St George's University Hospitals NHS
NHS Foundation Trust

# Standard Operating Procedure (SOP)
# Data Management

| | | | |
|---|---|---|---|
| **SOP ID number:** | JREOSOP0038 | **Effective** | 2/11/15 |
| **Version number and date:** | Final Version 2.0 16/10/15 | **Review Date:** | 01/11/17 |
| **Author:** | Debs Rolfe | **Title:** | Regulatory Assurance Manager |
| **Approved by:** | Lucy H H Parker | **Date:** | 16/10/15 |
| **Signature of Authorisor** | | | |

This is a controlled document.
The master document is posted on the JREO website and any print-off will be classed as uncontrolled.

Researchers and their teams are responsible for checking the JREO website for the most recent version.
They may print off this document for training and reference purposes.

| **SOP Chronology** | | |
|---|---|---|
| **SOP Version Number:** | **Reason for Change:** | **Author:** |
| v1.0 | Original Version | Lucy H H Parker |
| V2 | Updated trust logo and reference added to Data Management resource on University website | Debs Rolfe |


© St George's

| Table of Contents | Page |
|---|---|
| 1. Background | 3 |
| 2. Joint Research and Enterprise Office (JREO) Policy | 3 |
| 3. Scope | 3 |
| 4. Definitions | 3 |
| 4.1 Independent Data Monitoring Committee | 3 |
| 4.2 Interim Analysis | 4 |
| 4.3 Trial Steering Committee | 4 |
| 4.4 Case Report Form | 4 |
| 5. Responsibilities | 4 |
| 6. Procedures | 4 |
| 6.1 Data Management Process | 4 |
| 6.2 Data Management Software | 5 |
| 6.3 Coding CRF Responses | 6 |
| 6.4 Data Entry | 6 |
| 6.5 Double and Single Data Entry with Control Checks | 7 |
| 6.6 Data Cleaning and Validation | 7 |
| 6.7 Data Protection | 9 |
| 6.8 Data Processing SOP | 9 |
| 6.9 Data Backup Systems | 10 |
| 6.10 Independent Data Monitoring Committee | 10 |
| 7. References | 11 |
| 8. Appendices | 11 |


© St George's

## 1. Background

An essential element of conducting a clinical trial is efficient data collection and management. Only that data that is essential for the purposes of the study should be collected. It is advisable for investigators to seek advice from a trial statistician as early as possible in the trial design process to facilitate this.

ICH GCP Guidelines specify that appropriately qualified individuals should supervise the trial data handling, verify the data and conduct the statistical analyses (ICH GCP 5.5).

## 2. Joint Research and Enterprise Office (JREO) Policy

All JREO SOPs will be produced and approved in accordance with the JREO SOP on SOPs and must be used in conjunction with local NHS Trust and St George's policies and procedures.

The JREO acts as the representative of both St George's University of London (SGUL) and St George's University Hospitals NHS Foundation Trust (SGHT). St George's will be the official name used on all SOPs to represent either institution acting as Sponsor.

## 3. Scope

This SOP describes the process for data management for St George's Sponsored clinical studies. Specifically the processes involved with collecting, validating and analysing such data. This SOP describes the full data management process including: data entry; data cleaning; and resolving data queries. This SOP also describes the use of Data Monitoring Committees for assessing data during interim analyses, and how such a committee should operate.

## 4. Definitions

### 4.1 Independent Data Monitoring Committee

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


© St George's

### 4.2 Interim Analysis

An interrogation of intermediate results or study related data collected and evaluation based upon aims or objectives to be achieved during the course of the trial. The interim analysis may be predetermined at a specific time point of the study, e.g. after nn patients have completed 2 cycles of treatment.

### 4.3 Trial Steering Committee

The role of the TSC is to provide overall supervision for a trial on behalf of the Sponsor and Funder and to ensure that the trial is conducted to the rigorous standards set out in Good Clinical Practice. It does not handle the day to day management of the study. Its main focus is the progress of the trial, adherence to the protocol, patient safety and the consideration of any new relevant information.

### 4.4 Case Report Form (CRF)

A Case Report Form is a paper or electronic data collection tool specifically used in clinical trials. There is usually a case report form for each study visit, designed specifically to capture the information from that visit.. The Case Report Form should enable data to be collected and captured to facilitate the research question(s) posed in the Research protocol to be analysed and answered.

## 5. Responsibilities

This SOP should be read and understood by the Chief Investigator and all those in the research team that will be responsible for data management in order to understand their responsibilities in relation to the management of data during the lifecycle of the study.

## 6. Procedure

### 6.1 Data Management Process

The process of data management involves converting the data collected using data collection tools, most commonly Case Report Forms (CRFs), into electronic data that can then be statistically analysed. This SOP will need to be read and utilised according to the size and complexity of the specific trial being conducted, as smaller trials may not require full use of the processes described.


