

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

Researchers and their teams may print off this document for training and reference purposes but are responsible for regularly checking the RGIT website for more recent versions

<h1>Data Management</h1>	
SOP Reference: RGIT_SOP_020	
Version Number: 10.0	
Effective Date: 19 Oct 2020	Review by: 19 Oct 2023
Author: Funke, Erejuwa, Clinical Trials Monitor	
Approved by: Ruth Nicholson Head of Research and Integrity	Date:

Version	Date	Reason for Change
Version 1.0	31 May 2007	Change of office name from CRO to CRGO
Version 2.0	01 Sep 2007	Annual review
Version 3.0	20 Jun 2008	Annual review
Version 4.0	08 Feb 2010	Formation of Joint Research Office
Version 5.0	14 Jul 2011	Annual Review
Version 6.0	30 Nov 2012	Annual Review
Version 7.0	18 Feb 2015	Scheduled Review
Version 8.0	25 Oct 2017	Scheduled Review
Version 9.0	21 Jan 2020	Updated Data Protection Act from 1998 to 2018
Version 10.0	19 Oct 2020	Scheduled Review Templates removed and administrative changes to SOP. JRGO name change to RGIT.

Table of Contents

1.	PURPOSE.....	3
2.	INTRODUCTION.....	3
3.	PROCEDURE.....	3
3.1.	Data Management Process.....	3
3.2.	Coding CRF Responses.....	4
3.3.	Data Entry.....	4
3.4.	Data Cleaning and Validation.....	5
3.5.	Data Protection.....	6
3.6.	Data Processing SOP.....	7
3.7.	Data Backup Systems.....	7
3.8.	Independent Data Monitoring Committees (IDMCs).....	7
3.9.	Archiving.....	8
4.	REFERENCES.....	8

1. PURPOSE

This SOP describes the process for data management for Imperial College Academic Health Science Centre (AHSC) sponsored clinical trials. Specifically, the processes involved with collecting, validating, analysing and archiving such data.

2. INTRODUCTION

An essential element of conducting a clinical trial is efficient data collection and management. Only data that is essential for the purposes of the study should be collected as stated in the clinical trial protocol. 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 (DMCs) for assessing data during interim analyses, and how such a committee should operate.

ICH GCP Guidelines E6 R2 section 5.5.1 states “The sponsor should utilize 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.” This SOP will not describe procedures for developing the Case Report Form (CRF)/ data collection tool, as this is covered in the Imperial College AHSC Case Report Form SOP (RGIT_SOP_007).

3. PROCEDURE

3.1. 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 amended according to the size and complexity of the trial being conducted, as smaller trials may not require full use of the processes described.

3.1.1 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 this database could be a standard spreadsheet, or a more technical Data Management System (DMS) 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.

Under ICH GCP E6 R2 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 data entry to ensure an audit trail for the data is maintained (ICH GCP E6 R2 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 E6 R2 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 E6 R2 5.5.3).

For clinical trials involving Investigation Medicinal Products (IMPs) all AHSC sponsored trials must use the InForm electronic data capture system as their CRF unless data management and monitoring has been formally contracted to an external organisation with robust data management systems. Information about the system can be found at:

[InForm - Clinical Trials Unit](#) (*cited on 02nd June 2020*)

3.2. Coding CRF Responses

Before data entry to the database, the responses from the paper 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 and 1 for Yes and No respectively. 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 (AEs). 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. You can access WHO-ART and MedDRA through [WHODrug Portfolio](#) and [Medra](#) respectively (*cited on 02nd June 2020*).

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.

3.3. Data Entry

The study specific delegation log should specify which individuals are authorized to make data changes to study CRFs and the database respectively.

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 query responses 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. Data entry should be done 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 No Carbon Required Paper CRFs) going to the coordinating centre. 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 sprinkler. This should include checking that there are no water sprinklers above.

3.3.1. 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 Data 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 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.

3.4. 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. A CRF completion manual would assist with this (see RGIT_SOP_007). 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 independent monitors (see RGIT_SOP_015). 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 (3. 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 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:

- o 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.

3.5. 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 2018. 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.

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 Imperial College London networked computers must be stored in an anonymised form with no identifiable information.

Any data processing activity that could result in high risk for data subjects are required under GDPR to conduct a data protection impact assessments (DPIAs) in order to identify and minimise data protection risk.

To facilitate compliance, the Faculty of Medicine has developed a Research Data Privacy Impact Assessment (DPIA) Tool based in SharePoint. The Faculty mandates a completion of this tool for research projects that handle health and social care data.

For more information please visit: <https://www.imperial.ac.uk/admin-services/secretariat/information-governance/data-protection/data-assessments/>

Where any data is stored on a database supported by a web application, please see the College Database Management Systems policy for further information on special [Data Protection Act](#) requirements for such systems.

For further guidance on data protection please refer to the College Data Protection Policy available at: [Imperial College London - Data Protection](#)

3.6. Data Management 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
 - 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.

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

For servers that are used to exclusively support research data there is a charge for this backup service [Imperial - File recovery and backup link](#)

3.8. 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 review 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 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. A DMC Charter must be established to monitor members of the committee. The results should be reviewed at

regular intervals as sufficient data accumulates. Plans to establish a DMC must be described in the protocol.

If there is a Trial Steering Committee (TSC) for the study, the IDMC would normally make their recommendations for action through them (see RGIT_SOP_015).

3.9. Archiving

Chief Investigators are required to make adequate arrangements for the archiving of all study data (including paper CRFs) and essential documents. This service is provided by Imperial College Corporate Records Unit. Please refer to RGIT_SOP_019 Archiving for further information regarding archiving, this SOP which can be found on the [SOP, Associated Documents & Templates page](#). Chief investigators must also make themselves aware of where the study database is to be archived and by who; for Imperial College sponsored CTIMPs this will be arranged by the InForm database team.

4. REFERENCES

[Guideline on Data Monitoring Committees](#)

[DAMOCLES study group: A proposed charter for clinical trial data monitoring committees: helping them to do their job well](#)

[International Conference on Harmonisation \(ICH\) of Good Clinical Practice E6 R2](#) (cited on 19th August 2020)

RGIT_SOP_007_Case Report Forms

Data Protection Act 2018.

[Imperial College Data Protection Policy](#) (cited on 02nd June 2020)

RGIT_SOP_015_Monitoring

RGIT_SOP_019_Archiving