SOP22a: Standard Operating Procedure for Data Collection and Management

Authorship Team: Mel Storey and Stephen Allen for Joint SOP Group on Trial Techniques (viz Helen Snooks, Ian Russell, Bridget Wells, Frances Rapport, Julie Peconi, Moira Morgan, Simon Thompson, Sian Davies, Sinead Brophy, Steve Allen, Ceri Phillips)

Approved by WWORTH JMG (Ian Russell in chair)

Signature Date 07 May 2014

<table>
<thead>
<tr>
<th>Version Number</th>
<th>Effective Date</th>
<th>Reason for Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>SOP approved by NWORTH</td>
</tr>
<tr>
<td>0.1</td>
<td>04 Mar 2009</td>
<td>Adapted SOP discussed by JSOPG</td>
</tr>
<tr>
<td>0.2</td>
<td>01 Apr 2009</td>
<td>Reviewed by JSOPG</td>
</tr>
<tr>
<td>0.2</td>
<td>07 Jul 2009</td>
<td>Comments for review at JMG by KT</td>
</tr>
<tr>
<td>0.3</td>
<td>15 Jul 2009</td>
<td>Amendment to SOP post-JMG review</td>
</tr>
<tr>
<td>0.4</td>
<td>07 Aug 2009</td>
<td>Further amendments</td>
</tr>
<tr>
<td>0.5</td>
<td>15 Sep 2009</td>
<td>Amendments to SOP following 2nd JMG review for final approval by KT</td>
</tr>
<tr>
<td>1.0</td>
<td>16 Sept 2009</td>
<td>Approved in principle</td>
</tr>
<tr>
<td>1.1</td>
<td>15 May 2010</td>
<td>Minor formatting amendments</td>
</tr>
<tr>
<td>1.2</td>
<td>19 Apr 2012</td>
<td>Minor formatting amendments, addition of training section. Prepare SOP for JSOPG by MS. Egs in use CONSTRUCT, SAFER2, PRISMATIC, PROBAT</td>
</tr>
<tr>
<td>1.3</td>
<td>21 May 2012</td>
<td>Minor amendments following JSOPG</td>
</tr>
<tr>
<td>2.0</td>
<td>31 May 2012</td>
<td>Authorised for use by JMG</td>
</tr>
<tr>
<td>2.1</td>
<td>06 Jun 2013</td>
<td>Update header and formatting amendments</td>
</tr>
<tr>
<td>2.2</td>
<td>15 Jan 2014</td>
<td>Edits by MS. Inclusion of MACRO 4 software,</td>
</tr>
<tr>
<td>2.3</td>
<td>07 May 2014</td>
<td>Remove training log post JSOPG discussion - MS &amp; CS</td>
</tr>
</tbody>
</table>
1 Table of Contents

0 Version Record ............................................................................................................................................. 1
1 Table of Contents ...................................................................................................................................... 2
2 Glossary ...................................................................................................................................................... 3
3 Introduction .................................................................................................................................................. 3
4 Purpose ....................................................................................................................................................... 3
5 Roles and Responsibilities .......................................................................................................................... 4
6 Procedures ................................................................................................................................................... 6
   6.1 Procedural Flowchart ............................................................................................................................ 6
   6.2 Data Collection ..................................................................................................................................... 7
   6.2.1 Data analysis plan ............................................................................................................................. 7
   6.2.2 Database design ............................................................................................................................... 9
   6.3 Data Management ................................................................................................................................ 10
   6.3.1 Data transfer ..................................................................................................................................... 10
   6.3.2 Data storage ..................................................................................................................................... 10
   6.3.3 Data security ..................................................................................................................................... 11
   6.3.4 Post collection data validation ........................................................................................................ 11
   6.4 Quality Assurance ............................................................................................................................... 11
   6.5 End of trial ............................................................................................................................................ 12
7 Training Plan ............................................................................................................................................... 12
8 References .................................................................................................................................................... 13
9 Related SOPs ............................................................................................................................................... 14
2 Glossary

The full Glossary is in Swansea University H drive/Documents/526-WWORTH/Development Group/Glossary.

