STANDARD OPERATING PROCEDURE 15 part 4

Extraction of Data for Analysis and Data Lock

Version:	3.0	Effective Date:	16 Apr 2024
Issue Date:	09 Apr 2024	Review Date:	16 Apr 2026
Author:	Jill Wood, Quality Assurance (QA) Manager, Warwick Clinical Trials Unit (WCTU)		
WCTU Reviewers:	Louise Hiller, Associate Professor, WCTU		
Sponsor Reviewers:	Mathew Gane, Research Governance & QA Manager, Research & Impact Services (R&IS)		
WCTU approval:	Natalie Strickland, Head of Operations, WCTU		
Sponsor approval:	Carole Harris, Assistant Director, R&IS (Systems & Strategic Projects) & Head of Research Governance		
Review Lead:	WCTU Statistics Team		

Contents

Revision Chronology:	Effective date:	Reason for change:
V3.0	16 Apr 2024	Minor amendments to working around data cleaning expectations for the DMC to align with current version of SOP 8 on Statistical procedures. Biennial review date will remain the same.
V2.0	14 Feb 2023	Biennial review: Inserted guidance for datasets that are not held within an application maintained by the WCTU Programming Team. Minor readability changes and updated links.
V1.0	14 Dec 2020	New SOP.

STANDARD OPERATING PROCEDURE 15 part 4
Extraction of Data for Analysis and Data lock

1. Purpose and Scope

The purpose of this Standard Operating Procedure (SOP) is to describe the procedure for documenting the extraction of datasets from a study database. The SOP also details the checks and authorisations required to complete the final lock of the clinical study database. This SOP applies to all studies sponsored by the University of Warwick, or where the data management and statistical analysis have been delegated to WCTU by an external sponsor.

2. Definitions

Data freeze:	A picture of the data in the database that has not been subject to the full cleaning and checking process. Provides a view of the database as it currently is. This is often used for trial runs of analysis or to enable data cleaning activities or to get an idea of data completeness or quality. This is sometimes referred to as soft lock. Data freezes can be stored and used for routine reporting.
Data snapshot:	A picture of the data in the database that has been subject to cleaning and checking process. The datasets are stored and can be used for interim analyses and/or presentation/publication.
Data lock:	Noun: A formal final picture of the data in the database. Verb: The process whereby a database or subsections of a database containing cleaned and validated data have all edit permissions revoked to ensure the dataset has a constant state. No further changes can be made to the underlying data, nor can the datasets be exported without appropriate justification and authorisation being sought.
Dataset:	The data content or subset of data in a database.
Cleaned dataset:	A dataset that has been checked for quality and completeness and has had errors rectified.
Database:	Repository of data associated with a clinical research study. System used to enter, manipulate, and extract data is often referred to as the database application.
Interim analysis:	Any formal analysis that occurs prior to the final analysis of a trial.
Unblinding:	In the context of this SOP, unblinding refers to revealing the blinded allocation of each participant in the dataset to perform unblinded analysis.
End of trial declaration:	End of trial should be declared to the Research Ethics Committee (REC) and Medicines and Healthcare products Regulatory Agency (MHRA) (if appropriate), within 90 days after the definition of end of trial that is stated in the protocol for the study (usually last patient, last visit/data collection time point). It is expected that final analysis and results are available within a year of this date.

3. Background

The aim of any data management process is to provide a high-quality and appropriately clean final dataset that is suitable for statistical analysis. There should be a process for controlling this. Good

Clinical Practice (GCP) guidelines describe how it should be clear when a dataset or database is declared to be final and what checks were made to make that decision. The guidelines also state that storage and protection of 'final' datasets should be clear.

There are several reasons why datasets may need to be extracted from the study database, these include:

1. Analysis for final clinical study report
2. Protocolised Interim analyses
3. Making formal decisions about a study e.g. dose escalation
4. Publication in a peer-reviewed journal
5. Presentations at conferences
6. Making decisions and recommendations about the suitability of continuing a study
7. Pre-agreed data requests from funders or collaborators e.g. Adverse Event (AE) line listings
8. Preparation of annual reports e.g. Developmental Safety Update Reports (DSURs)
9. Quality Control checks on data entry
10. Preparation of reports for Data Monitoring Committee or Trial Steering Committee meetings

Clearly documenting the data lock procedures is important for maintaining the integrity of the final dataset by preventing unauthorised or unintended changes. In blinded Randomised Controlled Trials (RCTs), documentation of when the data lock was performed is critical for demonstrating the integrity of the blinding.

