STANDARD OPERATING PROCEDURE FOR DATA MANAGEMENT IN CLINICAL RESEARCH

SOP NUMBER:	TASC SOP053 v6
AUTHOR:	Marcus Achison
EFFECTIVE DATE:	25 Aug 2023
REVIEW DATE:	25 Aug 2025

1. PURPOSE

This Standard Operating Procedure (SOP) describes the management of clinical research data.

2. SCOPE

This SOP applies to clinical research studies sponsored or co-sponsored by the University of Dundee (UoD) and/or NHS Tayside (NHST). It is intended for use by those involved with the management of data in clinical research studies. It should be used in conjunction with other relevant SOPs, as agreed and approved by Sponsor.

3. RESPONSIBILITIES

Research data should be collected, recorded, and managed in accordance with the General Data Protection Regulation (GDPR) European Union (EU) 2016/679 (2018) and UoD and NHST policies. In addition, data for all studies involving the participation of human subjects must be handled according to the principles of Good Clinical Practice (GCP). GCP standards are a legal requirement for Clinical Trials of Investigational Medicinal Products (CTIMPs).

4. PROCEDURE

4.1 Data Management and Clinical Studies

Data management encompasses the receipt and tracking of clinical study data, data entry into the study database (or equivalent data capture system), generating and resolving data queries and database/system lock.

The development and management of a study-specific Data Management System (DMS) may be performed by the Tayside Clinical Trials Unit (TCTU), outsourced to an external third party or carried out by the study team. If the development and management of a DMS is outsourced to an external third party, the Chief Investigator (CI) should contact TASC Legal to prepare an appropriate collaboration or service agreement.
Each study requires documentation to prove that the DMS meets the requirements of the end users and that data can be stored and exported without alteration.

Study-specific documentation should include:

- A Data Management Plan (DMP)

- A DMS specification document based on the protocol and paper Case Report Form (pCRF), if one is to be used, any subsequent amendments and agreed data validations
- Evidence that the system has undergone and passed validation and that any study-specific configuration has also undergone and passed testing. This includes validation of the export process and user acceptance testing
- User instructions and evidence of user training
- Documented release of the system for "live" use.

4.2 DMP

The DMP describes the study-specific data management activities and shall be put in place for each study. It must be reviewed and signed off by the CI and study Statistician, and where applicable, persons responsible for study data management and study management.

A DMP typically details:

- All data sources and how they will be integrated (i.e. the data flow)
- The DMS to be used, e.g. Castor EDC, OpenClinica, Excel
- Responsibility and procedures for activities including:
    - Database development and testing
    - Approval of the template pCRF or study variables if a pCRF is not used, and any subsequent changes
    - Data entry
    - Data checking, querying and correcting
    - Any other quality control measures (e.g. database audit, including predetermined data entry error limit and source data verification)
    - Preparing data extracts
    - Database lock and archiving
    - Management and security of the DMS
- Any additional study-specific instructions
- Information on data requirements for Data Monitoring Committees and interim analysis (if relevant for the study)
- Data Transfer, if required
- Details of any external datasets that will form part of the analysis data.

4.3 DMS

The DMS should include the following:

- Reflect the layout and design of the pCRF, if used, as accurately as possible to facilitate ease of data entry
- Should not collect participant-identifiable data
- Date of birth may be collected in exceptional circumstances where it is vital to the study outcome. If this is the case, the use of date of birth should be included in the Sponsor's risk assessment and risks mitigated by employing suitable data security measures. Collecting age at consent instead of date of birth should be sufficient in the majority of studies
- Include data validation functionality, e.g. range checks and checks for missing or inconsistent data, to ensure the highest quality data

- Use unambiguous and unique identifying participant IDs. The code or file linking participants' names with their ID shall be kept secure and separate from the data used for study analysis
- For blinded studies, the study blinding should be safeguarded. Treatment allocation shall not be broken by day-to-day usage of the system and must support any unblinding procedures put in place to ensure participant safety. Data which may inadvertently unblind staff, e.g. laboratory data, must remain inaccessible to the study team
- Include an audit trail to ensure it is possible to determine when and by whom data has been originally entered or changed
- May also include the option to record electronic Participant Recorded Outcome (ePRO) data, where participants complete study data themselves via managed access to the database (email invites, mobile device, tablet etc.).

If data are transformed during processing, e.g. calculation of BMI from height and weight, or if a partial date field (text field) is transformed into a real date field it should always be possible to compare the original data with the processed data.

