ECTU Central Office WPD ECTU_DM_W2: Data Cleaning Procedures

Version No:	2.0
Issue Date:	13 Jun 2024
Effective Date:	10 Jul 2024

Authorship and Approval			
Name and Designation	Author/Reviewer /Approval/ Authorisation	Date	Signature
Lynsey Milne Data Manager	Author	13-Jun-2024	*Lynsey Milne (Jun 13, 2024 11:25 GMT+1)*
Christopher Flitcroft-White, Data Manager - Digitally Enabled Clinical Trials	Reviewer	13-Jun-2024	*Christopher White (Jun 13, 2024 11:45 GMT+1)*
Kenton D'Mellow, Data Management and Programming Team Lead	Approver	13-Jun-2024	*Kenton D'Mellow (Jun 13, 2024 16:03 GMT+1)*
Tanya Tharakan QA Manager	QA Authorisation	13-Jun-2024	*Tanya Tharakan (Jun 13, 2024 16:09 GMT+1)*

Document Revision History		
Version No.	Effective Date	Summary of Revisions
1.0	21-Aug-2020	• Initial Creation
2.0	10 Jul 2024	• Updated to new SOP template • Change to title of WPD • Minor changes to section 1 • Section 2.1.2 and 2.1.3 as now included in SOP ECTU_DM_11 Data Cleaning • New section 2.2 added • Previous section 2.2 now section 2.3 and additional sub-sections included • Previous section 2.3 now included in SOP ECTU_DM_11 Data Cleaning

The user of this document is responsible for ensuring it is the current version.

1. INTRODUCTION

This Working Practice Document (WPD) provides guidance on preparing for, completing and documenting Data Cleaning as detailed in ECTU Central Office SOP ECTU_DM_11 Data Cleaning

2. INSTRUCTIONS and GUIDANCE

2.1 Creating a Data Cleaning Plan

2.1.1 Using template DM012 Data Cleaning Plan to create the study-specific plan, the following will be included:

- **Proposed Data Cleaning Deadline and Start Date**
  A deadline date for the cleaned data to be submitted for analysis should be agreed with the Trial Manager or designee and Trial Statistician or designee. This is not necessarily the date of database lock.

- **Standard Checks and Reviews**
  The datasets listed in this section are standard datasets that will be applicable to most studies. The specific checks listed may not always be required (for example, some may already be included as part of the query management system or may not be applicable- e.g. Cause of Death not required for the particular study). This list can be edited according to the study requirements. These checks – particularly reconciliation of pharmacovigilence data – are applicable at the end of study cleaning but may not always be applicable for interim data cleaning.

- **Outcome/Endpoint-specific Checks and Reviews**
  Each Primary and Secondary Endpoint and/or Outcome should be listed along with any checks and reviews that will be completed in order to clean the data. Where it is not possible to clean the data for this outcome (e.g. scan data), the reason for this should be specified. For example, at interim analyses it is possible that a proportion of endpoint data will not have been collected and therefore only specific outcomes will be included in the checks.

- **Other Checks and Reviews**
  If any other data reviews or checks are to be completed during data cleaning, these should be specified here.

2.1.2 Once agreed and approved, the Data Cleaning Plan is not subject to a formal review schedule but should be updated by the Data manager as required, if any of the agreed scope changes throughout the cleaning process.

2.2 Identifying Findings during Data Cleaning

2.2.1 The person who will complete the data cleaning should consider the best methods for cleaning each review parameter and this should be stated in the Data Cleaning Plan in the "Checks and Reviews to be completed" sections.

The user of this document is responsible for ensuring it is the current version.

ECTU WPD Identifier   ECTU_DM_W2
Version No   2.0
Effective Date   10 Jul 2024

2.2.2   Data is cleaned primarily by manually checking data on the study database for discrepancies or anomalies. This may involve obtaining reports from the study database and reviewing this to look for any errors. For example, a report of all routine blood results for all study participants may be obtained and the results reviewed to look for any results that may look anomalous compared to the accepted normal ranges for that blood result, the other results recorded for that participant and compared to other results recorded at that site.

2.2.3   When cleaning and reconciling pharmacovigilance data held by ACCORD, the pharmacovigilance manager should be informed of any findings and a course of action should be agreed. All emails relating to this should be retained in the study specific folder in the shared drive.

2.3   **Recording and Resolving Findings Identified during Data Cleaning**

2.3.1   All findings will be recorded using DM014 Data Cleaning Log Template. The Data Cleaning Details section will include:

- Study Name
- Data Cleaning Plan Version and Date (all versions used throughout the cleaning process should be listed)
- Data Cleaning Completed By (Name and Designation of all personnel who completed the cleaning)
- Date Cleaning Started
- Date Cleaning Completed and Notification Sent
- Index of review points

2.3.2   The log should be structured according to the Standard, Outcome/Endpoint-specific and Other Reviews as stated in the Data Cleaning Plan, with a separate tab used to document each review point.

