UNIVERSITY OF WARWICK

STANDARD OPERATING PROCEDURE 15 Part 4

Extraction of Data for Analysis and Data Lock

Version:	4.0
Approval Date:	08/07/2026
Effective Date:	22/07/2026
Review Date:	21/07/2028
Review Lead:	Jo Grumett, QA Manager, WCTU
WCTU Reviewers:	Claire Jacques, Trial Manager, WCTU Anower Hossain, Assistant 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, Associate Director Research Culture, Governance and Compliance, R&IS

Revision Chronology:	**Effective date:**	**Reason for change:**
V4.0	22 Jul 2026	Biennial review. Change to branding & review of text to remove duplication.
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.

Uncontrolled when printed

UNIVERSITY OF WARWICK

CONTENTS

1 PURPOSE AND SCOPE 3
2 ACRONYMS 3
3 DEFINITIONS 3
4 BACKGROUND 4
5 RESPONSIBILITIES 4
6 WHEN 5
7 HOW 5
7.1 DATA FREEZE 6
7.2 DATA SNAPSHOT 6
7.3 DATA LOCK 6
7.4 UNLOCKING THE FINAL DATASET 7
8 ASSOCIATED TEMPLATES AND DOCUMENTS 7

Uncontrolled when printed

1 Purpose and Scope

To describe the procedure for documenting the extraction of datasets from a clinical study database. The SOP 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, as well as studies for which data management and statistical analysis have been delegated to WCTU by an external sponsor.

2 Acronyms

AE	Adverse Event
CI	Chief Investigator
CDMS	Clinical Data Management System
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
QC	Quality Control
RCT	Randomised Controlled Trial
REC	Research Ethics Committee
R&IS	Research & Impact Services
SOP	Standard Operating Procedure
SPM	Senior Project Manager
TM/TC	Trial Manager/Coordinator
TMF	Trial Master File
TMG	Trial Management Group
UK GDPR	UK General Data Protection Regulation
WCTU	Warwick Clinical Trials Unit

3 Definitions

Data freeze:	A picture of the data in the database that has not been subject to the full cleaning and checking process. It provides a current view of the database in its present state and is used for trial analyses; to enable data cleaning activities; or to review data completeness or quality. 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. Snapshots 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, with any identified errors rectified.
Database:	Repository of data associated with a clinical research study. System used to enter, manipulate, and extract data may be referred to as the database application or CDMS.
Interim analysis:	Any formal analysis that occurs at any point prior to the final analysis of a trial.
End of trial declaration:	End of trial should be declared to the REC and MHRA (if appropriate), within 90 days after the protocol-defined definition of end of trial. It is expected that results of the final analysis are available within 12 months of this date.

4 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 in place to control this. GCP guidelines require clear criteria for when a dataset is declared final, including the checks performed to support this, and stipulate that the storage and protection of final datasets must be clearly defined.

Clearly documenting the data lock procedures is important for maintaining the integrity of the final dataset by preventing unauthorised or unintended changes. Documentation will provide an audit trail to demonstrate reconstruction of analysis and timelines.

In blinded RCTs (See SOP 41: Blinding and Unblinding in Research Studies), documentation of when the data lock was performed is critical for demonstrating the integrity of the blinding.

Reasons for dataset extraction from the database 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. Conference presentations
6. Making decisions and recommendations about the suitability of continuing a study
7. Pre-agreed data requests from funders or collaborators e.g. AE line listings
8. Preparation of annual reports e.g. DSURs
9. QC checks on data entry
10. Preparation of reports for Data Monitoring Committee or Trial Steering Committee meetings

5 Responsibilities

Statistician/Health Economist:	Ensure capture of all data related to trial CRFs, coordination of the data cleaning process, analysis of the data and ensuring datasets are appropriately stored and protected.
TM/TC or SPM	Ensure data integrity and that all reasonable efforts to clean have been carried out prior to data snapshot or data lock. Ensure **T46 - Data**

Uncontrolled when printed

	snapshot/lock Confirmation Form has been completed and signed prior to the data snapshot or lock.
CI	Ensure 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 user edit permissions following authorisation of data lock.
Programming Team:	Revoke user edit permissions to the clinical study 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.

