

Standard Operating Procedure CCTU/SOP057

Providing a Dataset from a Clinical Trial Database

1. Scope

This SOP applies to datasets obtained from databases used for the capture of clinical trial data from Cambridge Sponsored CTIMPs or trials managed by the CCTU .

In most EDC systems, it is possible to view and export data via the data entry interface. Data accessed via this mechanism is not included within the remit of this document.

2. Purpose

The purpose of this document is to provide details of how datasets from clinical trial databases should be accessed, controlled and protected to ensure data is managed and reliable for its intended uses. It also includes details on how datasets should be provided to consumers of the data. These procedures should be followed irrespective of the database in use.

These uses are limited to;

- Datasets required during database locking
- Datasets used for interim analysis
- Datasets used for formal purposes e.g. for IDMC
- Snapshots of data obtained via a restricted interface that standard users of the system cannot access – for example, in MACRO™ only users with access to the Query module

The principles addressed are;

- The process is controlled
- The reason for producing a dataset governs how the dataset export should be made, how the dataset should be protected and how access to the dataset should be controlled
- Evidence of the processes followed are retained

3. Definitions and Abbreviations

The headings below contain the definitions of terms and meaning of abbreviations used within the document.

3.1. Definitions

Term	Definition
Cambridge Sponsored	Sponsored by Cambridge University Hospitals NHS Foundation Trust (CUH); or the University of Cambridge (UoC); or jointly by CUH and UoC OR Sponsored by: Cambridge University Hospitals NHS Foundation

	Trust (CUH) or CUH jointly with the University of Cambridge or Cambridgeshire & Peterborough NHS Foundation Trust (CPFT) or CPFT jointly with the University of Cambridge
Data consumer	An entity receiving and using a dataset. The entity is usually a person (such as a statistician) who will use software tools to manipulate the data in order to meet a specified objective – for example, providing an interim analysis report.
Database lock	Refers to the process whereby a clinical trial database has data cleaned and validated and all edit permissions revoked to ensure the dataset has a constant state. No further changes can be made to the underlying data system or exported dataset without appropriate authorisation being sought.
Clinical trial database	A repository of data associated with a clinical trial
Dataset	Refers to the data content of the clinical trial database
Checklist	The list of checks during the dataset export that itemises what processes should be undertaken in order to provide assurance of the quality of the dataset
Cleaned dataset	A dataset that has been fully cleaned and checked and for which there is documented evidence of this checking by the use of a checklist
Snapshot	A dataset that has not been subject to the full cleaning and checking process and provides a view of the database as it currently is
CRF Generator	A bespoke CUH process using XML that allows users to design the data capture for a trial and enables the rapid development of paper CRF and electronic database
Subversion	A software versioning and revision control system used to maintain current and historical versions of code and associated files
TortoiseSVN	An Apache™ Subversion (SVN) client®. This Windows client provides a Graphical User Interface (GUI) for the underlying Subversion system.
Interim Analysis	Protocol defined analysis for which datasets are required
Checksum	A method by which a small datum is recorded for the purpose of detecting errors that may have been introduced during the transmission or storage of data.

3.2. Abbreviations

Abbreviation	Meaning
CI	Chief Investigator
pCRF	A paper CRF
EDC	Electronic data capture system e.g. MACRO™
CUH	Cambridge University Hospitals NHS Foundation Trust
XML	Extensible Mark-up Language
IDMC	Independent Data Monitoring Committee
TSC	Trial Steering Committee
DSUR	Drug Safety Update Report

4. Undertaken by

The export processes are primarily undertaken by trial team members involved in providing datasets for database locks, interim analysis, other formal purposes and snapshots. These are mainly Data Managers and Programmers although anyone who has the required permissions to perform an export should also follow the procedures documented in this SOP.

Data requestors are primarily Data Managers, Clinical Trial Administrators or Coordinators

5. Items Required

- CCTU/FRM093 Data Export Checklist
- CCTU/FRM094 Database Lock Request and Data Extract Authorisation
- CCTU/FRM095 Interim Data Request