2.3.3   Each finding should be documented as stated on the log, including Study No, CRF on which the finding was identified, who identified the error and a description of the finding.

2.3.4   As much as possible, manual queries should be added to the study database to resolve any findings with sites. This should be documented in the Action taken/required section and should include the date the query was raised. These queries will be sent to site as part of the normal query procedures for the study. Refer to ECTU_SOP_DM_06 Query and Missing Data Management

2.3.5   Where a finding requires further investigation or discussion instead of or before referring to the site, this should also be documented in the Action taken/required section. This may include for example, stating an email was sent to the Trial Manager to confirm the appropriate action to be taken. The date the email was sent should be included and the email retained the shared drive.

The user of this document is responsible for ensuring it is the current version.

2.3.6 Once the finding is resolved, the resolution should be documented in the Comments/updates section. For example, this may include that site corrected the data on the database and the manual query was closed (including date of closure) or that Trial Manager confirmed via email (including date of email) that finding is correct and no further action is required.

2.3.7 The Data Cleaning Log is a living document. A study-specific version of this log will be created, used and maintained throughout the cleaning process. The study-specific version is not subject to version control.

2.3.8 Once all data cleaning is complete (as detailed in ECTU Central Office SOP ECTU_DM_11 Data Cleaning), the date and notification details should be recorded in the Data Cleaning Details section.

3. RELEVANT DOCUMENTS AND REFERENCES

On ECTU Website
- ECTU Central Office SOP ECTU_DM_11 Data CleaningECTU_SOP_DM_06 Query and Missing Data Management

On ECTU Shared Drive
- DM012 Data Cleaning Plan
- DM014 Data Cleaning Log Template

ACCORD
- POL012 Data Management Policy

The user of this document is responsible for ensuring it is the current version.

ECTU Central Office WPD Template v3.0
Page **4** of **4**

ECTU_SOP_DM_11 Data Cleaning v1.0 and ECTU_WPD_DM_W2 Data Cleaning Procedures v2.0

Final Audit Report   2024-06-13

Created:	2024-06-13 (British Summer Time)
By:	Tanya Tharakan (tanya.tharakan@ed.ac.uk)
Status:	Signed
Transaction ID:	CBJCHBCAABAATmOe0ADecQ7FQa3BHfSE1r-LS62l8SGm

"ECTU_SOP_DM_11 Data Cleaning v1.0 and ECTU_WPD_DM_W2 Data Cleaning Procedures v2.0" History

📄 Document created by Tanya Tharakan (tanya.tharakan@ed.ac.uk)
2024-06-13 - 11:16:25 GMT+1- IP address: 192.41.114.230

✉ Document emailed to Lynsey Milne (Lynsey.Milne@ed.ac.uk) for signature
2024-06-13 - 11:24:28 GMT+1

✉ Document emailed to Christopher White (christopher.white@ed.ac.uk) for signature
2024-06-13 - 11:24:28 GMT+1

✉ Document emailed to Kenton D'Mellow (Kenton.D'Mellow@ed.ac.uk) for signature
2024-06-13 - 11:24:29 GMT+1

✉ Document emailed to Tanya Tharakan (tanya.tharakan@ed.ac.uk) for signature
2024-06-13 - 11:24:29 GMT+1

📄 Email viewed by Lynsey Milne (Lynsey.Milne@ed.ac.uk)
2024-06-13 - 11:25:03 GMT+1- IP address: 104.47.11.126

✒ Document e-signed by Lynsey Milne (Lynsey.Milne@ed.ac.uk)
Signature Date: 2024-06-13 - 11:25:46 GMT+1 - Time Source: server- IP address: 192.41.114.230

📄 Email viewed by Christopher White (christopher.white@ed.ac.uk)
2024-06-13 - 11:45:16 GMT+1- IP address: 104.47.11.126

✒ Document e-signed by Christopher White (christopher.white@ed.ac.uk)
Signature Date: 2024-06-13 - 11:45:48 GMT+1 - Time Source: server- IP address: 86.168.3.125

Email viewed by Kenton D'Mellow (Kenton.D'Mellow@ed.ac.uk)
2024-06-13 - 16:02:48 GMT+1- IP address: 82.132.219.222

Document e-signed by Kenton D'Mellow (Kenton.D'Mellow@ed.ac.uk)
Signature Date: 2024-06-13 - 16:03:13 GMT+1 - Time Source: server- IP address: 82.132.219.222

Document e-signed by Tanya Tharakan (tanya.tharakan@ed.ac.uk)
E-signature hosted by Tanya Tharakan (tanya.tharakan@ed.ac.uk)
Signature Date: 2024-06-13 - 16:09:19 GMT+1 - Time Source: server- IP address: 192.41.114.230

Agreement completed.
2024-06-13 - 16:09:19 GMT+1

Adobe Acrobat Sign