4. Procedure

4.1 Responsibilities

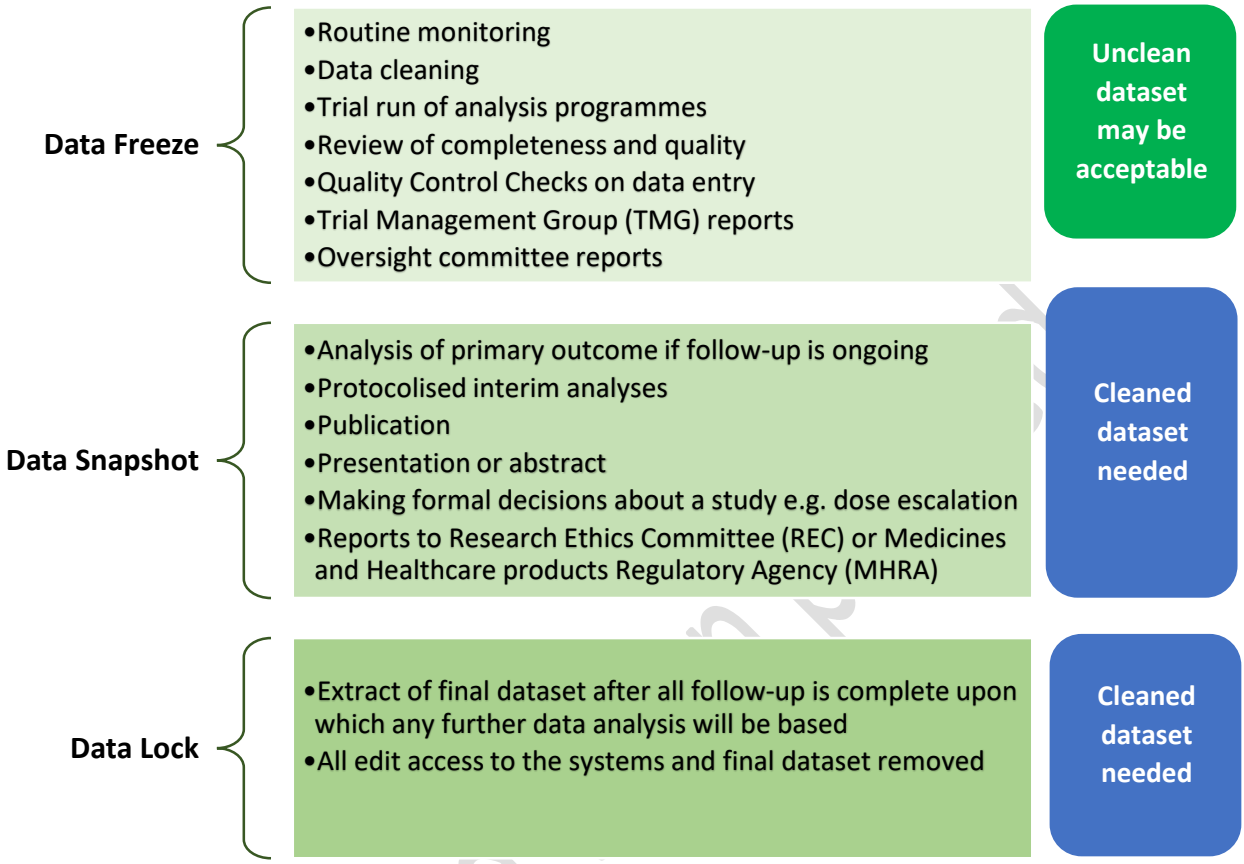
Statistician/Health Economist:	Ensuring full capture of all data related to trial CRFs, coordination of the data cleaning process, analysis of the data and ensuring the datasets are appropriately stored and protected.
Trial Manager/Coordinator (TM/TC) or Senior Project Manager (SPM):	Ensure the integrity of the data, that all reasonable efforts to clean have been carried out prior to data snapshot or data lock. Ensure the 'Data snapshot/lock Confirmation Form' has been completed and signed prior to the data snapshot or lock.
Chief Investigator (CI):	Ensuring adequate checks have been done prior to data snapshot or lock. Sign off **T46 - Data snapshot/lock Confirmation Form.**
System Owner:	Named owner of the database application responsible for submitting request to revoke access upon receipt of signed copy of **T46 - Data snapshot/lock Confirmation Form**.
Programming Team:	Revoking access to the clinical study web application upon the receipt of a WCTU Programming HelpDesk request from system owner.
Governance Committee:	To receive and review requests to unlock where there are exceptional circumstances.

4.2 When?

At any point when data are extracted from the clinical database (or other data sources) with the intention to use a dataset for any of the purposes stated above in section 3.

4.3 How?

Below is a summary and details can be found in the corresponding sections:



4.3.1 Data freeze

For instances where data are required for routine processes, a data freeze is an appropriate method for a statistician, health economist or other researcher to access data from the database. Good practice in line with relevant SOPs and the UK General Data Protection Regulation (UK GDPR) should be undertaken and no external formal publication or presentation from such datasets should be made. Data freezes can be used for routine reporting purposes, but it is good practice to take a Data Snapshot (see below). It is also good practise to provide audit trails for any data freezes that are used for routine reporting purposes. They should be clearly named and stored in accordance with Information Security Guidelines within the Statistics area of the WCTU Mdrive to provide an audit trail to enable reconstruction of analyses and timelines.