6. Summary of Significant Changes

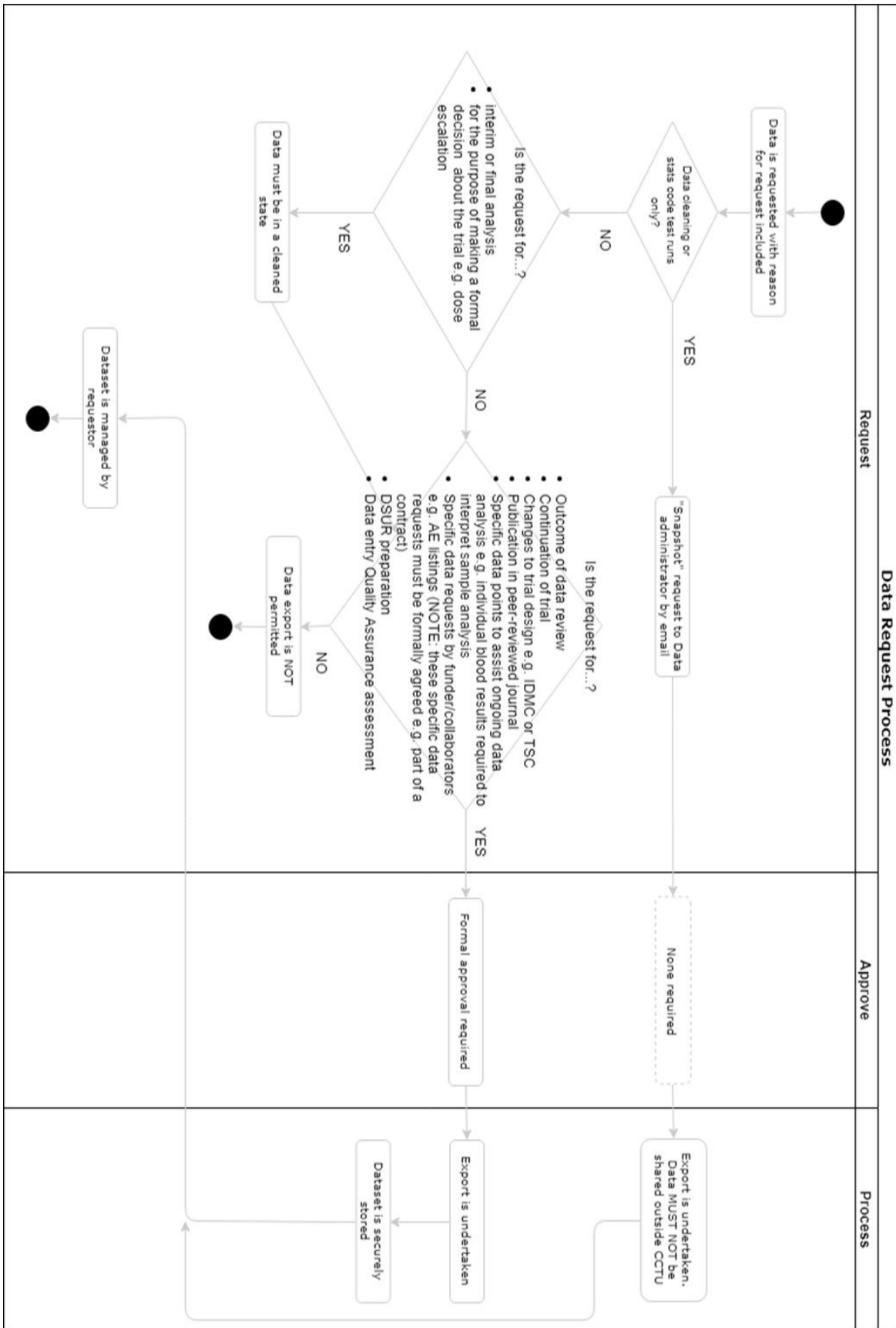
Revision of method

Inclusion of Data request swimchart for clarification of the process

7. Method

The following sections provide a description of the processes to be followed when implementing this document's procedures.

The swimchart below provides an overview with further detail provided in subsequent text.



7.1. Reasons for requiring dataset

A request for a dataset must be accompanied by a reason for the request.

Requests will be handled one of three ways;

1. As a request for a cleaned dataset, where cleaning is documented using CCTU/FRM093 Data Export Checklist and other documents – for example, documents that provide evidence that trial specific procedures were followed.

The following requests always require a cleaned dataset;

- Database locking
 - Formal Interim analysis
 - For the purpose of making a formal decision about the trial e.g. dose escalation
2. As a request where the dataset does not require formal cleaning although is required for a specific procedural purpose – these requests are referred to as 'formal' requests ;
 - Continuation of trial
 - Changes to a trial design e.g. IDMC or TSC
 - Publication in a peer-reviewed journal
 - Specific data points to assist ongoing data analysis e.g. individual blood results required to interpret sample analysis
 - Specific data requests by funders/collaborators e.g. AE listings (note: these requests must be formally agreed e.g. as part of a contract)
 - DSUR preparation
 - Data entry Quality Assurance assessment
 3. As a request for a snapshot dataset where the data in the database is exported 'as is' for the purpose of cleaning it or for the use by Statistics to test their code

If a request is made for any purpose other than those listed above then the export should NOT be made without prior confirmation from the Statistician.