3 Introduction

Standard Operating Procedures (SOPs) are succinct formal documents designed to achieve consistency in specified trial functions by specifying standard practice in performing those functions (GCP 1.55 & 5.1.1 – EMeA, 2002). While SOPs should cite relevant legislation & regulations, and key references & evidence, they need not expound theory.

WWORTH SOPs should accord with all relevant regulations, including the European Union Clinical Trial Directive[^1], ICH Good Clinical Practice (GCP)[^2] and the current NHS Research Governance Framework[^3]. They will seek to distinguish between regulations for CTIMPs and for other research.

This document describes the principles for collecting and managing data in trials. We are grateful to colleagues in the North Wales Organisation for Randomised Trials in Health for access to their SOP on Data Management (NWORTH, 2007). The overall aim of this SOP is to ensure that data management procedures:

- manage personal data from trial participants in a legally responsible manner
- result in error-free, high quality trial data for analysis
- are efficient and cost-effective

All staff must comply with the Data Protection Act (1998)[^4] and should comply with the MRC’s code on Personal Information in Medical Research[^5]. All staff are required to have read these documents, agree to uphold their principles and sign a log after completion of training in this SOP.

4 Purpose

The purpose of this SOP is to define the major processes in the collection and management of trial data held by each trial or on behalf of WWORTH and the responsibilities of individuals involved. It also aims to describe good practice in trial data collection and management techniques.

This SOP should be used when any form of data is collected, accessed, transferred or stored by a trial. MACRO 4 is the single data management system that will centralise, automate, streamline and enhance the process of data collection whenever possible (WWORTH SOP 21 IT Databases).
It should inform protocol development (see WWORTH SOP13 Protocol Development) and trial methods.

Types of data can include the following:

1. Pre-consent data (including basic, non-identifiable participant identification details held at trial sites but not by the Trial Office)

2. Post-consent data (interviews, questionnaires, laboratory results, etc.)

Random allocation information (held at research sites and at WWORTH; see WWORTH SOP24 Randomisation).

5 Roles and Responsibilities

The personnel responsible for using and implementing this SOP are listed below. Each trial should allocate the responsibilities linked to these roles.

Chief Investigator (CI) - responsible for overall management of the trial, and delegating team responsibilities.

Principal Investigator (PI) - responsible for the trial at a given site.

Trial Data Manager (TDM) - responsible for all aspects of the trial data once it comes into existence. This includes checks on the accuracy and timeliness of data collection, production of accrual figures for reporting to the required authorities, provision of data to trial statisticians, health economists and outcomes specialists and overall security of the data including secure data storage in accordance with the Data Protection Act 1998⁴.

WWORTH Manager - responsible for overseeing and providing general guidance to the TDM of each trial, receiving regular updates on data management issues and ensuring that all staff adhere to the principles and practice outlined in this SOP.

Senior Statistician – Overall responsibility for the mathematical integrity of the data, the data checking algorithms and procedures and for the implementing of suitable “fit for purpose” databases for all trials managed by WWORTH (in collaboration with the trial research team and the IT specialist)

Database IT Specialist – responsible for the writing, testing and implementation of the databases and IT systems, including the upgrade of all database software implemented. Also responsible for all aspects of electronic data security.
Trial Statistician – delegated responsibility for the mathematical integrity of the data, the data checking algorithms and procedures and for the implementing of suitable “fit for purpose” databases for the specific trial (in collaboration with the trial research team and the Database IT Specialist). The Trial Statistician should approve and sign-off all final CRFs and other data collection instruments. The Trial Statistician, in consultation with the TDM, will be responsible for “locking” the final version of the data for analysis in accordance with WWORTH SOP 28 Statistics.

Trial Manager/Trial Coordinator (TM/TC) – where a TDM has not been assigned on a trial, the TM / TC is responsible for overseeing data collection in accordance with the trial protocol and for coordinating the databases and paper based systems and developing, in consultation with the Trial Statistician and the Database IT Specialist, the details of the specification and testing.

