



## STANDARD OPERATING PROCEDURE FOR DATA MANAGEMENT IN CLINICAL RESEARCH STUDIES USING EXCEL

SOP NUMBER:	TASC SOP048 v8
AUTHOR:	Petra Rauchhaus
EFFECTIVE DATE:	12 July 2023
REVIEW DATE:	12 July 2025

### 1. PURPOSE

This SOP describes the process for data management in clinical research studies when the data management system in use is Excel.

### 2. SCOPE

This SOP applies to members of staff associated with and managing clinical research studies that are sponsored or co-sponsored by the University of Dundee and/or NHS Tayside (NHST).

It is intended for use by all staff who have the responsibility for designing, managing or administering the data management process for clinical research studies where an Excel-based data management system is in use.

### 3. RESPONSIBILITIES

To safeguard data quality and security in CTIMPs where Excel is used as the data management system, the trial team must adhere to the processes described in this SOP and the TASC Guideline 'When is Excel an acceptable data management system for a CTIMP?' (Doc. Ref. 097) This SOP should be used in conjunction with the other TASC Data Management SOPs as applicable.

### 4. PROCEDURE

#### 4.1 Spreadsheet Setup

- 4.1.1 The spreadsheet template will be used to provide rows and columns matched to the data required from the Case Report Form (CRF) or other source data. The Chief Investigator (CI), statistician or other delegate responsible for data analysis will check the spreadsheet template for accuracy and format before any data entry begins and it will be amended until all outcomes specified in the protocol can be entered in a suitable format for data analysis.

## 4.2 Spreadsheet Validation

4.2.1 The trial team may use Excel to manage the trial data and undertake some simple analysis. It is expected that there will be a documented check on formatting of the cells (for example, numeric, dates) and any formulae that have been put into the system. No scripts or macros should be added to the spreadsheet.

## 4.3 Data Entry

4.3.1 The trial team will agree a data entry procedure and document this in the Trial Master File (TMF).

4.3.2 When data are entered into Excel, any inconsistencies with source data documentation should be highlighted using the Insert Comment facility available on Excel - this should include the name of the person raising the query. Transcription errors which are subsequently corrected on Excel should be highlighted using the Insert Comment facility. Any changes to the paper CRF must be made with a single line drawn through, initialed and dated. The original data must be clearly visible.

4.3.3 Once data entry commences, to support the audit trail, the file must be saved regularly as a new spreadsheet. This could be daily or weekly depending on frequency of updating data; a suggestion for how this might be done is to save spreadsheets chronologically and record the identity of the member of the team adding to data, save with the date reversed, and initials. This method will automatically file the workbooks with the most recent version at the bottom of the list.

Example: **100624.HH**

**100629.HH**

See 4.5.1 - 4.5.10 for more details.

## 4.4 Data Validation

4.4.1 Data validation can be undertaken by a number of methods:

1. Double entry one person
2. Double entry two people
3. Single entry with second look
4. Single entry with a second individual reading aloud

4.4.2 The justification for the method selected should be documented. Methods 1 and 2 are self-explanatory.

For methods 3 and 4:

Method 3: after single data entry, the data recorded on the paper CRF (and other source documents) are visually compared what was entered into Excel.

Method 4: This involves an individual reading aloud the data on the paper CRF or in the spreadsheet, while another individual checks the corresponding data in the spreadsheet or on the paper CRF as appropriate.

- 4.4.3 Data recorded in the CRF that is not source data in itself should be consistent with the source documents, or the discrepancies should be explained.
- 4.4.4 Visual checks should be made on all missing values and values outside normal or expected ranges. Data type should also be checked for accuracy.
- 4.4.5 Logical checks should be performed to ensure consistent reporting between relevant fields and that there are no differences between fields: e.g. male/pregnant. All dates should be checked to ensure they are logical.
- 4.4.6 All queries are to be recorded by the Insert Comment facility on Excel.
- 4.4.7 Data checking should continue until all missing data and/or inconsistent values have been corrected or clarified.
- 4.4.8 The person responsible for analysis will not be given access to the data until the spreadsheet has been fully audited and all queries have been resolved.
- 4.4.9 It should always be possible to verify the electronic data with the source data.

See 4.5.13 - 4.5.15 for more details.

