Standard Operating Procedure:

Data Management

# Purpose

This standard operating procedure (SOP) describes the data management procedures that need to be followed for a clinical research project.

# Scope

The SOP is applicable to all clinical research sponsored by the University of Birmingham (UoB). Where clinical research is sponsored by another institution, this procedure should be followed as far as possible, and in line with the contractual agreement between the UoB and the other institution. This SOP also applies to clinical research approved by a UoB Research Ethics Committee (REC) that are required to follow the UoB Principles of Good Clinical Practice (GCP) for Clinical Research (UoB-GCP-POL-001). This SOP may be used as a guidance document in all other cases.

# Implementation plan

This SOP will be implemented in line with this document’s effective date for all clinical trials, and any clinical studies that are in set up. For existing clinical research that is already set up and in the recruitment phase (or further) at the time of implementation, this SOP will be implemented within three months of this document’s effective date.

# Stakeholders

Note that where the UoB takes on the sponsor’s responsibility for data management, the UoB will delegate the majority of these duties to the chief investigator (CI) or to a clinical trials unit (CTU), who may delegate these duties further to their trials team(s). All delegation of duties will be documented (e.g., using the CI declaration and/or the Clinical Trials Task Delegation Log (UoB-SPO-QCD-001)).

* CI: the CI may delegate activities to members of their trial team, although evidence of CI involvement and approval is still expected and may not be delegated where ‘no delegation allowed’ is indicated. The SOP will state where delegation is possible. For clinical research approved by UoB REC, the role of CI may be referred to as the principal investigator (PI), or the supervisor for postgraduate research students.

# Background and rationale

For the purposes of this SOP the terms ‘clinical research’, ‘research project’ or ‘project’ will cover clinical trials of investigational medicinal products (CTIMPs), other interventional trials (e.g. surgical trials, device trials and non-CTIMPs, and any other projects deemed to be ‘interventional’ by the Sponsor) and clinical studies.

The data generated from a clinical research project play a fundamental role in determining the outcome of the research and subsequently the direction of future research/treatments. It is vital that the data-management processes are robust enough to ensure that the data being collected and reported is of the highest standards, accurately reflecting the research being carried out. Data-management processes are required to ensure that during the trial the principles of ALCOA-CCEA (see below) are followed. This will ultimately lead to the use of accurate data in the final report or publication.

Clinical study/trial systems play a pivotal role in managing the data generated and are especially beneficial when dealing with large amounts of data from multiple sites. The complexity of the project and the use of bespoke or off-the-shelf software can vary, therefore careful consideration of the role of the clinical study/ trial system(s) and feasibility are required.

Data-management processes must adhere to the UoB Principles of GCP for Clinical Research (UoB-GCP-POL-001) Principle 21: ‘All information about treatment, care or other services provided as part of the research project and their outcomes is recorded, handled and stored appropriately and in such a way and for such time that it can be understood, where relevant, by others involved in the participant’s care and accurately reported, interpreted and verified, while the confidentiality of records of the participants remains protected, where appropriate.

# Procedure

This SOP describes the requirements relating to data management. For the purposes of the SOP, the aspects of study/trial systems management have been grouped together.

