Information transfer & storage procedure
  - Procedure for managing photographic images, video & audio recording of patients

4.7.13 Whilst a trial is active and after its conclusion, all data (including source data) should always be available when needed to authorised individuals such as monitors, auditors and inspectors, to meet their regulatory obligations.

### 4.8 Data Management Plan or Data Management SOP

4.8.1 For singles-site or small multi-site trials, the Principal Investigator will be required to complete a Data Management Plan (DMP) for the TMF, specifying the data management process before the trial starts. The DMP is Working Instruction 17 saved on the ClinicalGov Y drive in Y:\Research\GCP SOPs & forms.

4.8.2 For large multi-site trials, before the trial starts, it is recommended that a Standard Operating Procedure for Data Management specific for the trial, is produced and updated as necessary throughout the trial. The trial-specific SOP should address all the points in this SOP and include details relevant to the trial. The SOP should contain the following information:

- Roles, responsibilities, contact details of all trial staff with regard to data management, data access, data entry, data changes.
- Access controls and restrictions of each staff member.
- Details of the flow of data from the investigator site to the archiving site.
- Identify all vendors and parties holding associated trial data and metadata (e.g. randomisation service provider)
- Trial database including;
  - data audit trail
  - maintenance
  - disaster recovery plans
  - contemporaneous back-up of data
  - identify metadata
- Procedures on how to complete the CRFs (CRF completion manual).
- Data coding (include coding of AEs).
- Data monitoring e.g. central monitoring for missing data, unusual data patterns, identification of outliers, frequency of data monitoring.
- Data Monitoring Committee
- Data entry
  - How to use the data entry system
  - Double or single entry
  - Expected ranges for data values
  - Procedures in case of discrepancies
- Details of automatic data entry checks (Edit Check Specification Document)
- Description of data entry validation
  - Who checks the consistency of the data?
  - What logical checks are done?
- Data query process
  - Who sends data queries to Investigators?
  - What is the format of the data query form?
  - How many days are allowed to answer a data query?
  - Who decides that a data query is resolved?
- Data protection procedures (including electronic transfer requirements)
- Data storage (including confirmation that original/certified copy of pCRF or copy of eCRF is kept in/with the ISF by PI).
- Database lock process
- Statistical analysis process
- Archiving arrangements.

4.8.3 Although the above list is not exhaustive it provides a basis for the Trial-specific Data Management SOP that can be adapted and expanded as necessary.

4.8.4 The metadata must be identified in the Data Management Plan or SOP. Metadata is additional data to the CRF data and is information needed to understand and/or effectively use the CRF data e.g. the key to codes. The audit trail is also considered as part of the metadata. The trial data and metadata may be held by external vendors such as CTUs, randomization service providers, laboratories, IMP manufacturers. As for the trial data, the metadata must be obtained and saved in the TMF.

4.8.5 The Sponsor is required to review the Data Management Plan or SOP. The review should be documented in the Data Timelines Log (Working Instruction 19) and saved in the TMF.

### 4.9 Data backup

4.9.1 A data backup is the result of copying or archiving files or folders for the purpose of being able to restore them in case of data loss. Data loss can be caused by many things, for example, computer viruses, hardware failures, file corruption, fire, flood, theft etc.

4.9.2 There should always be a data backup system in place and data should be backed up contemporaneously. The Hull University Teaching Hospitals NHS Trust IT department has a backup service that provides a reliable means of protecting data held on Trust servers. There is regular daily backup of the data on Trust servers.

4.9.3 For servers that are used to exclusively support research data there may be a charge for this backup service and it is advised that the Trust IT department are consulted regarding IT requirements as early as possible.

### 4.10 Data Monitoring Committees (DMCs)

4.10.1 It is recommended for large multi-site or long-running trials that a Data Monitoring Committee (DMC) is set up to carry out reviews of trial data at regular pre-defined intervals during the trial.

4.10.2 The role of the DMC is to review interim results and determine whether or not there are any safety issues or any reason why the trial should not continue e.g. if interim results are showing strong evidence that the treatment/intervention is superior or inferior to the control.

4.10.3 The data reviewed by the DMC should be as up-to-date as possible and should be validated up to the point of the interim analysis to ensure it is of reliable quality.

