Psycho-oncology Co-operative Research Group

Study Documentation and Data Management

SOP ID: SOP_4.0.0-DM-121203

SOP development and approval

SOP developed by	Date	Associated document(s)
Michelle Peate Research Program Manager	04/06/2012	**SOP_4.3.1-DM-120604-Data_Managment_Template**
Michelle Peate Research Program Manager	04/06/2012	**SOP_4.3.2-DM-120831-Data_Dictionary_Template**
Michelle Peate Research Program Manager	04/06/2012	**SOP_4.3.3-DM-120604-SPSS_PODB_Template**
Michelle Peate Research Program Manager	04/06/2012	**SOP_4.3.4-DM-120604-Excel_PODB_Template**
Michelle Peate Research Program Manager	04/06/2012	**SOP_4.3.5-DM-120604-Access_PDB_Template**
Michelle Peate Research Program Manager	04/06/2012	**SOP_4.3.6-DM-120604-Archive_Label_Template**
Michelle Peate Research Program Manager	04/06/2012	**SOP_4.3.7-DM-120604-Archives_Register_Template**
Michelle Peate Research Program Manager	04/06/2012	**SOP_4.3.8-DM-120604-Data_Destruction_Form**
Michelle Peate Research Program Manager	04/06/2012	**SOP_4.3.9-DM-120604-Data_Managment_Checklist**
Michelle Peate Research Program Manager	04/06/2012	**SOP_4.3.10-DM-120604-File_Hierarchy**
Michelle Peate Research Program Manager	23/01/2013	**SOP_4.3.10-DM-130123-Archive_Metafile**

Review panel: Melanie Price, Melanie Bell, Monika Dzidowska, Haryana Dhillon, Mark Maclean

Approved by	Date	Signature
Phyllis Butow PoCoG Executive Committee Chair	04/06/2012	
Monika Janda PoCoG Scientific Advisory Committee Deputy Chair	04/06/2012	

Supersedes documents:

SOP Revisions

Approved by:	Date	**Signature** (only required for PoCoG hard copy)	Description of change(s)

Date Administered: 14/06/2012
Recommended date for review: 14/06/2014

THE UNIVERSITY OF SYDNEY

Foreword

The Psycho-oncology Co-operative Research Group (PoGoG) has developed a program of quality assurance for psycho-oncology research. PoCoG's Quality System requires documentation of both management and procedural activities. This guidance document Study Documentation and Data Management provides a standard working tool for outlining the data management process, including where the analyses will take place, how data will be entered on the database, and how data will be tracked, checked and audited.

Questions regarding this document should be directed to:
Research Program Manager
Psycho-oncology Co-operative Research Group (PoCoG)
Level 6, Chris O'Brien Lifehouse (C39Z)
The University of Sydney, NSW, 2006, Australia
E-mail: pocog.office@sydney.edu.au
Phone: +61 2 9036 5002
Fax: +61 2 9036 5292