7.2. Appropriate authorisation must be given

7.2.1. Permanent Extraction

All requests for a cleaned dataset for database locking or for any type of analysis should be accompanied by the appropriate authorisation;

- Database locking - use CCTU/FRM094 Database Lock Request and Data Extract Authorisation
- Interim analysis or for any of the formal purposes listed in section 7.1 bullet point 2 - use CCTU/FRM095 Interim Data Request

For all instances of authorised data extractions a permanent copy of the data should be retained. The CI should not have access to these datasets without the prior agreement of the Statistician.

7.2.2. Temporary Snapshots

The authorisation forms may be omitted only for data snapshots taken to aid data cleaning or for the purpose of testing Statistics code. In these instances

the snapshot can be requested by a suitably delegated member of the trial team who is employed by, or has an honorary contract with the Trust. The dataset should be suitably secured and must not be shared outside the confines of the CCTU trial team. The CI may not access or use a snapshot.

If the data extractor has reservations about the legitimacy of the request they should seek confirmation from the trial statistician and wider trial team.

7.3. It is the responsibility of the requestor to ensure authorisation has been given and documented. The disseminator should attach any caveats for data use at the time of transfer.

7.4. Preparing the data

For database locking, formal interim analysis or any formal decision making about the trial the trial team will undertake to provide a 'clean' dataset by following the necessary procedures as detailed in the Data Management Plan.

7.5. Checklist

Every trial should have a checklist that helps the data managers undertaking the cleaning process. Refer to CCTU/FRM093 Data Export Checklist for a default list of checks that can be followed. The list can be added to as required for each trial and consistently followed for every export requiring a dataset clean.

The checklist must be signed and dated by the completer and appropriately filed.

The checklist is not required for data snapshots or for formal requests not requiring a cleaned dataset.

7.6. Running the export

The export must be undertaken in accordance with the request made.

If the export is part of a database lock or an interim analysis, the whole dataset should be exported.

For snapshots or other formal purposes, the trial team may specify a subset of data for export or request all data. The data must be provided in a consistent manner and in accordance with the specifications of the request.

In MACRO™, it is possible to save the settings for an export. Wherever possible, the same settings should be used to ensure a consistent format for consumers of the data.

7.6.1. File naming conventions

The files created should identify the trial and date of export. If more than one format of export is undertaken, it should be reflected in the name – e.g. TRIALNAME_20160601_sas.txt. Naming conventions should be consistently applied.

For long term storage, file names should also document the type of export. When zipping the data (see below), the zipped file name can be used to determine the export type – e.g. TRIALNAME_20160601_hardlock.zip, rather than specifying the type in each individual file name (which is also acceptable).

7.7. Disseminating the dataset

The data should be zipped and password protected using a strong password. Refer to <https://strongpasswordgenerator.com/>).

If disseminated using email ensure that confirmation of receipt is received from the requester. Once confirmation is received the password can be sent in a separate email to the original requestor.

Alternatively a secure network location may be used. The data extractor should have read/write permissions to this location and consumers of the data should be restricted to read permissions only.

The sender should be satisfied that the receiver is storing the data in a controlled manner in accordance with good data handling practices, and the controls in place up until this point are not negated as a result of sending it on. Once sent, the receiver is responsible for the appropriate handling of the dataset and the sender no longer maintains that responsibility.

7.8. Storing the dataset

Extracted data must be stored in a controlled location and kept securely. As a minimum, datasets related to database locking, interim analysis and formal requests must be retained and appropriately stored. It is good practice to maintain a checksum of the dataset to aid verification of the integrity of the file at a later date.

For trials where data design is using TortoiseSVN, the password protected zip file should be placed in the repository and a separate, protected record of the trial name, query file name, location of data file and password should be maintained in order that the file can be accessed if necessary. The location of this record should be protected.

Whatever method is used it is essential that the master copy of the data cannot be edited and is protected from changes and accidental or deliberate deletion.

8. Monitoring Compliance with and the Effectiveness of this Document

a. Process for Monitoring Compliance and Effectiveness

As part of routine monitoring visits, audit and inspection

b. Standards/Key Performance Indicators

This process forms part of a quality management system and is reviewed according to CCTU procedures. Standard Operating Procedures are reviewed every two years.

9. References

The Institute of Clinical Research, Abbreviations used in Clinical Trials.
MHRA, Good Clinical Practice "Grey Guide"

10. Associated Documents

NA

11. Equality and Diversity Statement

This document complies with the Cambridge University Hospitals NHS Foundation Trust service equality and diversity statement.

12. Disclaimer

It is the user's responsibility to check against the electronic library that this printed out copy is the most recent issue of this document.

Review date	2 years (or earlier in light of new evidence) from approval date
Owning department:	CCTU QA
Supersedes:	CCTU/SOP057 V2
Local reference:	CCTU/SOP057 V3