4.3.2 Data Snapshot

For instances where cleaned data are required for pre-specified analyses, publications etc., a snapshot of the data is taken from the database and clearly named and stored in accordance with Information Security Guidelines within the Statistics area of the WCTU Mdrive to provide an audit trail to enable reconstruction of analyses and timelines. Folder or file names should include the study name, purpose and extract date. **T46 – Data snapshot/lock Confirmation Form** needs to be completed for each data snapshot.

4.3.3 Data Lock

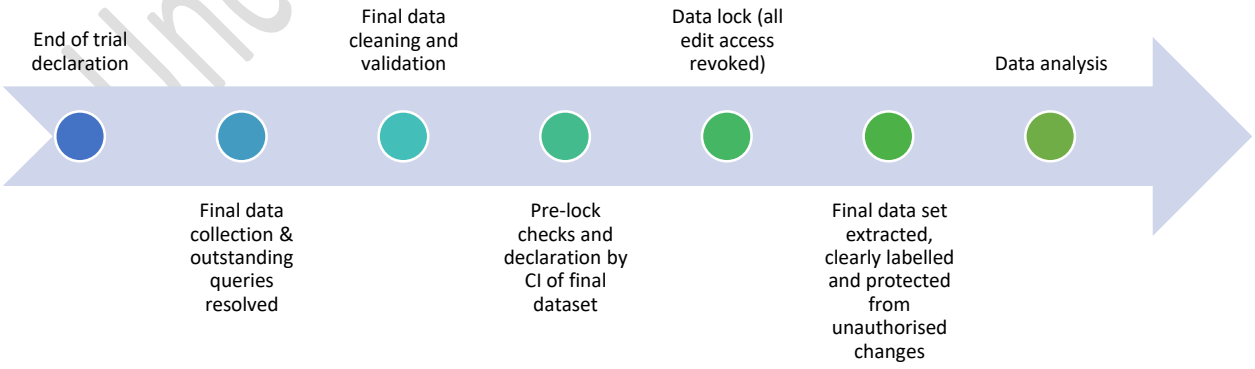
The figure below shows the order of activities from end of trial declaration up to the point of data analysis to ensure the integrity of the study results. Evidence that the process below occurred in the correct order should be available in the S/TMF. Evidence should be documented using **T46 -Data snapshot/lock Confirmation** form associated with this SOP, which records the following:

- Confirmation of when pre-lock checks were performed
- Confirmation by the CI of when data were declared 'final'
- Confirmation of when removal of edit access was completed = Date of data lock
- Location and file name of the final dataset

File/folder naming for the final dataset should make it clear that it is the final dataset and for which study. The file should be stored in accordance with Information Security Guidelines within the WCTU Mdrive (see part 1 of this SOP 'Security, Protection and Management of Data for Clinical Research'). The final dataset should also have sufficient protection from being edited or deleted. This could take the form of restricted folder access or password protection on the file. Where unblinding is applicable to a study, an additional unblinding authorisation form should be completed (see SOP 41 'Blinding and Unblinding in Research Studies').

At the point of data lock, as many queries should have been resolved as possible with the trial resources available, prioritising in a risk proportionate way. If there is justification for some queries to remain unresolved, or despite escalation some have received no response, justification/explanation should be documented on the confirmation form.

If data are held in repositories outside the trial database and will be used for the data analysis, then this should be made clear in the Trial Master File (TMF) Index or Map so that the primary data source can be located and secured. Where there are different phases of the study, data lock should be performed in sub-sections or a partial lock. If possible, edit access should be removed from these subsections. If this cannot be done, the date the data management team were informed not to edit should be recorded on **T46-Data snapshot/lock Confirmation form**. When the final lock occurs, audit trail reports can be checked to confirm no further edits have been made. A record of this check should be made available in the S/TMF.



4.3.4 Unlocking the final dataset

There may be exceptional circumstances where errors or inconsistencies are noted after the final data lock which may be critical for the integrity of the final dataset. In these circumstances, there may be sufficient justification to temporarily unlock the database to correct the data.

In these cases, **T47 - Request to unlock Form** should be completed and sent to the **QA team** who will organise for review and authorisation of the request by the **Governance Committee**. The decision to unlock will consider risk to bias and impact on the integrity of the trial results and the justification documented on the request form. Decisions will be recorded in the minutes of the Committee or via email correspondence.

A record of the changes made during the unlock period should be recorded on the Request to unlock Form and available in the S/TMF.

List of abbreviations

AE	Adverse Event
CI	Chief Investigator
DSUR	Developmental Safety Update Report
GCP	Good Clinical Practice
GDPR	General Data Protection Regulation
MHRA	Medicines and Healthcare products Regulatory Agency
QA	Quality Assurance
RCTs	Randomised Controlled Trials
REC	Research Ethics Committee
R&IS	Research & Impact Services
SOP	Standard Operating Procedure
TM/TC	Trial Manager/Coordinator
TMF	Trial Master File
TMG	Trial Management Group
WCTU	Warwick Clinical Trials Unit

Template Documents

T46 – Data snapshot/lock Confirmation Form

T47 – Request to Unlock Form