ECTU Central Office SOP_DM_11: Data Cleaning

Version No:	2.0
Issue Date:	19 Feb 2026
Effective Date:	05 Mar 2026

Authorship and Approval			
Name and Designation	Author/Reviewer /Approval/ Authorisation	Date	Signature
Lynsey Milne, Data Manager	Author	19-Feb-2026	Lynsey Milne (19-Feb-2026 12:28:45 GMT)
Ruth Armstrong, Data Manager	Reviewer	19-Feb-2026	Ruth Armstrong (19-Feb-2026 13:19:44 GMT)
Joyce Thomson, Chief Operating Officer	Approver	19-Feb-2026	Joyce Thomson (19-Feb-2026 12:51:14 GMT)
Tanya Tharakan QA Manager	QA Authorisation	19-Feb-2026	Tanya Tharakan (19-Feb-2026 13:38:45 GMT)

Document Revision History		
Version No.	Effective Date	Summary of Revisions
1.0	04 Jul 2024	Initial Creation/New document
2.0	05 Mar 2026	Removal of reference to TM responsibilities regarding TSC/DMC from section 2

1.0 PURPOSE

This Standard Operating Procedure (SOP) describes the procedure for cleaning study data at the end of study and prior to data analysis by the Trial Statistician or designee.

2.0 SCOPE

This SOP applies to all studies where the ECTU Data Management Team will be responsible for data cleaning at the end of study data analysis or for any interim data analysis by the Trial Statistician or designee. This includes cleaning of data on ECTU REDCap and bespoke databases.

Delegation of this task to the ECTU Data Management Team will be established at the start of the study as per ECTU Central Office SOP ECTU_DM_01 Data Management Procedures.

3.0 RESPONSIBILITIES

Where data management has been delegated to the ECTU Data Management & Programming Team, the designated Data Manager, Assistant Data Manager or designee will be responsible for implementing and maintaining the Data Cleaning Plan.

All data cleaning requirements will be referenced in the Data Management Plan (DMP)

4.0 PROCEDURE

4.1 Preparing for Data Cleaning

4.1.1 Data Cleaning consists of manual checks to ensure that study data is complete and accurate before any analysis takes place.

4.1.2 Study data must be cleaned at the end of the study. Cleaning must be completed before the database can be locked prior to final analysis.

4.1.3 Additional periods of data cleaning may be required if any interim data analysis is planned. This may only involve certain datapoints or participants within the database.

4.1.4 Depending on the size and nature of the study, a period of 1-2 days per week for 3-6 months is recommended for data cleaning. This is a guideline only and actual timeframes may differ according to the study.

4.1.5 A Data Cleaning Plan will be implemented to define the scope of the cleaning at the end of study or for interim analysis. The manual checks performed will also be described in this.

4.1.6 It is recommended that a discussion between the Trial Manager or designee and the Data Manager or Assistant Data Manager and Trial Statistician or designee on the end of study data cleaning starts once recruitment is complete as the last participant last visit (LPLV) date will then be known. However, this discussion may be required sooner depending on the nature of the study. The Trial Manager or designee will keep the Data Manager or Assistant Data Manager informed on recruitment, and trial dates relating to recruitment and follow-up.

4.1.7 Where interim analysis is planned, this should be detailed within the study protocol. A discussion on the interim data cleaning should take place as soon as possible, involving the statistician and unblinded statistician, allowing enough time for cleaning to be completed before any analysis is due to begin.

4.1.8 The Trial Manager or designee, the Data Manager or Assistant Data Manager and Trial Statistician or designee will agree to the terms of the end of study or interim data cleaning to be included in the Data Cleaning Plan.

4.2 Scope of Data Cleaning and Proposed Timeline

4.2.1 The scope of the data cleaning and the proposed timeline will be agreed and documented as detailed in section 4.1.8.

4.2.2 The scope of end of study cleaning will include the following as standard (where applicable to the study):

- Review of all Adverse Event data
- Review and reconciliation of all Adverse Event and Serious Adverse Event data
- Review of all Death data
- Review of all Change of Status data

4.2.3 In addition to the standard criteria, any data included in the Primary and/or Secondary Endpoints and Outcomes will also be included in the scope of the data cleaning.

4.2.4 Other data unrelated to the standard or endpoint/outcome checks may be included in the cleaning if required, and will be agreed within the Data Cleaning Plan.

4.2.5 The scope of interim study cleaning will depend on the requirements of the analysis, as specified in the Statistical Analysis Plan, and agreed within the Data Cleaning Plan. For interim study cleaning, the participants who are included in the cleaning will depend on the requirements of the analysis.

4.2.6 For end of study data cleaning, all participants on the study database will be included in the cleaning. All participants may not require the full scope of the data cleaning depending on their status at the end of study. For example, a participant may withdraw prior to the collection of endpoint data.

4.2.7 The proposed start and end date of the data cleaning period will be agreed and specified within the Data Cleaning Plan before the cleaning begins. It is important to consider any publication deadlines, workload restrictions and the analysis period when agreeing the completion deadline for the data cleaning.

4.2.8 Where possible, it is recommended that a two-week buffer period is given between the end of data cleaning and database lock at the end of study or start of data analysis period to allow the Trial Statistician or designee to review the data and identify any additional issues to be addressed prior to database lock or the start of analysis. This period should be allowed in addition to the agreed data cleaning period.

