

ECTU Central Office WPD ECTU_DM_W2: Data Cleaning

Version No:	1.0
Effective Date:	21 Aug 2020

Authorship and Approval			
Name and Designation	Author/Reviewer/Approval	Date	Signature
Lynsey Milne, Assistant Data Manager	v1.0 Author	06 Aug 2020	See retained approval email dated 06 Aug 2020
Tony Wackett, Assistant Data Manager	v1.0 Reviewer	06 Aug 2020	See retained approval email dated 06 Aug 2020
Michelle Steven, Data Manager	v1.0 Approval	06 Aug 2020	See retained approval email dated 06 Aug 2020

Document Revision History		
Version No	Date	Summary of Revisions
1.0	21 Aug 2020	Initial creation/New document

1. INTRODUCTION

Data Cleaning is performed to ensure that the study data is accurate and robust prior to being used in any analysis. Where this has been delegated to the Data Management team, QC Checks form part of the Data Quality procedure outlined in ECTU Central Office SOP ECTU_DM_05 Data Quality Assurance. This Working Practice Document (WPD) provides guidance on preparing for, completing and documenting Data Cleaning.

2. INSTRUCTIONS and GUIDANCE

2.1 Creating a Data Cleaning Plan

- 2.1.1** The Data Cleaning Plan will be written using the DM012 Data Cleaning Plan Template. This template is used for both Interim and Final Data Cleaning Plan.
- 2.1.2** The Data Management team will ensure that a Data Cleaning Plan is in place before Data Cleaning begins. The proposed Data Cleaning periods that have been agreed will be stated in the DMP.
- 2.1.3** It is recommended that the Final Data Cleaning Plan is agreed and in place at least six months before the End of Study Date (as defined in the protocol) although this may vary according to the needs of the study. The Data Cleaning Plans should be agreed as early as practical.
- 2.1.4** The scope and content of the Data Cleaning Plan will be agreed with the Trial Manager and the Trial Statistician as specified on the template. This includes:
- **Proposed Data Cleaning Deadline and Start Date**
A deadline date for the cleaned data to be submitted for final analysis should be agreed with the Trial Manager and Trial Statistician. This is not necessarily the date of database lock. Once a deadline date has been agreed, the Data Management team will agree a date for the data cleaning to begin. Consideration should be given to the type and amount of data to be cleaned, follow-up with sites and the potential for other issues to be identified during the clean which may extend the process. It is recommended that the duration of time required for data cleaning is overestimated to allow for unexpected issues to be addressed.
 - **Standard Checks and Reviews**
The datasets listed in this section are standard datasets that will be applicable to most studies. In some cases, the query management system may check for discrepancies throughout the study and further checks may not be required during data cleaning. This list can be edited according to the study requirements.
 - **Outcome-specific Checks and Reviews**
Each Primary and Secondary Outcome should be listed along with any checks and reviews that will be completed in order to clean this data. Where it is not possible to clean the data for this outcome (e.g. scan data), the reason for this should be specified.
 - **Other Checks and Reviews**
If any other data or checks are to be completed during data cleaning, these should be specified here.

2.1.5 Once agreed and approved, the Data Cleaning Plan is not subject to a formal review schedule. The Data Management team is responsible for ensuring that the Data Cleaning Plan is maintained throughout the data cleaning process and any revisions are documented as appropriate.

2.2 Recording and Resolving Issues Identified during Data Cleaning

2.2.1 All issues and errors will be recorded using DM014 Data Cleaning Log Template.

2.2.3 Any action taken to resolve an issue will also be recorded on the Data Cleaning Log, including any issues discussed with Trial Managers or Trial Statistician. Any emails documenting resolution of an issue must be retained.

2.2.4 As much as possible, manual queries should be raised and followed-up to resolution on the study database to query data errors or discrepancies identified during the data cleaning activity. Where manual queries are opened and closed for this purpose, this should also be recorded on the Data Cleaning Log.

2.3 Completion of Data Cleaning

2.3.1 Data Cleaning will be considered to be complete after the following:

- All issues documented on the Data Cleaning Log have been addressed and resolved
- All queries (including manual queries) on the database have been addressed and/or closed
- All missing data has been entered or documented as known to be missing at the end of study

2.3.2 Once the Data Management team is satisfied that Data Cleaning is complete, this will be confirmed to the Trial Manager via email.

2.3.3 The Data Management will retain the Data Cleaning Log and all associated documentation (including any reports obtained, emails etc.) in the study specific data management file.

3. RELEVANT DOCUMENTS AND REFERENCES

SOP and WPD

- ECTU Central Office SOP ECTU_DM_05 Data Quality
- ECTU Central Office SOP ECTU_SOP_DM_07 Data Management Version Control and Document Review

Available on the ECTU website - www.ed.ac.uk/usher/edinburgh-clinical-trials/supporting-trials/governance/standard-operating-procedures

Templates

DM012 Data Cleaning Plan Template

DM014 Data Cleaning Log Template

Available from the Data Management Team – dm.ectued.ac.uk