QA Officer is responsible for ensuring that all staff adhere to the principles and practice outlined in this SOP.
6 Procedures

6.1 Procedural Flowchart
6.2 Data Collection

6.2.1 Data analysis plan

An analysis plan will be written and agreed by all investigators and the DMEC / advisory members at the beginning of the study and prior to any data analyses taking place (see WWORTH SOP 28 Statistics). All data sources e.g. health economics, routine data, questionnaires and qualitative data should be included in the analysis plan.

Case Report Form development and completion

Case Report Forms (CRFs) and any other relevant data collection tool are completed for each trial participant who has consented and contain the data that will be used for trial analyses. CRF data entry must be verifiable using source documents e.g. patient notes, and stored safely and securely on site as these documents are confidential, although ownership lies with the trial Sponsor (see WWORTH SOP31 Sponsorship and Adoption).

CRFs are developed at the beginning of a trial based on the funding bid and the protocol which will describe the data to be collected for analysis by the trial statistician. Typically, the CRF or any other data collection tool will contain patient demographical data, clinical data, details of medical and surgical histories, etc, as necessary depending on the design of the trial. Each page of the CRF/ source data should have the patient study ID and a space for the person completing the page to enter their name and the date it was completed.

Subject to the provisions of the Data Protection Act, and Caldicott guidelines the CI and PI should agree who will be responsible for making data entries and amendments to the CRF and for storing them appropriately, and record them in the study delegation log. Anyone not authorised to collect / amend / file CRF data should not do so.

The trial may ask the PI or other responsible person at the study site to do one of the following with Paper based CRFs/ source data:

Fax the CRF/ source data to the Trial Office. A “fax send report” is filed with the CRF. /source data.

- Post a copy of the CRF /source data to the Trial Office, marked as confidential with the return address indicated and a copy of the covering letter is filed with the original CRF

- Scan and send the CRF /source data

- University FaxPress the CRF/ source data
The sponsor may come and collect the CRF after verification of source data.

Transfer of eCRFs to the trial office should be carried out following File Transfer Protocol (FTP) as outlined in WWORTH SOP 22b Data Protection and Confidentiality.

Completion of paper-based CRFs should be done clearly and any errors crossed through once, dated and signed off by the person making the amendment. Correction fluid is not permitted. Changes to eCRFs must be auditable so that any corrections can be traced back (see WWORTH SOP 21 IT Databases).

CRFs are subject to quality control checks during a monitoring visit by a member of the Trial Office, usually the TM or TDM. The contents of the CRF are compared with source documentation (see WWORTH SOP 17 Monitoring and WWORTH SOP 18a Quality Management). They may also be audited or inspected by an external regulatory body at any time. The procedures involved in this are described in WWORTH SOP 18b Preparing Audit Inspection.

All eCRFs must be saved in a password-protected electronic form to protect against loss and should be printed to create paper CRFs at the end of a trial. All forms of CRF should be archived at the end of the trial in accordance with WWORTH SOP 08 Archiving.

Other forms of data collection
Treatment allocation forms, questionnaires, logs and any other forms to be completed for a trial on each participant or study site should be designed and completed in accordance with the relevant SOPs. Data must be anonymised and encrypted as soon as practicable.

Trials using data from multiple sources, may have more than one value for the same variable. It may be necessary to design a variable list which will outline all the available data sources as well as a prioritisation log - which will outline how to use source data.
6.2.2 Database design

Data should be collected, with verification of entry, using standardised data collection tools. Wherever possible, data should be entered directly into the trial database. Data transfer should be as close as possible to the point of collection (e.g. using a hand-held computer or laptop). Each trial is guided by TMG responsible for ensuring the quality of data entry for example verification of data.

Data transfer into databases is described in more detail in WWORTH SOP21 IT Databases.

In situations where data from paper CRFs are scanned into databases, data capture schedules for scanned data should be designed by trained staff, and checked for accuracy on sample forms before use in the trial. They should incorporate suitable validity checks wherever possible, and either prevent or force operator verification of unexpectedly missing or suspect entries in accordance with WWORTH SOP18a Quality Management.

Databases may be designed for direct electronic data entry. Where this occurs, suitable data validity checks should be incorporated at the point of data entry to prevent “out of range”, “missing” or other checkable data entry errors (see WWORTH SOPs 18aQualityManagement and WWORTH SOP21 IT Databases).