4.10.4 The membership of the committee should include experienced trial investigators and statisticians. If all committee members are independent to the research team, then the DMC is termed an Independent Data Monitoring Committee (IDMC). The IDMC protects against potential bias from investigators. In many cases, an independent Data Monitoring Committee is recommended.

4.10.5 If there is a Trial Steering Committee (TSC) for the trial, the DMC would normally make their recommendations for action through the TSC.

4.10.6 More guidance on Data Monitoring Committees is available in the R&D GCP SOP 11 Trial Oversight.

### 4.11 Final dataset lock

4.11.1 The CI/PI is required to provide the Sponsor with a copy of the final locked dataset following completion of the trial and **before** the statistical analysis is performed. The final dataset must be 'locked' and password protected to ensure access is restricted for final analysis and report. The dataset must not be altered in anyway after being locked.

4.11.2 This is not a physical lock but is when the CI/PI has ensured that all the following activities are completed:

- All clinical trial subjects must have completed their final visit and any follow-up visits prior to dataset lock.
- All coding of clinical events must be completed.

- All outstanding queries have been resolved and the dataset updated.
- Any Serious Adverse Events (SAE) queries have been resolved and the dataset up-dated.
- All members of the study team and R&D monitor/QA manager have been notified of the proposed date of lock.
- The final checked dataset is protected from deletion and editing by password protection and restricted access to authorised staff only.

4.11.3 For double-blind randomised trials, the Sponsor must receive a copy of the 'locked' final dataset and signed end of trial notification form before asking the randomisation service to release the randomisation list only to the R&D QA staff. The randomisation service will be informed at the trial set-up phase that they can only release the list to the Sponsor when the trial has finished and that the request will come from R&D QA staff. R&D will then send the list on to the statistician and CI/PI to reveal which trial patient received which treatment and to enable the statistical analysis to be performed.

## 4.12 Statistical analysis

4.12.1 This list is not exhaustive, please follow R&D GCP SOP 20 Statistical Analysis.

4.12.2 There should always be pre-specified statistical methodology documented for a trial, either in the protocol or in a Statistical Analysis Plan.

4.12.3 For small single or multi-site trials, it is usually sufficient for a statistics section to be in the protocol. For large multi-site trials, a Statistical Analysis Plan (SAP) should be produced and finalised prior to database lock and prior to any interim analysis in blinded trials. The SAP will need to be version controlled so that it is clear which is the final version. The Sponsor is required to review the Statistical Analysis Plan and the review must be documented in the Data Timelines Log and saved in the TMF.

4.12.4 The order of events from the spreadsheet/database going live to the statistical analysis must be documented in the Data Timelines Log. This log is Working Instruction 19 and saved on the ClinicalGov Y drive in Y:\GCP SOPs & forms.

4.12.5 The SPSS (statistical analysis software) output will need to be date and time stamped. The date and time that the SPSS output was produced will be recorded on the Data Timelines Log.

4.12.6 All statistical analyses for the trial will be performed by the trial statistician or if performed by the investigator will be checked by the trial statistician.

4.12.7 For all trials, the results of the statistical analysis will need to be reviewed by the Sponsor and this review documented. A suitably qualified colleague in HUTH or the University of Hull will be asked to review the analysis on behalf of the Sponsor and to sign off the Clinical Summary Report (CSR) QC form (Working Instruction 13).

## 4.13 Clinical study report (CSR) and publications

4.13.1 For single and multi-site trials, the Sponsor is required to review the CSR and publications. The review must be documented and saved in the TMF. The Sponsor review of the CSR will be documented in the CSR QC form (WI 13).

### 4.14 Archiving data

4.14.1 The archiving of trial documents and data in paper and electronic format is described in the Archiving SOP 14. Electronic documents and data are archived on the Trust's servers with restricted access to R&D QA staff only. The servers are backed up nightly to disk media.

## 5 Acknowledgements

We thank Imperial College and York Foundation Trust R&D departments for their kind permission to adapt and reproduce their SOPs.

## 6 References

- MHRA Good Clinical Practice Guide – first published in 2012 by The Stationery Office
- HUTH CP134 Confidentiality and Information Security Policy

## 7 Implementation

Implementation of this SOP will conform to the process outlined in R&D SOP 01 Management of SOPs.