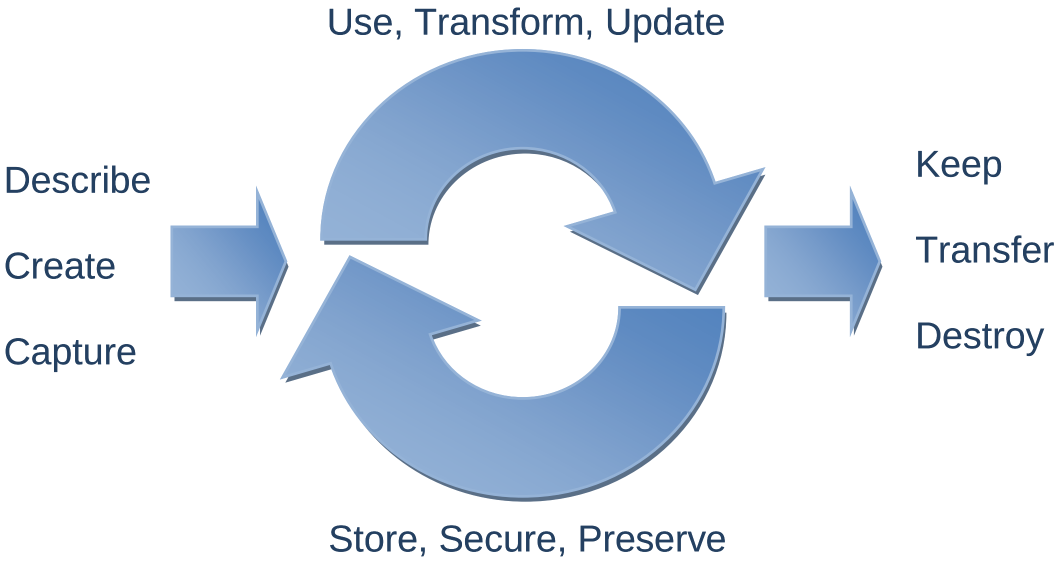
Table of Contents

Overview

Good research conduct requires that the research methods and results are open to, and able to withstand, scrutiny and debate. Research data and records therefore need to be accurate, complete, authentic and reliable. Data and records need to include sufficient detail to establish their authenticity and confirm the validity of conclusions. Data transformations, study set up, operation logs, codebooks, and forms that are used to generate data may be crucial elements in establishing validity. Thus, to facilitate good research, each study must establish formally documented procedures to be followed by anyone involved in the study.

By definition, research data management encompasses the design, collation, cleaning, and management of all participant and other information, observations and measurements. Efficient and effective data management is necessary to minimise risk of error. Data management is an integral part of the research process and should be considered in conjunction with development of the full study protocol. Data management involves not only data collection and entry, but also the ongoing management of data, data preparation, analysis and publication, data archiving and destruction (Figure 1). Ideally, a data manager should be consulted in this process, although this may not always be feasible. A minimal requirement for all studies is a Data Management Plan (DMP) to ensure the quality of research data and outputs, integrity and repeatability, appropriate access to data, and appropriate reuse of data for subsequent research and to record the person(s) responsible for various aspects of data management.

Figure 1: Data management lifecycle



The most important ‘rule’ in data management is documentation – there should be an audit trail through which changes in information, and alterations to data, can be traced back to source, should anything need to be investigated. This documentation is a safety net for the data collected as part of the study.

The key elements of data management are:

- Data must be recorded in a durable and appropriately referenced form that complies with relevant regulations and policies.
- The data must be retained for sufficient time to allow reference. Recommendations for the duration of data retention have been prescribed in national, and international, guidelines.
- Data reported in publications should be available for discussion with other researchers, without breaching confidentiality.

Purpose

This SOP describes the PoCoG processes for research data collection and management (including procedures for quality control (QC), data query resolution, record retention and archiving). This document will guide the development of a Data Management Plan, and is written so that anyone involved in the study can understand the procedures with relative ease.

Scope

This SOP applies to the data management of studies subject to Australian and international regulations, for all phases of development. In particular, this document is directed towards the data management in psycho-oncology research and specifically for PoCoG studies.

For the purposes of this document, research data management includes the:

- Definition of required study data
- Design of databases and case report forms (CRFs)
- Data entry processes
- Development of a data dictionary
- Access and permissions to documents and data
- Collection and storage of participant details and study data
- Quality control and assurance of data
- Data archiving and destruction

Guiding Principles

Research studies are required to be conducted in accordance with applicable legislation and regulatory standards. Thus, this document has been guided by the following resources:

- the International Conference on Harmonisation documentation on Good Clinical Practice (ICH-GCP) and the Therapeutic Goods Administration (TGA) annotated version.
- the National Health and Medical Research Council National Statement on Ethical Conduct in Human Research (henceforth referred to as the 'national statement'.
- The Australian Code for the Responsible Conduct of Research.
- the Therapeutic Goods Administration's Australian Clinical Trial Handbook.
- the Declaration of Helsinki.
- the Australian Law Reform Commission's review of the privacy legislation.
- Australian Bureau of Statistics, *National Statistical Service (NSS) Handbook*.
- Australian Psychological Society, 2007. *APS Code of Ethics*.
- The Australian Sociological Association. *Ethical guidelines for research*.
- Kelman CW, Bass AJ, Holman CD. Research use of linked health data - a best practice protocol. *Aust N Z J Public Health.* 2002;26(3):251-5.
- Statistical Information Management Committee, 2007. *Guidelines for the Use and Disclosure of Health Data for Statistical Purposes*.
- Electronic Medical Records in Australia and Clinical Trials.

Definitions and Abbreviations

Case report form (CRF)

A paper or electronic questionnaire specifically used in clinical research. The CRF is the tool used to collect data. All data on each participant participating in a clinical trial are held and/or documented in the CRF, including adverse events.