© St George's

## 6.2 Data Management Software

Once the CRF has been designed in accordance with the protocol; the database to store the information collected should be designed. If an electronic data capture (EDC) system is used, it is important to ensure that the database has been validated and has been deemed fit for purpose through User Acceptance Testing (UAT). Depending on the size and type of study this database could be a standard spread sheet, or a more technical Data Management System may be required. When developing a database, points to consider include:

- ease of setting up and maintaining data entry screens;
- the ability for more than one user to use the system at the same time; and
- the ability to store and retrieve all data required for the study efficiently.
- A full audit trail

A useful resource that SGUL investigators should access together with this SOP is http://library.sgul.ac.uk/researchers/research-data-management

The Chief Investigator (CI) should decide whether to have the database developed in-house or whether to purchase a system. If a decision is made to purchase a system, the CI must ensure that they speak with Procurement to ensure the system is fit for purpose. (Refer to JREOSOP0030 Contracts).

Under ICH GCP there should be a specific SOP for managing the study database in place. The database should allow changes to be made to the data in a documented manner, and should not delete the original data entry to ensure an audit trail for the data is maintained (ICH GCP 5.5.3). The database should be secure, with appropriate password-protected access to prevent unauthorised access to the data, with a list identifying those individuals permitted to make changes to the data. ICH GCP also requires that there is adequate backup for the data, and that if there is blinding involved in the study, that the data entry and processing systems allow this to be maintained (ICH GCP 5.5.3).

For clinical trials involving Investigation Medicinal Products (CTIMPs) there must be an automatic electronically generated audit trail, rather than a handwritten log of


© St George's

database modifications. There are many electronic data capture systems (EDC) that can be purchased and modified to suit a particular study.

### 6.3 Coding CRF Responses

Before data entry to the database, the responses from the CRFs must 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. codes 0, 1 for Yes, No. Codes should also be in place for answers such as 'not known' or 'not applicable' e.g. 999 to show missing data. It is important to make sure that whatever value is chosen to represent missing data, that value would be unfeasible as an actual response, in order to prevent errors.

Clinical data also needs to be coded for recording of all adverse events. The World Health Organisation Adverse Reaction Terminology (WHO-ART) and Medical Dictionary for Regulation Activities (MedDRA) both have a system of coding to assist with this categorised by System Organ Class. A code is assigned for each disease and adverse reaction. Investigators can access WHO-ART and MedDRA through http://bioportal.bioontology.org/ontologies/WHO-ART?p=classes&conceptid=root

and www.medra.org respectively.

The coding can be done at various stages of the trial such as: during the initial data collection from the participant by the investigator or research nurse; after the data collection, but prior to entering the data on the database; or after data has been entered onto the database.

### 6.4 Data Entry

On initial receipt of CRF, the form should be date stamped and checked for initial missing or incomplete responses. If any inconsistencies are found these should be queried with the investigator and a record or log should be kept of all queries sent out. Instructions for sites to respond to data query responses should include no use of correction fluid, not to obscure the original data entry and to initial and date any amendments made.

Once the paper CRFs have been completed, the data must then be entered onto the database. Data entry should be performed by trained data entry staff. For multicentre studies where the CRFs are being sent to a coordinating centre for data entry, a copy of the CRF should be retained by the Investigator, with the originals (usually 2 copies from NCR (No Carbon Required) Paper CRFs) going to the coordinating centre. If NCR is used, it should be checked to ensure readability. If in doubt, the NCR should be


© St George's

photocopied as the data recording will fade over time. The coordinating centre must keep a log of all CRFs received, maintained by the Data Manager for the study.

All stored CRFs should be kept in a secure environment such as a locked filing cabinet in a locked room. Secure also means protection against environmental damage such as damp or fire, without water sprinklers.

### 6.5 Double and Single Data Entry with Control Checks

During data entry by trained staff, an average error rate of 5% is expected. Two methods can be used to reduce the risk of errors: Double Data Entry or Single Data Entry with control checks.

**a. Double Data Entry**

This method involves two people entering the same CRF data onto the database independently of each other. Depending on the software used, the data may be entered twice onto 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.

**b. Single Data Entry with Control Checks**

This method may be more suitable for smaller single centre studies with fewer staff available for data entry and/or less sophisticated database software. Once the data has been entered, a visual check can be done between what is recorded on the paper CRF, and what was entered on screen. One person should enter the data and a different person should do the visual check.

### 6.6 Data Cleaning and Validation

An integral part of the data management process is validation; to ensure the most accurate 'clean' set of data is provided for the statistical analysis. Data validation can be carried out at three stages during the trial:

**a. When CRFs are completed by the investigator**

To improve accuracy at this stage all staff completing CRFs should be sufficiently trained in their completion. Validation should also be carried out as part of the on-going monitoring of the study, either by members of the research team or by


© St George's

independent monitors (see the JREO SOP on monitoring JREOSOP0009). Validation via monitoring is done through Source Data Verification (SDV). SDV involves checking the data entered into the CRFs against that in the original source records e.g. patient's hospital files for accuracy.