1. During project set up, the CI (or delegate) will decide on and document the data-management processes, to include the process of data collection, data clarification, data location and transfer, quality check procedures, and decide the need for a study/trial system.
* The data management processes will be set up to ensure the ALCOA-CCEA principles relating to data integrity are met, these being:
* attributable - data collection/amending is attributable to the person performing the task
* legible - data recorded is legible and permanent
* contemporaneous - data collection is contemporaneous with the procedure
* original - original records or certified copies are available
* accurate - data being acquired is accurate and not erroneous
* complete - data being collected is without omission and any changes made to the already collected data is traceable
* consistent - data collection is in sequence to the trial events
* enduring - relevant data is recorded securely on the appropriate trial-controlled tools (e.g. not on a post-it)
* available - data is available when required and throughout the archiving period (e.g. accessible for review/audit).
1. The CI (or delegate) will ensure that the data management processes is documented. The process may be documented in the protocol (see either Protocol Template - CTIMP (UoB-CLN-PRO-QCD-002) or Protocol Template - Non-CTIMP & Studies (UoB-CRT-NCTM-QCD-001)). For more complex projects, it is recommended that a data management plan be created, based on the project risk assessment.
2. The CI (or delegate) will confirm their agreement with the data management processes in writing (e.g. through CI approval of the protocol or data management plan). Any subsequent changes to the data management process deemed significant will follow the same review/approval process.
3. The CI (or delegate) will consider the need for defining a list of critical data items. Critical data items are information that is essential for the successful analysis of the endpoints of a project.
4. Where a project requires data to be provided via NHS Digital, the CI (or delegate) will work with the [Research Governance Team (RGT)](https://intranet.birmingham.ac.uk/finance/rss/ethics-and-governance/research-governance/index.aspx) to ensure:
* appropriate registration of the project
* the [UoB information security training](https://intranet.birmingham.ac.uk/it/security/training.aspx) has been completed by the research team
* a data sharing agreement (DSA) is in place and appropriately authorised.
* It is important that this data is curated in line with the requirements of the DSA and when this expires, a destruction certificate (facilitated and authorised by [UoB IT Services](https://universityofbirmingham.service-now.com/itportal/)) is provided to NHS Digital.

### Develop data management tools

1. The CI (or delegate) will develop tools to help facilitate and evidence data management processes (e.g. data collection tools, databases, data clarification forms (DCFs)). They will ensure that the following is covered.
* Collecting data:
* data collection tools may be paper (e.g. case report forms (CRFs), see the CRF development SOP (UoB-CRT-CRF-SOP-001) for more information) or electronic (e.g. electronic CRFs or computer applications)
* consideration should also be made to the storage location of data, who has access to it and what security is in place, and that this is in line with what participants have provided consent for (see the Participant Engagement and Informed Consent SOP (UoB-PEI-SOP-001)). This is of particular importance if identifiable personal information is being collected. Particular care should be used when using electronic data collection tools, where data may be stored off site (e.g. on a third party server or cloud based system).
* Tracking the movement of data:
* the movement of any data is compliant with (but not limited) the [Data Protection Act (DPA)](https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted), [Caldicott Principles](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/942217/Eight_Caldicott_Principles_08.12.20.pdf), [Health Research Authority (HRA) guidance](https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/) and applicable UoB policies (see [Data Protection Resources](https://intranet.birmingham.ac.uk/executive-support/legal-services/what-we-do/data-protection/data-protection-resources.aspx) and [UoB IT Policies, Standards and Guidance](https://bham.sharepoint.com/sites/IT/SitePages/Policies-and-procedures.aspx))
* the transfer and receipt of any data is appropriately logged to avoid data being misplaced (e.g. if CRFs are to be posted to the coordinating centre, a log of receipt is completed on its arrival).
* Amending and clarifying collected data:
* the coordinating centre (where applicable) does not amend the data received from the sites unless it has been classed as a self-evident correction (see below for information on self-evident corrections) or the site has confirmed in writing that data should be amended
* appropriate records are kept when dealing with any data queries (e.g. the use of DCFs)
* any changes made to the research data already collected are clearly identified (i.e. an audit trail is present in electronic systems showing the changes and changes in paper systems should not obscure the original entry)
* the data records kept at the site match the records kept at the coordinating centre including any updates, e.g. through DCFs.

Note: The number and type of tools required can vary depending on the complexity of the research project, the quantity of data being collected and the movement of data. As a minimum, it is expected for there to be a data collection tool.