Data dictionary

A catalogue of information about data collected for the study. This includes the variables, relationships to other data, origin/ question, values, usage, and format.

Data Management (DM)

The development, implementation and supervision of policies relating to the management of study data. This includes mechanisms to protect the data.

Data Management Plan (DMP)

The plan that defines details of policy and implementation of the management of data.

Human Research Ethics Committee (HREC)

Human Research Ethics Committees responsible for the ethical review of research studies involving humans.

Nominal variable

A categorical variable that has no order to it (the assignment of numbers to categories is purely arbitrary).

Ordinal variable

A categorical variable in which the categories have an obvious order (e.g. strongly disagree, disagree, neutral, agree, strongly agree).

Participant Consent Form (PCF)

The form which participants sign to indicate their consent to be involved in the study.

Participant Information StatementStatement (PIS)

An information sheet that outlines the study and the risk, benefits, and what participation in the study involves.

Participant Database (PDB)

Database for sorting participant contact and follow-up information.

Participant Outcomes Database (PODB)

Database for storing participant outcome data.

Principal Investigator (PI)

The investigator responsible for the coordination of Chief Investigators and Site Investigators in a multicentre study. Often the Principal Investigator is also Chief Investigator A (CIA) on grant and ethics applications and will lead the *Management Committee*. Also: Coordinating Investigator.

Quality Control (QC)

Detection activities which focus on detecting ‘problems’ with research study processes. The operational techniques and activities undertaken within the Quality Assurance (QA) system assess whether the requirements for quality of the study-related activities have been fulfilled.

Research Assistant (RA)

A person who assists with conducting the study.

Sponsor	A person or organisation that provides funds or support for a project or activity carried out by another. In the context of this document, PoCoG is a sponsor for PoCoG endorsed studies by providing support for endorsed studies.
Standard Operating Procedure (SOP)	Detailed written instructions designed to achieve uniformity of the performance of a specific function.
Unique Participant Number (UPN)	A number that is allocated to a participant and is unique to that participant.
Variable	A factor that is likely to vary or change.

Qualifications and Responsibilities

This SOP applies to those members of the research team involved in data collection, transcription and management. This includes:

- Principal Investigator (PI)
- Chief Investigators (CIs)
- Study manager/ coordinator
- Research and support staff
- Data manager
- Study statistician

1. Procedure(s) – Developing a Data Management Plan

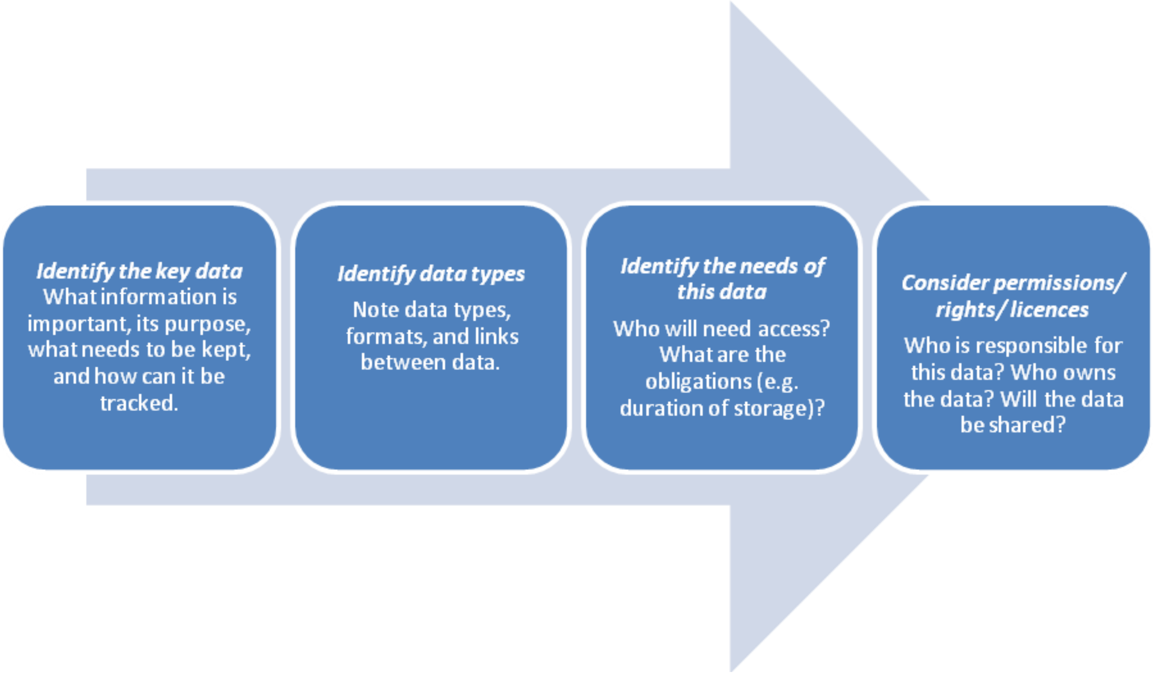
The most important concept in data management is to KEEP GOOD RECORDS. The role of documentation is to support research integrity, repeatability and allow for tracking back to source, if necessary. This is best achieved by having a data management plan (SOP_4.3.1) which should:

1.1. Identify and describe key data

Identifying key data

The first step is to consider the data carefully in terms of what is being collected, why, and how it will be used. In many psycho-oncology studies this will include data from medical records, self-administered questionnaires (http://www.pocog.org.au/content.aspx?page=podlaunchpage) and health professional derived data. Address the different elements shown in Figure 2. Once you have a clear idea about the data consider the lifecycle of these key elements.

Figure 2: Steps to identify and describe key data.



Identify the key data What information is important, its purpose, what needs to be kept, and how can it be tracked.

Identify data types Note data types, formats, and links between data.

Identify the needs of this data Who will need access? What are the obligations (e.g. duration of storage)?

Consider permissions/ rights/ licences Who is responsible for this data? Who owns the data? Will the data be shared?

For example, if an objective of your study is to assess the Body Mass Index (BMI) of participants:
Key data - W*hat information*: height in metres and weight in kilograms (i.e. two different data points)
- *Purpose*: To calculate BMI (i.e. third data point)

Type of data - Type: Numeric
- *Format*: To two decimal places
- *Links*: BMI = height/ (weight)^2

Needs: Accessible by study research staff. To be stored for the duration of the trial and as regulations stipulate.
Permissions: Study coordinator will be responsible for this data which is owned by the investigators. Data will only be shared in summary form and any identifiable data is to be de-identified.

Cancer Australia minimum data set

PoCoG endorsed studies are required to collect a minimum data set. Please refer to the PoCoG website (http://www.pocog.org.au/content.aspx?page=minimumdataset) for up-to-date details of the data that is to be collected.

Describing key data – the data dictionary

This information needs to be included in the study data dictionary (SOP_4.3.2) to catalogue every individual data item (variable), the question text, data type and the possible values (precision and range) it can have. For example, the variable for age may describe the inclusion criteria such as 18 – 40 years, whereas a Likert scale variable might indicate a range from one to five. Any standardised questionnaires you use should also be attached and cross-referenced to the data dictionary along with the suite of CRFs/ questionnaires used for the study.

Tips for setting up a data dictionary:

1. In designing your variables, use numeric values even if they are nominal or ordinal. For example, this may describe how the variables will be coded into “Values” (SPSS) or via Proc Format (SAS) (e.g. men = 0, women =1).
2. Use consistent nomenclature for variables and values. Variables representing items from a scale, for example, should have a common and time indicative prefix (T0FG1, T0FG2, etc for the first and second questions to the FACT-G at baseline (T0)).
3. If using a standardised scale such as the HADS or the FACT-G, match codes to values used for scoring the scale (e.g. 0-3 for the HADS and 0-4 for the FACT-G).
4. Describe logic checks that will be referenced and implemented upon data entry or during data cleaning prior to analysis. For example, has the person had cancer longer than they have been alive? Does a man have ovarian cancer? Is the date of the second assessment before baseline?

1.2. Outline the mechanisms to capture the data

The steps and procedures for data collection and data entry should be outlined in detail and include the responsible person for each step of the process. For instance, some data may be collected through self-report questionnaires completed by participants and collected by the clinical trial nurses who post them to a central location where they are checked and entered. Treatment data may be extracted from medical records by the clinical trials nurse and entered directly into an online CRF at the site.

Development of case report forms (CRFs)

Depending on the data to be collected and the method of collection, CRFs (including questionnaires) should be developed for each study and include details of how these should be processed. Templates for different forms and measures databases are available on the PoCoG website (www.pocog.org).