**b. When data are entered in the database by data entry staff**

During data entry the two methods for validation described above can be used e.g. single data entry. Where data entry checks are used, if the study database has software enabled for automatic data entry checks, an Edit Check Specification (ECS) document should be put together by the clinicians/statisticians/data staff involved with the study. The ECS should provide full details of the data entry checks that have been set up, and all checks should be tested before the trial begins. This is done by User Acceptance Testing (UAT)

Depending on the database software, it is also advisable to set up warnings to alert data entry staff when values are entered outside of the expected range, or if the type of value entered is incorrect e.g. a numeric value entered rather than text. It is also useful to set up alerts for missing values where possible.

**c. When data have been entered and are available for the data manager**

At this stage it is advisable to carry out systematic post-entry computer tests. Lists should then be created (either through automatic database software system, or manually) of the following data queries:

- All missing values will 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 differences between fields.

All checks should be defined before the study 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.


© St George's

### 6.7 Data Protection

During the entire data management and validation process it is essential that all study data are kept in a secure location and in accordance with the terms of the Data Protection Act 1998. Participant confidentiality must be maintained at all times and all study records should be kept in pseudonymised form identifying participants by their study code rather than name, initials or hospital number.

Any paper CRFs should be kept in locked filing cabinets in locked rooms only accessible by authorised personnel. The key to the participant code list should be kept separately to these documents, again in a locked, secure location. If paper CRFs must be transferred to a coordinating centre for data entry, they should be sent either by courier or registered post to minimise the risk of losing data. A log should always be maintained of documents sent and received at each centre involved. If electronic data transfer is used, this should be via a secure system, password protected and encrypted where possible. At **no stage**, unless **specifically** consented to by the participant, should patient identifiable data be sent out of the originating NHS trust.

The database itself should be password protected, with each data entry staff member having their own password. If data entry is performed at the investigator site it is essential that the investigator does not have access to the whole database, to protect against biases occurring due to investigators making decisions based on interim data.

Any data that is stored on SGUL networked computers must be stored in an anonymised form with no identifiable information.

For further guidance on data protection please refer to the SGUL or SGHT Data Protection Policies (see references)

### 6.8 Data Processing SOP

Before the study starts, it is essential that a Standard Operating Procedure for data processing, management, and validation is put together, and updated as necessary throughout the study. The SOP should contain information on the following:

i. Contact details for all study staff
ii. Details of the flow of data from the investigator site to archiving
iii. Procedures on how to complete the CRFs
iv. Monitoring plans e.g. frequency, how Source Data Verification will be done, expected ranges for data values
v. Data Entry


© St George's

a. How to use the data entry system
b. Double or single entry
c. Roles and responsibilities of study staff with regard to data
d. Procedures in case of discrepancies

vi. Details of edit checks

vii. Description of Post Data Entry Validation System
a. Who checks the consistency of the data?
b. Who queries the Investigator?
c. What is the format of the query form?
d. How many days are allowed to answer a query?
e. Who decides that a query is resolved?

viii. Data Protection procedures, including a back-up system

Although the above list is not exhaustive it provides a basis for the Data Management SOP that can be adapted and expanded as necessary. Refer to http://library.sgul.ac.uk/researchers/research-data-management for guidance and resource.

### 6.9 Data Backup Systems

Whatever the format of the database software used to manage the study data, there should always be a back-up system in place to guard against loss of data due to software or environmental disasters. ICT may be able to advise whether the proposed backs up plans are fit for purpose.

### 6.10 Independent Data Monitoring Committees (IDMCs)

It is recommended for large, complex trials that an Independent Data Monitoring Committee (IDMC) is set up to carry out reviews of trial data at staged intervals during the study. The data reviewed by the monitoring committee 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 sufficient quality. The membership of the committee should include experienced trial investigators, statisticians and clinicians; all of whom must be independent to the research team. The results should be reviewed at regular intervals as sufficient data accumulate. If there is a Trial Steering Committee (TSC) for the study, the IDMC would normally make their recommendations for action through them.

## 7. References

International Conference on Harmonisation (ICH) of Good Clinical Practice.


© St George's

Data Protection Act 1998.

Trust Data Protection Policy

http://stginet/Procedural%20Documents/Corporate%20-%20IT/Information Governance and Management/org%20 1 4 2 July 07%20.pdf

JREOSOP0009 Routine monitoring on CTIMPs

http://library.sgul.ac.uk/researchers/research-data-management

## 8. Appendices

None associated with this SOP.


© St George's