## 4.5 Quality Control

- 4.5.1 PCs should be password protected and kept in a room which can be locked to prevent unauthorised access.
- 4.5.2 A trial-specific Excel spreadsheet should be opened and used to record only outcomes that are specified in the protocol. The spreadsheet should be password protected and available to delegated trial staff only. The folder containing the spreadsheet should be accessible only to named trial individuals.
- 4.5.3 The spreadsheet should be saved on a secure server which is backed up daily and where a disaster recovery plan is in place. The folder containing the spreadsheet should be accessible only to named trial individuals.
- 4.5.4 Data should be transcribed to Excel from paper CRF, prior to any data transfer to any other programme for analysis (as stated in protocol). As 4.3.2, all queries on CRF must be clarified with the CI or Principal Investigator (PI) and any changes to the data in the CRF must be made with a single line drawn through, initialled and dated. The original data must be clearly visible.
- 4.5.5 Data should be entered on the same day as the participant visit or as soon as possible afterwards.
- 4.5.6 The data entry person should initial and date each page of the CRF and/or procedure after it is entered into the spreadsheet.
- 4.5.7 The spreadsheet should be saved at the end of each data entry session and the date of entry recorded to ensure an audit trail on both the CRF and the spreadsheet.
- 4.5.8 Participants should be identified only by unique study number, initials and/or age at consent. No patient identifiable information should be stored on the spread sheet.
- 4.5.9 Participants should be cross checked with the Screening Log and/or Randomisation Log, as appropriate.
- 4.5.10 The data on the Excel spreadsheet must be always verifiable against the CRF.

Uncontrolled when printed. Please visit the [TASC website](#) for the latest version of this SOP.

- 4.5.11 When data is fully entered the password should be released to those who are delegated the task of data checking. The spreadsheet should be given a new password and again saved according to date.
- 4.5.12 The data on the Excel spreadsheet should be re-checked against the CRFs to confirm accurate entry. Queries should be identified by using the Insert Comment Function in Excel and resolved and or explained as soon as possible.
- 4.5.13 A final check on the accuracy of the data should be performed using one of the methods outlined in section 4.4.1, where possible, by individuals not involved in the original data entry. The individual(s) performing these checks should initial and date each page of the CRF and/or procedure to confirm that a final check has been carried out.
- 4.5.14 Checks should be made on all missing values and values out with normal or expected ranges. A check should also be made that values entered are of the correct type: i.e. numerical instead of text. Logical checks should be performed to ensure consistent reporting between relevant fields: i.e. male/pregnant.
- 4.5.15 Data checking must continue until all missing data and/or inconsistent values have been corrected or clarified.
- 4.5.16 When data checking is complete, a database pre-lock meeting will be arranged and will include those responsible for trial management, data management and statistical analysis. Refer to the TASC SOP on Locking Clinical Study Databases. As a minimum, the CI and person delegated to perform statistical analysis should provide authorisation to lock the data. The meeting will be documented.
- 4.5.17 Upon authorisation, the database should be locked using the Protect Worksheet function of Excel, and the file named FINAL RESULTS. No further additions or alterations to data can occur after this point.

The locked Excel database should then be transferred to a USB memory stick or CD, sealed in an envelope labelled with the EudraCT number and the date of the data lock and filed in the TMF. The spreadsheets can also be printed and filed alongside the USB stick in the TMF. In parallel, the Excel database should remain archived on a secure server as specified in the TASC Archiving SOP.

## 5. ABBREVIATIONS & DEFINITIONS

CI	Chief Investigator
CRF	Case Report Form
CTIMP	Clinical Trial of an Investigational Medicinal Product
NHST	NHS Tayside
PI	Principal Investigator
SOP	Standard Operating Procedure
TASC	Tayside Medical Science Centre
TMF	Trial Master File

## 6. ASSOCIATED DOCUMENTS & REFERENCES

Doc Ref 097: When is Excel an acceptable data management system for a CTIMP?

## 7. DOCUMENT HISTORY

*History prior to 2021 is in the archived SOPs available from TASC Quality Assurance Dept.*

Version Number:	Reviewed By (Job Title):	Effective Date:	Details of editions made:
6	Tracy Petrie (Quality Assurance Support Officer)	18/01/2021	Uploaded to new TASC SOP template which shows the new TASC website in the footer. Physical scan converted to electronic pdf as a requirement for upload to new TASC website.
7	Petra Rauchhaus (Clinical Trials Statistician)	12/07/2021	Scheduled review. Change of title from CTIMPs to clinical research studies.
8	Petra Rauchhaus (Clinical Trials Statistician)	12/07/2023	Scheduled review. No changes required.

## 8. APPROVALS

Approved by:	Date:
Dr Valerie Godfrey, TASC Quality Assurance Manager, on behalf of TASC Clinical Research Guidelines Committee	11 July 2023