Pre-trial testing of data management procedures should occur to ensure that they are appropriate for the purposes of the trial, are efficient, cost effective and result in high-quality, error-free data before the launch of the main trial. Pre-trial testing is likely to require initial feasibility testing, pre-
pilot testing (e.g. testing of specific data management procedures) and pilot testing of the final procedures (see WWORTH SOP16 Site Setup).

6.3 Data Management

6.3.1 Data transfer

This includes:

- Any means of electronic data flow including File Transfer Protocols, emails, database to database transfers, portable devices and storage mechanisms.

- Paper flow, e.g. on forms, filled in by researchers, research participants or medical professionals and transferred by person to person or by fax or postal means or scanned CRFs sent by email.

- Paper to electronic flow including scanning and manual data entry.

- Data transfer is legitimate if it is auditible, secure and reliable. All methods of data transfer should be done in accordance with the principles of data confidentiality set out in the Data Protection Act 1998 and MRC's guidelines (Personal Information in Medical Research), both of which are described in WWORTH SOP22b Data Protection Confidentiality.

6.3.2 Data storage

This includes:

- Electronic data however stored (for example databases, computer hard drives, pen drives, remote server or other temporary and permanent media). All such data must be password protected.

- Paper documents which may include source data (for example CRFs and questionnaires) and all documentation relating to the management of the trial. Further information on documents stored in Trial Master and Site files can be found in WWORTH SOP03 Master Site File.

Systems of record keeping and database management must prevent the loss, missing or unreadable information that compromises future data analysis. Interim and final archiving of documents should be done in accordance with WWORTH SOP08 Archiving.
6.3.3 Data security

Data accessibility should be done in accordance with delegation log (see WWORTH SOP21 IT Databases and WWORTH SOP22b Data Protection Confidentiality).

Data security protocols for trials should comply with the principles of data confidentiality set out in the Data Protection Act 1998\(^4\) and MRC’s guidelines Personal Information in Medical Research\(^5\), both of which are described in WWORTH SOP22b Data Protection Confidentiality. GCP guidelines must be followed to ensure patient data is secure.

Paper versions of data collection forms will not contain any patient identifiers and should not be held together with patient identifiers, i.e. should be stored in separate filing cabinets, not in different drawers of the same filing cabinet. Electronic data containing patient identifiers should be held on two password-protected databases with restricted access to both:

- one with only identifiers and a study ID ONLY and
- one with all relevant study data (clinical measures, patient-reported measures) and study ID only

This will prevent patient data being identifiable.

6.3.4 Post collection data validation

Routine statistical validity checks should identify missing or suspect entries at agreed defined intervals within the design of the trial so that correct data can be obtained from the research staff, participants or their records if needed. Thorough data cleaning should be performed before any data analysis (see WWORTH SOP18a Quality Management and WWORTH SOP28 Statistics).

Manual data entry checking should be done when data is entered manually from a source document into a database MACRO 4. Where resources allow, this should be done separately by two independent data entry clerks and then validated to identify data punching errors. Where double entry of all data is not possible, a randomly-selected sample of at least 10% of the data should be double entered and validated. A decision can then be made regarding further validation and correction, depending on the number of errors identified in the quality sample (also see WWORTH SOP18a Quality Management).

6.4 Quality Assurance

Database documentation should be clear and logical. A paper or electronic audit trail should be kept of all deletions, corrections/changes or recodes.
that are performed after data capture, including version changes due to protocol amendments (see WWORTH SOP01b documentation control, WWORTH SOP18a Quality Management, WWORTH SOP18b Preparing Audit Inspection and WWORTH SOP21 IT Databases).

The TDM should check the accuracy and reliability of the raw data (e.g. against original data sources such as hospital records) at pre-defined regular intervals throughout the study. Similarly, instruments used in the study (e.g. weighing scales, laboratory instruments) should be quality controlled at pre-defined intervals throughout the study, either during a monitoring visit or remotely by authorised site staff in accordance with WWORTH SOP17 Monitoring and WWORTH SOP18a Quality Management.