1.3. Outline the infrastructure and mechanisms to store the data

The data management plan should also describe the mechanisms by which the databases for storing participant data should be set up, for example, instructions for naming variables. Other options for data storage may include using online workspaces, spreadsheets or text documents. Specific instructions on data entry are described in section 1.5.

Physical and digital data storage

Ideally, data storage for multicentre studies should be centralised. A register (meta-file) should specify the location of the different types of data (paper, electronic, and audio-visual) and include a description of the data and the name of the researcher responsible.

Items to consider:

- Is the data going to be stored centrally or at sites?
- What is the timeline for data collection and storage?
- How much data storage is needed?
- How is the system secured (locked filing cabinets, password protected databases)?
- Have you created a log for cataloguing the movement of data?
- Where is data stored?
- In what format will it be stored and why this format has been chosen?
- Is any specific software required to read, analysed or process the data?
- Who is responsible for the data?
- What are your institution’s data management policies?
- How will remote data be stored?

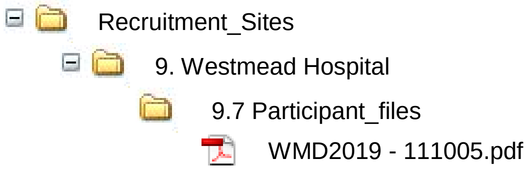
A file tree that is currently used for PoCoG administered study documents is available (SOP_4.3.10). In regards to the data described in this SOP, the folder numbered 19 and labelled “Study_Data” is where the participant related information should be stored.

Storage on a local level

For users of this folder the data management plan should outline:

a. The use of a standard naming procedure for electronic files, which includes either the date or a number for versioning, both in the file name and within the document itself. PoCoG recommends using reverse date [yymmdd] for this purpose. For example, your database may be named: ‘RAVESDA-PtData-110425.xls’, and when you make changes on the 27th of April, the next version will be saved as ‘RAVESDA-PtData-110427.xls’.

b. Instructions for file nomenclature should include something that identifies what the document is, who it refers to (if applicable, such as ID number or location code), and the version. For example, ‘RecruitmentForm-WMD2019-111005.pdf’ might be used as the file name for the recruitment form for Westmead Hospital participant ID 2019 who was recruited on the 5th October 2011. Alternatively, you may prefer to use the folder structure to identify content or activity, for the same example:



Recruitment_Sites

9. Westmead Hospital

9.7 Participant_files

WMD2019 - 111005.pdf

c. Also determine the limit for the number of versions that will be kept at any one time. It must also be clearly indicated that there can only be ONE current version.

1.4. Describe data security

Though not research oriented, the Payment Card Industry Data Security Standard (PCI DSS) is a useful standard to refer to in regards to protecting participant data. The aims are to:

1. Build and Maintain a Secure Network (including use of confidential passwords and documenting an audit trail to capture changes to information).
2. Protect Participant Data (e.g. de-identification of personal information and use of participant IDs).
3. Maintain a Vulnerability Management Program (e.g. using anti-virus software and having a system of regular backups).
4. Implement Strong Access Control Measures (e.g. restricting access and using unique IDs for each person who accesses data, criteria for using electronic signatures).
5. Regularly Monitor and Test Networks (e.g. tracking and monitoring access and testing security systems).
6. Maintain an Information Security Policy (i.e. the system should be configured so that it is capable of restricting the access of various subscribers/ sponsors to **ONLY** those records for participants in the trial who have consented to have their records monitored).

Refer to the "Electronic Medical Records in Australia and Clinical Trials" document (http://www.pharmacouncil.com.au/news_pdfs/Electronic%20Medical%20Records%20in%20Australia%20and%20Clinical%20Trials%20V%203.0.pdf) for more information. This document contains a useful list of questions in regards to electronic medical records.

The organisation in which the coordinating centre is situated will have policies about data security which will apply to your data, in the case of PoCoG administered studies the University of Sydney policies apply. Please refer to these for general 'rules' in regards to the physical safety of data (to damage or loss, whether it be unintentional or intentional), and it will most likely suffice to reference this policy. Ultimately, you need to consider how important the data is and what is considered an acceptable loss. For most studies loss of data will have a serious impact and thus it will be important to have a process to minimise the potential for this data loss happening (including using backing up of digital data – see section on data back up below).