1. The CI (or delegate) will consider what will be defined as source data.
* For CTIMPs, the definitions of source data will be documented, e.g. in the protocol, a source data agreement, or at site initiations.
* For all other research, documenting the definition of source data will be considered.
1. Prior to the start of data gathering, the CI (or delegate) in association with PIs (where applicable) will specify and document what is considered as acceptable self-evident corrections.
* Self-evident corrections will be kept to a minimum, and where a correction is made the appropriate site (where applicable) will be informed.
* Where further self-evident corrections need to be defined after recruitment has begun, the CI (no delegation) in association with PIs (where applicable) will review and approve these before they are applied.
* Examples of self-evident corrections could include:
* at the start of a New Year where the month January is correctly recorded but the previous year is obviously incorrectly quoted
* where data is recorded on the CRF but in the wrong location. For example, the subject’s date of death is written in a comments box on the death form but not provided in response to the question “Date of Death”
* the data collection form lists concomitant medications taken by a patient but the box stating, ‘Are there any medications this cycle?’ is left blank.
1. The CI (or delegate) will define and document the plan for performing quality checks on data and validation checks on the project database. The CI (or delegate) will approve the data quality check/data validation plans prior to the beginning of the data management process. The CI (or delegate) will ensure the following is covered as part of quality check/validation process:
* quality check plans are in line with the requirements outlined in the risk assessment (where applicable)
* data systems are appropriately validated before use and data validation checks are being performed throughout the project
* the need to include version control on the data validation specification and to review the specification following protocol/CRF amendments
* the need to monitor that data gathering is contemporaneous, and any queries are addressed in a timely manner
* a description of where data validation ends, and the data is finalised and provided for statistical analysis
* the need for deviations of the data management process, including the above, to be dealt with appropriately and resolved in a timely manner.
* Where project-specific issues are significant, involvement from relevant oversight committees (e.g. Study/Trial Management Group, Study/Trial Steering Committee, Data Monitoring Committee) may be required. See also Project Oversight & Quality Management SOP (UoB-POS-SOP-001).
1. The CI (or delegate) will ensure that staff involved in managing and processing the data are appropriately trained and informed of their roles before undertaking their respective tasks.
* An individual’s access to the research data will be in line with the activities the individual is to undertake (e.g. only observational, full editorial access)

### Set up and maintain study/trial systems

1. The CI (or delegate) will decide the need for setting up a study/trial system, and the complexity of the system required to support the research.
2. The CI (or delegate) will define the process, ensuring each individual section of the system is fit for purpose (validated) and has been successfully tested before being implemented in the project. Typically, this would involve an outline of the requirements of the system and user requirement specification. It is expected for there to be documented evidence of approval from the CI (or delegate).
* For laboratory systems, see the Laboratory Analysis SOP (UoB-CRL-SOP-004) for further information.
1. The CI (or delegate) will consider the need for a system validation plan using a risk-based approach and document their conclusion.
* A system validation plan is expected for bespoke systems but may not be necessary where commercial off-the-shelf systems are used or have been configured to be specific for the project.
* A documented system validation plan is required for all laboratory systems. See the Laboratory Analysis SOP (UoB-CRL-SOP-004) for further information.
1. The CI (or delegate) will implement appropriate version control on the system, and on the system validation plan (where one is required).
2. The CI (or delegate) will ensure any changes to the project are appropriately incorporated into the system where relevant, and that this is suitably managed, appropriately version controlled and successfully tested before implementation.
3. The CI (or delegate) will ensure the system has appropriate back-up and security measures in place.
4. The CI (or delegate) will ensure that the study/trial system is recorded on the university’s [data asset register](https://asi.bham.ac.uk/data_asset_register), and amend the register accordingly should any amendments to the system be required.
* For login details and further guidance contact the Head of Data Governance in the [Strategic Planning and Performance Insight Team](https://intranet.birmingham.ac.uk/executive-support/planning/contact/index.aspx).
1. The CI (or delegate) will ensure relevant staff are appropriately trained in using the system prior to dealing with the research data and that any modifications to the data are being logged (i.e. there is an audit trail).
2. The CI (or delegate) will ensure only appropriate staff have access to the research data/files and that a record of this and the access is up to date and being maintained, including where access is revoked.
3. The CI (or delegate) will ensure that the system and data within is stored on a UoB server configured to UoB IT policies, and that back-ups and storage of data adheres to [UoB IT policies](https://bham.sharepoint.com/sites/IT/SitePages/Policies-and-procedures.aspx).
4. The CI (or delegate) will ensure the system is accessible throughout the research, including the archiving period (see the Archiving SOP (UoB-ARC-SOP-001) for more information).
5. The CI (or delegate) will ensure the process for storage and retention of data is defined and documented (see the Archiving SOP (UoB-ARC-SOP-001) for more information).
6. The CI (or delegate) will ensure the system functionality is documented and being validated throughout the project.

