South Eastern Health and Social Care Trust

Standard Operating Procedure
(SOP)
Research and Development Office

Title of SOP:	Data Management for Research Projects / Clinical Trials
SOP Number:	9
Version Number:	2.0
Supercedes:	1.0
Effective date:	August 2013
Review date:	August 2015

Author:	Alison Murphy, Research Manager Endorsed by Paul Carlin
Approved by:	Dr David Hill
Signed:	
Date:	01 August 2013

Version History:

Version No.	Date	Author	Reason for Change
1.0	Oct 2007	Paul Carlin	N/A
2.0	July 2011	Paul Carlin	Regional SOP

Table of Contents

1. INTRODUCTION

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

ICH GCP Guidelines specify that appropriately qualified individuals should supervise the trial data handling, verify the data and conduct the statistical analyses (ICH 5.5). This SOP will not describe procedures for developing the Case Report Form data collection tool.

2. OBJECTIVE

This SOP describes the data management processes for South Eastern Health and Social Care Trust (SEHSCT) sponsored clinical trials and other research projects, specifically the processes involved with collecting, validating and analysing such data.

3. SCOPE

This SOP applies to all clinical trials and other research projects sponsored by the SEHSCT. Where the data management has been outsourced to a third party, such as a Clinical Trials Unit, it may be appropriate to follow their SOP for data management.

This SOP does not apply to commercially funded research or research sponsored by an external non-commercial organisations.

4. PROCEDURE

4.1 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.

4.1.1 Data Management Plan

Before the study starts it is essential that a Data Management Plan for data processing; management; and validation is put together; and updated as necessary throughout the study. The SOP may 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
vi. Data Entry
    a. How to use the data entry system
    b. Double or single entry
    c. Roles and responsibilities of study staff with regard to data
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 Plan that can be adapted and expanded as necessary.

4.1.2 Data Management Software

Once the CRF has been designed in accordance with the protocol; the database to store the information collected should be designed. Depending on the size and type of study a MS Excel spreadsheet may be sufficient or an alternative Data Management System may be used. 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.

The database should allow changes to be made to the data in a documented manner, and should not delete data entry without recording this to ensure an audit trail for the data is maintained (IH 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 blinding is 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 (IMPs) an automatic electronically generated audit trial, rather than a handwritten log of database modifications, is preferred.

4.2 Coding CRF Responses

Before data entry to the database, the responses from the CRFs may 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, 2 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.

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 Regulations 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. You can access WHO-ART and MedDRA through http://www.umc-products.com/graphics/3149.pdf and http://meddramsso.com/MSSOWeb/index.htm 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 when the data are entered onto the database.

4.3 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 should be kept of all queries sent out. Instructions for sites to respond to data queries should include no use of Tippex, not to obscure the original data entry and to initial and date any amendments made.

Once the paper CRFs are completed, the data must then be entered onto the database. Trained data entry staff should do data entry. 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 original (usually 2 copies from No Carbon Required Paper CRFs) going to the coordinating centre. The coordinating centre should keep a log of all CRFs received.

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.

4.3.1 Double and Single Data Entry with Control Checks

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

i. 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.

ii. Single Data Entry with Control Checks

This method may be more suitable for smaller single centre studies with less 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. A record of all errors detected and changes made should be maintained.

4.4 Data Cleaning and Validation

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

i. When CRFs are completed by the investigator

To improve accuracy at this stage all staff completing CRFs should be sufficiently trained in their completion. A CRF completion manual would assist with this. Validation should also be carried out as part of the on going monitoring of the study, whether by members of the research team or by independent monitors. 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.

ii. When data are entered in the database by data entry staff

During data entry the two methods for validation described above (4.3.1) can be used i.e. data entry checks or double 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.

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 entered data is incorrect e.g. a numeric value entered rather than text. It is also useful to set up alerts for missing values where possible.

iii. 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 systems, 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 e.g. male and pregnant.

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.

4.5 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.

4.6 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 disaster. The SEHSCT IT service has a data backup service that provides a reliable means of protecting data held on departmental and research groups file servers. IT does not backup files on local desktop machines. Owners of such machines are responsible for protecting local files.

4.7 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 role of the IDMC is to view interim 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.

The data reviewed by the IDMC should be as up to date as possible and should be validated up to the point of interim analysis to ensure it is of sufficient quality. The membership of the committee should 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 or alternatively to the Chief Investigator (CI).

5. REGULATIONS, GUIDELINES, REFERENCES, SOP LINKS etc.

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

Data Protection Act 1998.

SOP05 Case Report Forms

SOP07 Computerised Systems for Clinical Trials

6. APPENDICES

None