4.3 Creating a Data Cleaning Plan and Completing and Documenting Data Cleaning Findings

4.3.1 A Data Cleaning Plan will be implemented for both end of study cleaning and interim cleaning (where applicable) before the data cleaning process begins. (See sections 4.1.4 and 4.1.6)

4.3.2 Template DM012 Data Cleaning Plan will be used to create the Data Cleaning Plan.

4.3.3 The Data Cleaning Plan will be approved by the Trial Manager or designee, Data Manager or Assistant Data Manager and Trial Statistician or designee.

4.3.4 Template DM014 Data Cleaning Log will be used to document any errors, issues and findings identified during the data cleaning. This is a living document that is continually updated throughout the cleaning process.

4.3.5 All documentation relating to data cleaning should be retained in a designated Data Cleaning folder within the Data Management section of the study-specific folder on the ECTU shared drive.

4.3.6 Further instruction on creating a Data Cleaning Plan and documenting data cleaning findings is detailed in ECTU Central Office WPD ECTU_DM_W2 Data Cleaning Procedures.

4.4 Completion of Data Cleaning

4.4.1 Data Cleaning will be considered to be complete when:

- All issues identified and documented in the Data Cleaning Log have been resolved or a non-resolution reason is documented
- All outstanding queries on the database have been closed either due to resolution of the issue or documenting a reason for non-resolution
- All missing data has been entered or if not obtainable, a reason for this is documented

4.4.2 Once the cleaning is complete, this will be confirmed to the Trial Manager or designee and/or Trial Statistician or designee by email.

4.5 Retention of Documents

4.5.1 A Data Cleaning folder will be created within the Data Management section of the study-specific folder on the ECTU shared drive**.**

4.5.2 The Data Cleaning folder will be used to save the following study-specific documents (sub-sections can be created as required):

- Data Cleaning Plan (all versions)
- Data Cleaning Log
- Emails (any emails that are relevant to the data cleaning process and documented in the Data Cleaning Log should be retained here)

- Reports (any reports obtained from the database or from the Sponsor should be saved here)
- Any other documents relevant to the data cleaning process

4.5.3 The signed Data Cleaning Plan can be printed and retained in the paper TMF, if required. The Data Cleaning Log, Emails and any reports will not be printed and held in paper versions as these may be particularly large files. A file note will be provided to document where these are retained on the drive.

5.0 RELEVANT DOCUMENTS AND REFERENCES

ECTU Website

- ECTU Central Office WPD ECTU_DM_W2 Data Cleaning Procedures
- ECTU_SOP_DM_06 Query and Missing Data Management

ECTU DataStore (ECT Unit\SOPs\Finalised SOP and WPD\DM\Supporting Documents and Templates\Current)

- DM012_Data Cleaning Plan
- DM014_Data Cleaning Log

ACCORD Website

- POL012 Data Management Policy

ECTU_SOP_DM_11 Data Cleaning v2.0

Final Audit Report 2026-02-19

Created:	2026-02-19 (Greenwich Mean Time)
By:	Tanya Tharakan (tanya.tharakan@ed.ac.uk)
Status:	Signed
Transaction ID:	CBJCHBCAABAAFy3ACT_tiRBkLVKDudjKjtiZHxZrQwDm

"ECTU_SOP_DM_11 Data Cleaning v2.0" History

Document created by Tanya Tharakan (tanya.tharakan@ed.ac.uk)
2026-02-19 - 12:24:02 GMT- IP address: 192.41.114.230

Document emailed to Lynsey Milne (Lynsey.Milne@ed.ac.uk) for signature
2026-02-19 - 12:26:31 GMT

Document emailed to Ruth Armstrong (Ruth.Armstrong@ed.ac.uk) for signature
2026-02-19 - 12:26:32 GMT

Document emailed to Joyce Thomson (joyce.thomson@ed.ac.uk) for signature
2026-02-19 - 12:26:32 GMT

Document emailed to Tanya Tharakan (tanya.tharakan@ed.ac.uk) for signature
2026-02-19 - 12:26:32 GMT

Email viewed by Lynsey Milne (Lynsey.Milne@ed.ac.uk)
2026-02-19 - 12:28:34 GMT- IP address: 104.47.11.126

Document e-signed by Lynsey Milne (Lynsey.Milne@ed.ac.uk)
Signature Date: 2026-02-19 - 12:28:45 GMT - Time Source: server- IP address: 192.41.114.230

Email viewed by Joyce Thomson (joyce.thomson@ed.ac.uk)
2026-02-19 - 12:50:40 GMT- IP address: 104.47.11.254

Document e-signed by Joyce Thomson (joyce.thomson@ed.ac.uk)
Signature Date: 2026-02-19 - 12:51:14 GMT - Time Source: server- IP address: 94.5.9.56

Email viewed by Ruth Armstrong (Ruth.Armstrong@ed.ac.uk)
2026-02-19 - 13:18:40 GMT- IP address: 86.167.119.238

Document e-signed by Ruth Armstrong (Ruth.Armstrong@ed.ac.uk)
Signature Date: 2026-02-19 - 13:19:44 GMT - Time Source: server- IP address: 86.167.119.238

Adobe Acrobat Sign

Document e-signed by Tanya Tharakan (tanya.tharakan@ed.ac.uk)

Signature Date: 2026-02-19 - 13:38:45 GMT - Time Source: server- IP address: 192.41.114.230

Agreement completed.

2026-02-19 - 13:38:45 GMT

Adobe Acrobat Sign