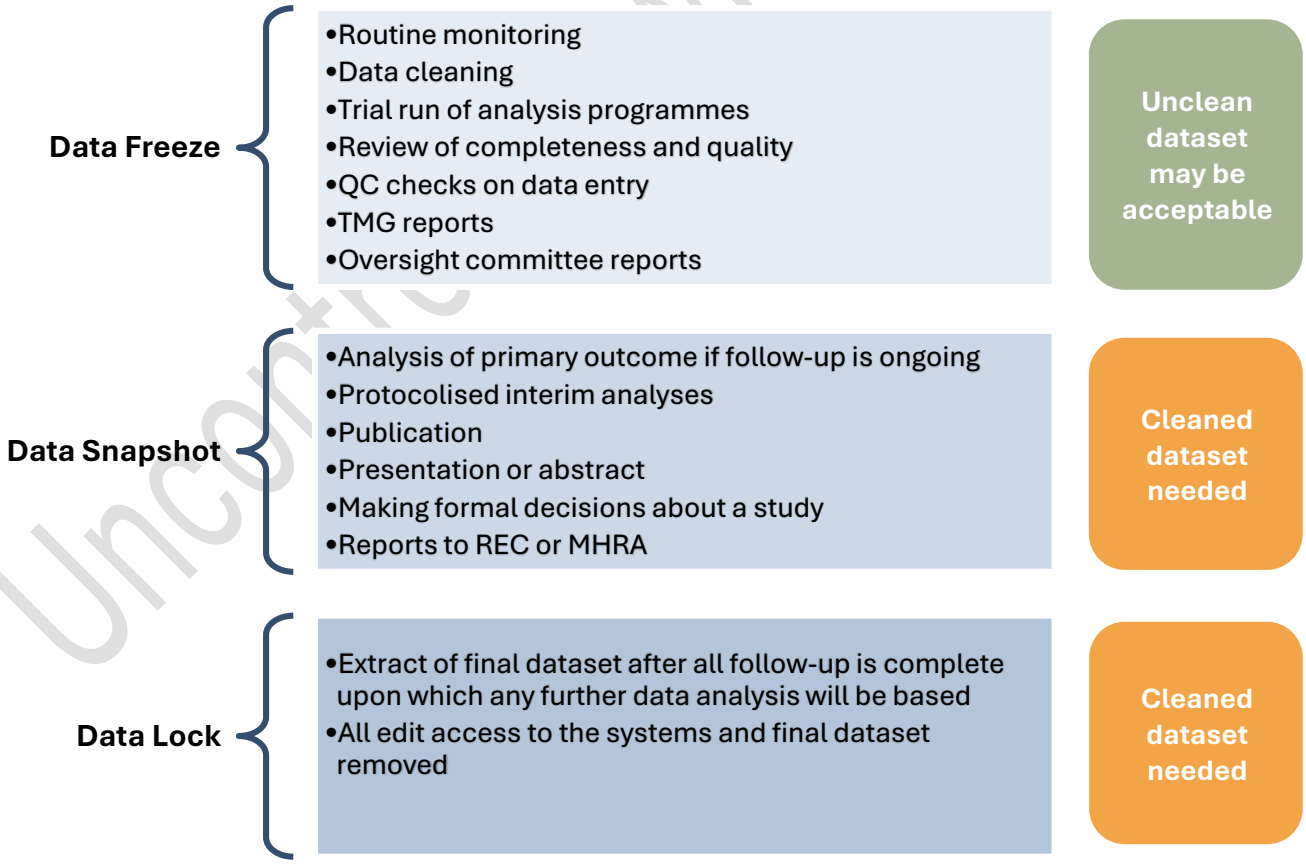
6 When

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

For WCTU studies, the process, timelines and required deadlines for working with the Programming time should be defined and agreed.

7 How

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



Uncontrolled when printed

7.1 Data freeze

Where data are required for routine processes, this 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 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 practice to provide audit trails for data freezes used for routine reporting purposes. They should be clearly named and stored in accordance with Information Security Guidelines. For WCTU studies, storage should be within a pre-agreed and secure Statistics/Health Economics location (e.g. M:Drive).

7.2 Data Snapshot

For instances where cleaned data are required for pre-specified analyses, publications etc., a snapshot is taken from the database and clearly named and stored in accordance with Information Security Guidelines. For WCTU studies, storage should be within a pre-agreed and secure Statistics/Health Economics location (e.g. M:Drive). The 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.

7.3 Data Lock

The flowchart 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 TMF. Evidence should be documented using **T46** -Data snapshot/lock Confirmation form 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. For WCTU studies, storage should be within a pre-agreed and secure Statistics/Health Economics location (e.g. M:Drive). 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.

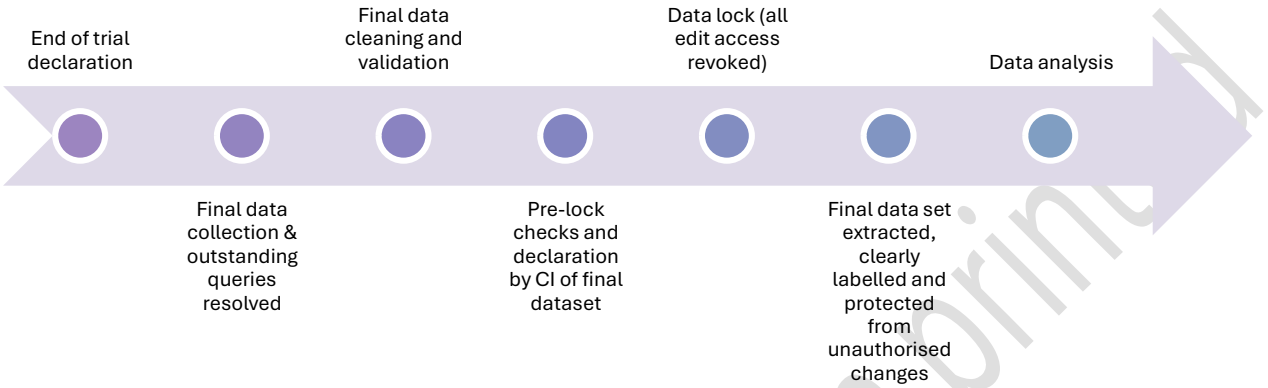
Where unblinding is applicable to a study, an additional unblinding authorisation form should be completed (see SOP 41: Blinding and Unblinding in Research Studies).

At the point of data lock, as many queries should have been resolved as possible with the trial resources available, prioritised 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 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

Uncontrolled when printed

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 TMF.



7.4 Unlocking the final dataset

There may be exceptional circumstances where errors or inconsistencies are identified after the final data lock which may be critical for the integrity of the final dataset. In such 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 submitted to the QA team who will facilitate review and authorisation of the request by the Governance Committee. The decision to unlock the database will take into consideration the risk to bias, the 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 and/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 TMF.

8 Associated Templates and Documents

T46	Data snapshot/lock Confirmation Form
T47	Request to Unlock Form

Uncontrolled when printed