Standard Operating Procedure

STUDY DATA

SETTING Trustwide

AUDIENCE All R&I and research staff involved in collecting, entering, checking, correcting, transferring and analysing data for UH Bristol sponsored trials

ISSUE This SOP relates to collecting, entering, checking, correcting, transferring and analysing data generated by UH Bristol sponsored trials.

QUERIES Contact Research & Innovation (R&I) department : Ext 20233 or research@uhbristol.nhs.uk

Document History

SOP number	SOP 012	**SOP Version**	1.3
Effective Date	21/FEB/2018	**Review Date**	21/FEB/2020

Review date	Version number	Version date	Effective date	Author/ Reviewer	Authorised by
-	V1.0	19/OCT/2015	03/NOV/2015	Diana Benton	Diana Benton
02/DEC/2015	V1.1	02/DEC/2015	23/DEC/2015	Jess Bisset	Diana Benton
November 2016	V1.2	02/DEC/2016	05/APR/2017	Jess Bisset	Diana Benton
25/JAN/2018	V1.3	25/JAN/2018	21/FEB/2018	Trusha Rajgor	Jess Bisset

Version Number	Reason for change
Original V1.0	N/A – original
V1.1	Minor update to clarify 'Protocol sign-off' in section 5.1
V1.2	Annual review – minor updates and clarifications.
V1.3	Annual review – minor updates and clarifications.

1. Introduction

In accordance with Good Clinical Practice (GCP) 'All clinical trial information shall be recorded, handled, and stored in a way that allows its accurate reporting, interpretation and verification' (Schedule 1, Part 2 (10) SI 2004/1031).

For clinical trials of investigational medicinal products, the need to be able to robustly defend the source of the data and the systems through which it passes until publication is paramount, and is underpinned by the law through the Medicines for Human Use (Clinical Trials) Regulations (SI/2004/1031) and any amendments. Similarly, robust systems to document the effects of investigational medicinal products on human subjects must be in place.

The standards with which to comply with are also referenced in the MHRA Grey Guide (2012).

The data generated through research may be used to influence or drive changes in clinical practice. Therefore the standards are in place to ensure that both robust data are generated and patients are safe.

2. Purpose

The purpose of this SOP is to describe the standards required for collection, entry, checking, correction, transfer and analysis of data generated by UH Bristol sponsored research.

3. Scope

In Scope: Data systems and processes for Clinical Trials of Investigational Medicinal Products sponsored by UHBristol. Data systems and processes for other research sponsored by UHBristol.

Out scope: Research sponsored by other organisations, hosted by UHBristol.

4. Responsibilities

The R&I department has a responsibility as sponsor to ensure that staff delivering UH Bristol sponsored research are fully aware of the required data management standards which must be complied with and must maintain a level of sponsor oversight which is proportionate to the level of risk.

All research staff undertaking UH Bristol sponsored research who process data are responsible for ensuring the applicable data management standards are met as described in this SOP, the applicable regulations and with GCP.

5. Abbreviations and Definitions

Abbreviations	
CI	Chief Investigator
CRF	Case record/report form
DMP	Data Management Plan
eCRF	Electronic Case record/report form
GCP	Good Clinical Practice
pCRF	Paper Case record/report form
RPM	Research projects Manager
SAE	Serious Adverse Event
TMF	Trial Master File

Definitions	
CRF	Document used to record the required data as defined by the protocol for each participant throughout their participation in the study
Data Management Plan	The main document that describes and defines all data management activities throughout the lifecycle of a research study
Data Validation	Checks on data quality
Data Verification	Checks on accuracy of data entered into a database

6. Procedure

6.1 Protocol

- For **all studies sponsored by UHBristol**, the protocol must clearly describe which data will be collected, at what time points, where it will be stored and how it will be used.
- The version of the protocol in use must be signed off by the sponsor. 'Sign-off' constitutes one or all of the following: electronic signature on the protocol; wet ink signature on the protocol; authorisation on the application for approval to the Research Ethics Committee Health Research Authority.