## Data collection and processing

1. The CI (or delegate) will ensure that appropriate data quality checks/data validation plans, and the data management processes are being followed throughout the project and that any deviations to the process are appropriately managed (see the Deviations and Serious Breach Reporting SOP (UoB-DSB-SOP-001) for further information on serious breaches).
2. The CI (or delegate) will ensure that where copies of research data are taken for analysis purposes, a copy is preserved so that it may be reproduced if required later.
3. Where interim analyses are to be performed on the data, the CI (or delegate) will ensure that a snapshot has been taken of the relevant part of the dataset before any further amendments are made.

## End of data collection

1. The CI (or delegate) will lock the data from any further changes once data collection has been completed to allow for statistical analyses to begin (see the Statistics SOP (UoB-STA-SOP-001)).
2. If the dataset needs to be unlocked due to significant changes needing to be made, the CI (or delegate) will ensure the following are documented and filed prior to the dataset being unlocked.
* Appropriate justification as to why the dataset has been unlocked.
* Any changes made will need to be in accordance with the approval and visible (i.e. original data is not permanently deleted).
* Written approval from the CI (no delegation) and the statistician (where applicable).
* Where relevant, the effect on the statistical outcome has been assessed by the statistician.
* It is also expected that the study report will include details of all relevant changes made to the database while it is unlocked.
1. The CI (or delegate) will ensure that the dataset is re-locked in a timely manner.

# List of expected outputs

* A documented record of the data management process describing data collection, data clarification, data transfer, quality check procedures and documented evidence of the process being followed.
* Evidence of a documented record of what is considered acceptable self-evident corrections with evidence of CI approval (and PI where applicable) being in place prior to the beginning of recruitment, and evidence that appropriate sites (where applicable) have been notified where a correction is made. Where self-evident corrections are added after recruitment has begun, evidence that the CI and PI (where applicable) have approved these before these being applied.
* Tools to facilitate and evidence the data management process. As a minimum, it is expected for a data collection tool to be present.
* Documented evidence of data quality check/validation plans, to include information on when data validation ends, and the database is finalised ready for statistical analyses and being approved by the CI (or delegate) prior to the beginning of the data management process and evidence of plans being followed.
* Evidence indicating staff have been informed/trained on their respective roles (e.g. training logs, written instructions).
* Where data collection has been completed, evidence showing that the dataset was locked to any further changes prior to the start of the final statistical analyses.
* Where the dataset needs to be unlocked, appropriate justification, written approval from the CI (no delegation) and the statistician (where applicable), with a documented assessment from the statistician on any effect to the statistical outcome prior to the dataset being unlocked (where relevant).

### Study/trial systems

* Evidence indicating the study/trial system is fit for purpose and has been tested appropriately before implementation, with appropriate version control applied.
* A version-controlled system validation plan and documented evidence of implementation across the duration of the project.
* Evidence indicating staff have been informed/trained on their respective roles prior to using the system (e.g. training logs, written instructions).
* Evidence showing that access to the system is appropriately restricted, and this is being maintained/up to date.
* Able to demonstrate that the system has been registered, and its record updated as required, on the university’s [data asset register](https://asi.bham.ac.uk/data_asset_register).
* Able to demonstrate that the system is stored on a UoB server configured to UoB IT policies.
* Where amendments have been made to the project: evidence that these have been incorporated in the system where relevant and that successful testing and that version control has been relevantly applied.
* Where copies of data have been taken for analysis purposes: evidence that a copy has been preserved.
* As part of project closure, evidence that the system has been stored on a UoB secure network drive for the stipulated duration and is retrievable.
* Evidence that the system has been stored as per the Archiving SOP (UoB-ARC-SOP-001).

# Related documents

* UoB-ARC-SOP-001 Archiving
* UoB-CLN-PRO-QCD-002 Protocol Template - CTIMP
* UoB-CRL-SOP-004 Laboratory Analysis
* UoB-CRT-CRF-SOP-001 CRF development
* UoB-CRT-NCTM-QCD-001 Protocol Template - Non-CTIMP & Studies
* UoB-CRT-STA-SOP-001 Statistics
* UoB-DSB-SOP-001 Deviations and Serious Breach Reporting
* UoB-GCP-POL-001 UoB Principles of GCP for Clinical Research
* UoB-PEI-SOP-001 Participant Engagement and Informed Consent
* UoB-POS-SOP-001 Project Oversight and Quality Management
* UoB-SPO-QCD-001 Clinical Trials Task Delegation Log

Note the UoB QMS documents can be found on the [CRCT website](https://www.birmingham.ac.uk/research/activity/mds/mds-rkto/governance/index.aspx). Internal work instructions can be obtained from the CRCT (crct@contacts.bham.ac.uk) and/or from the RGT (researchgovernance@contacts.bham.ac.uk).