Since the security of data is not merely restricted to loss of data due to technological failure, it is also recommended that you describe who has access to the data. By restricting access, the risk of incorrect use of the data can be limited. There are some basic steps that should be used for digital data:

- Use firewall servers which should be physically protected (e.g. locked up).
- Do not use default usernames and passwords – use unique IDs for each user (so that you can monitor access).
- Use encrypted transmission over the internet (e.g. VPN, SSL, SSH, GridFTP, S/MIME email).
- Have updated antivirus and antimalware software.
- Restrict access to sensitive data.

As the majority of psycho-oncology research is participant based, special care is needed in considering the safety of this material. Researchers have an obligation to be aware of and comply with privacy legislation and policy and how this may affect the data collection, storage, use and disclosure of the information they wish to collect. The key principle is that one should never store identifiable data (name and address etc) without permission. As the Human Research Ethics Committees (HRECs) are responsible for ensuring that participant privacy is protected according to the Commonwealth Privacy Act (http://www.privacy.gov.au), it is essential that ethical approval is obtained. In light of this particular attention needs to be paid to informed consent and data de-identification.

Informed consent

All the pertinent information to obtain informed consent is often provided to participants in the form of a Participant Information StatementStatement (PIS) and Participant Consent Form (PCF), and most HRECs will have templates and guidelines available to researchers. Refer to PoCoG's SOP on ethics for some guidelines specific to PoCoG research projects.

Key procedures to be described in the data management plan:

- Signed PCFs are to be kept with the PIS, as proof of informed consent and stored separately (and securely) from the de-identified data, for an appropriate period (the University of Sydney requires data to be kept for a minimum of 7 years, refer to: http://sydney.edu.au/research_support/ethics/human/guidelines/data.shtml).
- A copy of the PIS should be stored together with the collected data (e.g. completed questionnaires), so that it is apparent what was obtained through consent.
- If verbal consent is obtained the date should be recorded on the file.
- If no consent form was received but data was collected (questionnaire or interview) this should also be noted in the database.
- Who is responsible for management of consent records?

De-identification of participant data

De-identification is the process for removing all identifying information from the data to protect the privacy of individuals, necessary for publishing or sharing data. This is often achieved by stripping, encrypting, or recoding names, addresses, or any other identifiers.

More specifically de-identification involves the removal of any information that will allow someone to determine the identity of another individual, such as name, address and other contact details, balanced against retaining enough information to confirm who that participant is if necessary. The use of unique participant numbers (UPNs) at recruitment will allow the re-identification of data if needed. However, to confirm that this is the correct participant you may also wish to retain date of birth and site recruited.

A master-file of names and other identifiable data should be stored securely, and separately, from study data in locked/ password-protected databases with passwords kept separately.

Sharing data

The type of data that is confidential should be described, and sharing of data should fall within regulatory and legislative requirements. Refer to the study's research integrity document (a separate document designed specifically for the study that outlines the principles for sharing, publication and dissemination) that describes when and how data can be shared with other researchers and specifies any embargoes.

1.5. Standardising data entry, checking and validation

The database should be developed based on the information in the data dictionary. As mentioned above, try to avoid 'dirty data' by using consistent variable nomenclature and using numeric values for each variable. Template databases for participant reported outcomes are available in SPSS and Excel to PoCoG study researchers for use as a starting point for their data (SOP_4.3.3 and SOP_4.3.4) and include the PoCoG minimum data set. A participant contact database template is also available (SOP_4.3.5). Use of data management software should be considered (e.g. OpenClinica or REDCap) to reduce risk of error.

Data entry

The data management plan (SOP 4.3.1) should describe the routine for data entry, how missing variables are to be coded, how logic checks (as described in the data dictionary) will be implemented as data are entered, and the process for querying any inconsistencies. If resources are available, double data entry should be considered.

Updating data

The data management plan should include procedures for alterations and updating of data. This should include details as to who is responsible for making these changes and how details should be logged. For example, once a week the research assistant (RA) might be scheduled to enter all the questionnaires received in that week into the database. The RA will date and initial the front of each questionnaire as it is entered. Some databases will record details of access and dates of data entry (ideal) however, for studies using less sophisticated databases it is suggested that the date data was entered is logged in the separate participant contact database.