6.2 Data Management Plan

- For **UHBristol sponsored CTIMPs** a *Data Management Plan (DMP)* (*TMPL_041*) must be completed. This document will provide an overview of the data flow processes for the study as well as detail all of the data management activities for a study.
- The DMP is a live document that may change throughout the duration of a study. The first version should be sent to the Research Projects Manager (PRM) for review during study set up. The RPM will ensure that adequate data management is described and liaise with the Trial Manager to finalise the document. The DMP must be stored in the Trial Master File (TMF). Any amended versions will also need to be reviewed by the RPM prior to finalisation and all versions stored in the TMF.
- For **other interventional trials** a risk based decision should be made regarding the need for a DMP. The detail provided in each section should be proportionate to the size and complexity of the study. Refer to *TMPL_041 Data Management Plan*. If agreed by UH Bristol as sponsor, an alternative data management plan template may be used.

6.3 CRF

- The CRF should be designed to support collection of the data required by the protocol, and may be paper (pCRF) or electronic (eCRF). See *GD_007 Key elements to be considered in the design of a CRF.*
- Any verification of data which must be done by particular members of the research team (e.g. inclusion/exclusion criteria and safety data by medically qualified staff) must be evidenced. For paper CRFs, this would usually take the form of a signature and date; for eCRFs, this may be carried out by means of audit software incorporating particular logins, or could be documented separately within the source data.
- It is good practice to include a wide range of staff in CRF review. The CRF should be reviewed against the protocol to ensure all necessary information is captured and to ensure the design and flow of the questions allows good quality data collection, and so that the data tables that will be generated support the planned data analysis.
- The CRF and any amended versions must be signed off by the sponsor prior to implementation, or on a case by case basis, if agreed with the sponsor, by the CI.
- Original CRFs form part of the TMF, as an essential document. Provision must be made for sites to retain a copy of the CRF at site, independent of the sponsor, in order to ensure that the sponsor cannot manipulate the site data. This might be by using worksheets to collect data ('shadow' CRFs) when eCRFs are in use or providing copies of the data back to site on DVD/CD at the end of the trial.

6.4 Source Data

- Source data is the first place that a piece of information is recorded, prior to transcription into a CRF. Source data can take many forms and must remain at the location at which it was generated. Refer to *GD_008 Key elements to include in source data* for further information. Types of source data include handwritten and typed paper and electronic notes, clinical systems, hard copy or electronic images. On occasion, the CRF may act as the source.

- It should be possible to establish that the source information for data collected in the CRF existed at the appropriate point in time. That is, electronic systems used to record source data should have appropriate audit software providing the date, or be saved in a version controlled manner, and paper records should include a date and signature.

6.5 Authorising changes to the data

- There should be an agreed system and process in place for authorising changes to data. The CI should agree with the research team what changes are acceptable, and document this. For example, the CI may agree that any member of the research team can amend clear transcription errors where the source data have not been transcribed correctly into the paper CRF. Other items, such as medical assessments and safety data changes should only be authorised by the CI.

6.6 Database

- A database is a repository for electronic data.
- Databases vary widely, depending on the size, type and complexity of the research being carried out. For a simple, small study, an excel spreadsheet can be used, provided it meets the standards described in *SOP_011 Validation and Backup of Computer Systems*. At the other end of the spectrum are complex databases which have automated audit software and consistency checking capability, as well as the ability to generate data queries.
- A database must reflect the CRF so that the data required by the protocol can be collected. The chief investigator must check that the database meets the needs of the study by reviewing and testing it, and documenting that the database meets the required specifications (user acceptance testing).
- As more data is entered, or changes are made, it is important that an audit trail of the changes is available, so that previous versions of the datasets can be accessed if necessary. For sophisticated databases, the mechanism may be by using the database software to record changes to data fields and the associated logins that carried out the change(s); for a simpler database this might be by saving subsequent copies with a version number and date and a form of identification of the person who modified the file (e.g. initial and last name).