# References and frameworks

* ALCOA-CCEA: <http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/08/WC500095754.pdf>
* Caldicott Principles: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/942217/Eight_Caldicott_Principles_08.12.20.pdf>
* Data Protection Act: <https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted>
* General Data Protection Regulation (GDPR): <https://gdpr-info.eu/>
* Good Clinical Practice Guide compiled by the Medicines and Healthcare products Regulatory Agency (MHRA). First edition published in 2012
* HRA guidance on GDPR: <https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/>
* ICH-GCP: <https://www.ich.org/page/efficacy-guidelines>
* UoB Information Security training: <https://intranet.birmingham.ac.uk/it/security/training.aspx>
* UoB IT Policies, Standards and Guidance: <https://bham.sharepoint.com/sites/IT/SitePages/Policies-and-procedures.aspx>
* UoB Strategic Planning and Performance Insight: <https://intranet.birmingham.ac.uk/executive-support/planning/contact/index.aspx>

# Abbreviations and definitions

| Term | Description |
| --- | --- |
| Case report form (CRF) | A printed, optical, or electronic document designed to record all of the protocol required information to be reported to the sponsor on each participant. |
| Critical data item | Information that is essential for the successful analysis of the primary and key secondary endpoints of a clinical research project. |
| Data | Facts, figures and statistics collected together for reference or analysis. |
| Data clarification | A data clarification is a query generated by the sponsor and sent to, and is completed by, the investigator site (or if applicable the participant) as part of the data validation process. |
| Data clarification form (DCF)  | A document used to formally record a data clarification.  |
| Data collection forms | All forms used to collect data for the research dataset. This may include CRF, DCF, Quality of Life (QoL) questionnaires, health economic measures, participant diaries and/or correspondence with the site. |
| Data management | [The administrative](http://www.businessdictionary.com/definition/administrative.html) [process](http://www.businessdictionary.com/definition/process.html) by which [data](http://www.businessdictionary.com/definition/data.html) are acquired, processed, cleaned for analysis (verified), stored, and protected to ensure that the results are accurate and collected in accordance with the protocol and Good Clinical Practice. |
| Data sharing agreement (DSA) | Data sharing agreements set out the purpose of the data sharing, cover what happens to the data at each stage, set standards and help all the parties involved in sharing to be clear about their roles and responsibilities. |
| Data validation | The process of checking data for such elements as logical inconsistency, protocol deviations (e.g. subject ineligibility) and missing/incorrect/implausible data.  |
| Self-evident correction | A correction to the CRF that can be made by the sponsor without the requirement for a referral of a data clarification to the investigator.  |
| Statistician | A qualified statistician who contributes to the design, analysis and interpretation throughout the clinical research life cycle and takes responsibility for the statistical aspects of the project. The statistician may be unblinded to the allocation and evolving results whilst performing interim analyses but will not disclose these to the rest of the research team. There may be a senior statistician supervising a more junior statistician. |
| Study/trial data set (database) | (Personal) study/trial (relational) database, Excel spread sheet, Word documents etc. used to store data for the purposes of analysis. |
| Study/trial system | The study/trial system describes the software and database used to store and manage clinical research project data used for the analysis of outcome measures as defined in the research protocol. This may include databases containing data, contact databases and data tracking systems. |
| System testing | The process of ensuring that the software/system and database meets the requirements outlined in the user requirement specification. |
| System validation | The process of ensuring that a program operates on clean, correct and useful data through documented design, build and testing. For more information in the appropriate levels of testing please refer to the MHRA Good Clinical Practice Guide, section 14.5.2. |
| User requirement specification (URS) | A document that contains the requirements for the trial system. |

See also the [Glossary of Terms](https://www.birmingham.ac.uk/research/activity/mds/mds-rkto/governance/Glossary-of-Terms.aspx).