Cleaning and validation

Data cleaning is an important part of quality assurance and control. Steps to minimise dirty data are described above, but there are also steps that can be taken to clean the data after it has been entered. It is important to be vigilant over your data at all times and there are some steps that can be taken ***prior*** to analyses:

1. *Check for invalid character values.* If possible, build these checks into the data entry system. Alternatively, this can be done by using frequency tables.
2. *Check for invalid values for all Likert scales*. This may be achieved by only allowing data entry to be selected from that Likert code-list or also by using frequency tables.
3. *Check valid range,* e.g. physical measures have an expected range of values outside which you should query the entered data.
4. *Check univariate statistics of all continuous variables*, including derived scores: Min, Max, (out of range values), missing value rates. Percentiles and histograms can also aid in finding invalid values, as can printing values that are out of range.
5. *Check for repeated IDs*.
   a) There are a few ways to do this in SPSS, firstly sort the data, then look for repeats. Alternatively, this can also be done in frequency tables. Both of these approaches can be cumbersome when datasets are large.
   b) In SAS, a dataset set by id, then using something like
   ‘`IF FIRST.id NE LAST.id THEN repeat=1`’.
   will find repeats.
6. *Check dates for validity.* For example, check that date of diagnosis is before the treatment date, or assessment dates, by creating a variable that calculates the difference in time and checking for negative values.

1.6. Outline a strategy for backing up data

The data management plan should outline the processes of backing up data. For example, data may be stored on a server which is backed up daily. Details here may include the review and upgrades of data format. In regards to paper data, scanning key documents as the study progresses is a form of back up. Digital back up may also occur off-site or at multiple sites to provide better security of the data.

Questions you may have for your system administrators include:

- How frequently is the data being backed up?
- How will disaster recovery be dealt with?

1.7. Auditing data

Regular audits of the data should occur throughout data collection. Most often, these will be self-audits and should be periodical. Some research groups may wish to request an external audit from another research group, some sponsors will conduct audits, and of course there is always the chance of a formal/ regulatory audit for example from your ethics committee. The purpose of an audit is to ensure that research is being conducted as specified in the protocol approved by the ethics committee and to identify any errors that may be occurring and why they are occurring. Details in your data management plan may include how often an audit will be conducted, the types of checks which will be performed and what are the acceptable rates of error. An example of a planned self audit can be found in the template data management plan.

Like every other aspect of data management, it is important to document audits (for example by having an audit report) and log activities.

1.8. Using the data – preparation for analyses and dissemination

This is the preparation of the data for analyses by the statistician. At this stage consider the *Quality of data.* It is worthwhile to repeat the data cleaning checklist described in Section 1.5 again prior to commencing locking the data (the process where the data can no longer be altered) for analyses.

Other considerations at this time include:

- How do data cleaning decisions influence the distribution of variables?
- Review of missing observations - are there many missing values, and does there appear to be any pattern in how the data are missing?
- Review of outlying observations, potentially returning to participants’ clinical files for double-checking.
- Comparison and correction of differences in coding schemes.

1.9. Outline the plan for archiving, long term storage, and destruction of data

Research data and records should be maintained as specified by legislative and regulatory requirements. Generally, clinical trial data should be kept for 15 years from the date of publication as recommended by the NHMRC (refer to http://www.nhmrc.gov.au/guidelines/publications/r39). The recommended length of time to retain data varies depending on the type of study and location; refer to funding body specific requirements when designing the data management plan.

Archiving and long term storage

Data should be stored in such a way that they are quickly and easily identified and retrieved when required – it must be possible to demonstrate that the data can be retrieved. A register should be maintained specifying the location of the storage – this meta-file should also include a description of the data and the name of the researcher (SOP_4.3.7).

Not all data needs to be archived. Data that is archived should be of value to the research project, keeping in mind state, national and international regulations. It is crucial that consent forms are kept and, wherever possible, data kept de-identified. Data and records should be boxed and labelled with the project title, date of publication (or date of transfer to the central storage area), PoCoG endorsement number, name of the researcher who is responsible for the data, date for destruction, and number of boxes (see SOP_4.3.6 for labelling of boxes).

It is important to be forward thinking in regards to storing data. For instance, microfilm and non-acidic paper last for over 100 years compared with such mechanisms as magnetic media (10+ years), optical media (20+ year), or hard drives (of which 2-10% fail annually). The other consideration is the speed at which software and hardware become outdated. The National Archives of Australia recommend that digital data be converted into a standard, stable format, such as eXtensible Markup Language (XML) as this enables records to be read with computers in the distant future.