6.7 Data entry

- Entry of data into the database should be performed by fully competent staff that are appropriately qualified and have received any necessary training.
- Data entry may occur during or after each participants visit or at the end of the study.
- The data should be entered as recorded on the CRF without modification.

6.8 Quality control

- Systems to ensure that data entry is accurate or that errors are identified and corrected must be in place. This can be supported in a number of ways:
    - Dual data entry, and discrepancy checking of the two entries
    - Using consistency/logic checks to ensure expected answers are entered – e.g. blood results fall within expected reference ranges, ages match dates of birth, etc;
    - Source data verification - quality checks carried out against the source data in patient records and against CRF entries.
- If the mechanism of carrying out these checks is not automated, checks must be documented when they are completed. For automated checks, an audit trail must be available. Records of checks and audit trails must be retained as part of the essential documents.

- Data held on the database may be validated or monitored/reviewed centrally in order to ensure the data are complete and accurate. This can be done in a number of ways, including review of print-outs and comparison either across visits, or with paper CRFs to ensure consistency and accuracy. For more complex trials, this may be done using consistency/logic checks as described above. All checks must be documented fully. Any queries which are generated as a result of this process must be addressed with the PI (or delegate) in order to resolve the query and make changes to the data. These must be documented fully.

6.9 Reconciling clinical and pharmacovigilance databases

- If serious adverse events recorded are on a separate pharmacovigilance database they must be matched with the clinical research database. For long/complex studies this should be conducted throughout the lifecycle of the study; for simpler, short studies with low numbers of SAEs this can be carried out at the end of the study.
- As part of the DMP, an approach to reconciling the SAE data must be agreed at the start of a trial.

6.10 Data coding

- Data coding may be appropriate for larger trials. Plans for management of coding should be incorporated into the data management plan.

6.11 Final data quality

- It is good practice to check the final data quality of a database prior to lock. For small, low risk trials it might be appropriate to combine a number of checks which might have been conducted throughout the data management process into this final check. However, there is a risk in this approach that the time lapse since collection of the data makes clarification of any queries that arise more difficult. Any decision around the approach should be documented at the study start.
- A data quality check is an assessment of a proportion of the data, comparing it to the data in the paper CRF. That proportion may be 100% when it relates to primary endpoint data, or lower for other data. A threshold of 'acceptable' error rates should be agreed, below which further checks should be carried out until the quality is deemed to be acceptable. The process of identifying and resolving errors must be documented. For eCRFs the checks should have been completed via source data verification earlier in the process of data management, in accordance with the data management plan.

6.12 Locking and unlocking the database

- Database locking is the process by which it is declared and identified as final. No changes to the data should be made once the database has been locked, and arrangements should be put in place to control access to the data and protect it. The files should be protected from editing and deleting, and the decision about the approach to doing this should be made in a risk based way.

- Unlocking the database should take place only under exceptional circumstances, and requires due consideration by the sponsor and consultation with the statistician. Written approval for data unlocking, the justification, the changes that will be made and the impact on the analysis must be recorded in the trial master file prior to unlocking.

6.13 Release of the final database/datasets

- Data should be extracted securely from the locked database to carry out the final analysis. The process to do this should be adequately described documenting how the data will be

protected from alteration. Test extracts may be made, and these must be stored in a separate location to the extracted datasets on which the analysis will be performed.

7. Dissemination and training in the SOP

This SOP will be disseminated to applicable research staff (including R&I) and will be available on the R&I website.

All staff whose activities are subject to this SOP should ensure that they read and understand the content of the SOP. The personal training log of the individual (and the Investigator Site File/Trial Master File if required) should be completed to document that the content of this SOP has been read and understood as described in *SOP_ 007 Research Training*.

8. Related documents

- GD_007 CRF design: Key elements
- GD_008 Source data: Key elements to include, with dates
- SOP_007 Research Training
- SOP_011 Validation and Backup of Computer Systems
- TMPL_041 Data Management Plan