Regular, frequent monitoring of data quality throughout the data collection phase of the trial is important (see WWORTH SOP17 Monitoring, WWORTH SOP18a Quality Management and WWORTH SOP28 Statistics). Trial documentation data and agreed amendments should be retained for audit purposes. A paper or electronic audit trail should be available with amendment dates of versions, & audits, as described in WWORTH SOP18a Quality Management and WWORTH SOP18b Preparing Audit Inspection. At the end of trial, the amendments will be archived in accordance with WWORTH SOP07 Site Closure and WWORTH SOP08 Archiving.

### 6.5 End of trial

Once a trial has finished, the final, cleaned databases will be locked, dismounted and archived for audit and storage in accordance with WWORTH SOP07 Site Closure and WWORTH SOP08 Archiving. The study site data will be checked against the random allocation sequence (for WWORTH managed trials), discrepancies will be documented and resolved, so that they can be taken into account in intention to treat analyses. A copy of the random allocation information will be provided to the TM if requested (see WWORTH SOP24 Randomisation).

These data will be maintained for the period agreed in the trial protocol and records will be archived in accordance with WWORTH SOP08 Archiving. The databases will be available for inspection by interested legitimate parties in accordance with WWORTH SOP18b Preparing Audit Inspection.

### 7 Training Plan

All WWORTH staff involved with trials must undertake the appropriate generic and trial-specific training to ensure that they meet with the specific employers’
mandatory training requirements and the specific requirements of the trial. For example, for SU staff, all new employees must attend induction, fire and safety training (as well as role-specific training courses, e.g. laboratory safety). For new staff, additional training requirements should be identified alongside the specific role requirements and the WWORTH Unit Manager should make provision for the new staff member to attend the necessary courses as soon after appointment as is practicably possible.

It is the responsibility of the WWORTH Unit Manager (alongside the CI or TM) to identify all the SOPs that are relevant to a specific trial and in which the new member of staff should be trained. The WWORTH UM or the SOP author will provide group training for trial staff and/or one-to-one training, as required for new staff in relation to the specific SOPs identified. Training records should be filed both by the main employer and the staff member, in accordance with the specific employer requirements. Trial specific training should be filed in TMF or TSF as appropriate and every individual involved in a trial should have an individual training record (see WWORTH SOP02 Training).

Where the tasks specified in the individual SOPs are delegated to WWORTH staff, CIs/PIs or TMs, these delegated staff must ensure that they have attended a training course on GCP and keep up-to-date through attending refresher courses.

It is the responsibility of the CI/PI to ensure that all staff allocated duties on the study delegation log template of responsibilities are suitably trained in the activities linked to those duties (see WWORTH SOP16 Site Setup, Appendix 9 and Appendix 10).

Each trial should maintain a central training log and ensure that WWORTH has access to that log, not least to integrate the logs of staff who work on more than one trial. Similarly trials should ensure that each site maintains a local training log, not least to integrate the logs of staff who work for more than one sponsor.

8 References


5. Personal Information in Medical Research Medical Research Council’s Ethics Series. Available at www.mrc.ac.uk


9 Related SOPs

WWORTH SOP01a on SOPs
WWORTH SOP01b Document control
WWORTH SOP02 Training
WWORTH SOP03 Master Site File
WWORTH SOP04 Trial Supplies Labelling
WWORTH SOP05 Patient Info Consent
WWORTH SOP07 Site Closure
WWORTH SOP08 Archiving
WWORTH SOP10 Project Management
WWORTH SOP16 Site Setup
WWORTH SOP17 Monitoring
WWORTH SOP18a Quality Management
WWORTH SOP18b Preparing Audit Inspection
WWORTH SOP21 IT Databases
WWORTH SOP22b Data Protection Confidentiality
WWORTH SOP24 Randomisation
WWORTH SOP25 Outcome Measures
WWORTH SOP26 Health Economics
WWORTH SOP27 Qualitative data
WWORTH SOP28 Statistics
WWORTH SOP31 Sponsorship and adoption
WWORTH SOP32 Detecting & Managing Misconduct, Serious Breaches & Deviations from GCP/Protocol