Most organisations have a system for archiving. For example, the University of Sydney has the Archives and Records Management Services (ARMS) who provide a recordkeeping system. This repository is compliant with the NSW State Records Act, NSW Government Information (Public Access) Act and the NSW privacy legislation. Refer to the ARMS website (http://sydney.edu.au/arms) for more information and for contact details. It is also recommended that you speak to an ARMS staff member about the options available and the services that they provide and the processes involved in using these services. Alternatively, the University also has the Research Data Store (RDS) for the storage of research data (http://sydney.edu.au/research_support/data/where-to-store-data.shtml). This is a public access data repository and may not be appropriate for all data.

Sharing of data

Consider whether the data will be made available to other researchers. This may occur after publication of results, after a period of embargo or the data may never be shared. In the situation where sharing is permitted, address the access conditions. This may include:

- Open licences
- All rights reserved
- PI to be contacted to negotiate conditions of access and re-use

Destruction of data

Data destruction must be recorded, noting the dates and authority on which this action was taken (SOP_4.3.8 is a data destruction form). When confidential research data and records are destroyed it should be done in such a way as to ensure complete destruction of the information. Confidential research data and records in paper format should be shredded. Confidential research data and records in electronic format should be destroyed by reformatting or overwriting. ‘Delete’ instructions are not sufficient to ensure that all systems pointers to the data incorporated in the system software have also been destroyed. For audio-visual tapes a ‘magnetic field bulk eraser’ should be used to degauss the tape (i.e., remove the recording). At the time of destroying confidential data and records, researchers should ensure that they employ the most effective method since this may change over time with technological advances.

For PoCoG administered studies, the destruction of research data and records should be authorised by the Executive Director, on recommendation of the PI. A record of the recommendation and approval must be maintained. Note that most institutional archives have an established process for data destruction, and in many cases the archive staff will take responsibility for the destruction of the data. It is recommended that you contact your organisation's archives department to confirm how they manage the data.

Some data may be considered for permanent retention. This usually occurs for data that:

- Is controversial or of high public interest
- Would be costly or impossible to reproduce
- Relates to the use of or supports the development of an innovative intervention
- Supports a patent application or other services
- Has long-term heritage, historical or cultural value
- Is of significant value to other researchers

2. Safety procedures and records management

Refer to section 1.3, 1.4, 1.6 and 1.9 of this document.

In regards to personnel safety, refer to your organisation's occupational health and safety guidelines. Additionally, all staff should be trained for the tasks they are expected to conduct. If staff are exposed to or managing upsetting information that may result in distress, they should have access to specific support and supervision to prevent, plan for and manage the distress.

3. Quality Assurance (QA) – templates, forms and checklists

3.1. Template: Data Management Plan

The purpose of this template (SOP_4.3.1) is to assist in the development of a data management plan which describes how the data will be managed throughout the duration of the study.

3.2. Template: Data Dictionary

The purpose of this template(SOP_4.3.2) is to assist in the development of a data dictionary to describe all the data that is collected for the purposes of this study.

3.3. Template: Participant Outcomes Database (PODB) in SPSS

This is an SPSS database template (SOP_4.3.3) for the collection of participant reported outcomes. It includes PoCoG's minimum data set.

3.4. Template: Participant Outcomes Database in Excel

This is an Excel spreadsheet template (SOP_4.3.4) for the collection of participant reported outcomes. It includes PoCoG's minimum data set.

3.5. Template: Participant Database in Access

This Access database template (SOP_4.3.5) has been designed for the collection of participant contact information and to log the completion of questionnaires.

3.6. Template: Archive Box Labelling Template

The Archive Box Labelling Template (SOP_4.3.6) provides a standard label for archive boxes.

3.7. Template: Archives Register Template

This document (SOP_4.3.7) has been designed to log the location of documents that have been archived.

3.8. Form: Data Destruction Form

This form (SOP_4.3.8) is to be completed for PoCoG studies for approval from the Executive Director if archived data is to be destroyed.

3.9. Checklist: Data Management Checklist

A checklist (SOP_4.3.9) of the items to be included in the data management plan.

3.10. File Hierarchy

This documents a suggested filing system for a research study (SOP_4.3.10).

3.11. Template: Archive metafile

This document is a template for listing details of archived files.

4. References

http://www.dama.org/files/public/DI_DAMA_DMBOK_Guide_Presentation_2007.pdf "DAMA-DMBOK Guide (Data Management Body of Knowledge) Introduction & Project Status"

http://www.unimelb.edu.au/records/research.html Policy on the Management of Research Data and Records