4.4 Validation and Functional Testing

The DMS must undergo a process of validation, functional testing and user acceptance testing to demonstrate that it is fit for purpose and performs consistently.

4.5 Training

All DMS users must be trained by an appropriate trainer and provided with user guides. Training must be documented both in the individual's training log and also in the study-specific Data Management folder/Trial Master File (TMF).

4.6 User Access

A record of DMS users, their access levels and when access was granted and revoked shall be kept. This may be held within the DMS, if this function is available, or in the study TMF. Access will be revoked when access is no longer required.

4.7 Data Quality

4.7.1 Data Entry and Data Queries

Data must be entered into the DMS exactly as recorded in the pCRF (where used) and shall be verifiable against source data.

Data queries may be raised as a result of visual inspection of the data or by point of data entry checks present in the DMS used to check data. This can be performed as the data is entered or by data validation processes such as batch validation, which is used to perform checks that cannot be carried out at the point of data entry.

Where pCRFs are used, the data recorded in the pCRF and DMS must be kept consistent with one another. If a correction is made to the DMS as a result of the data querying process, the pCRF must also be updated by crossing through the original value without obscuring it, writing the new value and initialling and dating the correction. If a change is

made to the pCRF, a reason must also be provided on the pCRF to explain why the change was necessary.

Where data entry is not carried out at the site, the original pCRF should remain at site with a copy being sent elsewhere for data entry. Scanned copies of pCRFs may be transferred electronically via secure methods to facilitate data entry. Any changes to the original pCRF must be notified to the data entry team by means of an updated copy of the original being sent to the data entry team. Any copies of the pCRF should be checked and any participant identifiable data removed prior to transfer.

4.7.2 Identification of Source Data

Data points for source data verification (SDV) should be selected using a risk-based approach. This normally includes safety data and data points linked to the main outcomes of the study.

Data recorded in ePRO systems is considered source data. Changes to this data should be rare and should be supported by documentation in the participant's medical record.

The DMS, including ePRO systems, should have an audit trail, ensuring that when data are changed, the original value is retained as well as who changed what data, when and why. The DMS should prompt users to enter a reason for changing the data. Typical reasons for changing data include data entry error or response to query.

Data from randomisation systems or external datasets may not be imported or entered into the DMS but will be merged with the data from the DMS for later analysis. This should be reconciled against the related DMS data to ensure that all expected data are present. Where the same data are held in both systems they should be reconciled, and any inconsistencies identified and resolved to ensure data accuracy.

Where data is imported into the DMS, consideration must be taken to ensure that data accuracy and integrity is maintained, e.g., data is verifiable against source and that suitable data management and monitoring processes are in place to deal with data inconsistencies/queries.

4.8 PI Oversight of Study Data

For studies which use a pCRF or eCRF, all data entries and updates made to the study data must be performed by a delegated individual using the procedures described above in section 4.7.1. When all data have been entered and confirmed as correct, a final sign-off of the study data must be provided in the pCRF or the eCRF by the site investigator.

4.9 Data Lock, Unlock and Archive

The DMS will be locked when data entry is complete, and the data have been cleaned to the satisfaction of the study Statistician and the CI. In studies using pCRFs this includes quality checks of the data to ensure a predetermined data entry error limit has been met.

Database unlock shall be strictly controlled and only carried out in exceptional circumstances, with the approval of the CI, Sponsor, and study Statistician.

Refer to TASC SOPs on Database Lock and Archiving.

4.10 Ensuring Site Investigators maintain control of site data (eCRF studies)
For eCRF studies, it has to be ensured that sites have continuous access to their data. Therefore, study data should be transferred to the site as soon as possible after database lock. PIs should have read-only access to their study data in the eCRF until the study data and associated documentation is transferred to the site. This is required to ensure that investigators retain control of their data and to facilitate archiving of their study data according to the local site archiving procedure.

4.11 Data Transfer to third parties
Where data is to be transferred to a third party (i.e. a party outwith the UoD and NHS Tayside), the study team must ensure with TASC Legal, where appropriate, that Data Transfer Agreements are in place with the recipient of the data to ensure that it is used and stored appropriately.

Datasets shall be encrypted, using an industry-standard encryption mechanism, such as AES-128, prior to transfer (whether electronic or on removable media). Alternatively, a web-based secure data repository may be used (e.g. LabKey). In this case, the software should be installed on a secure UoD server with access restricted to authorised personnel only.

All methods of data transfer should be carried out in accordance with the principles of data confidentiality set out in the GDPR (EU) 2016/679 (2018).

5. ABBREVIATIONS & DEFINITIONS

CI	Chief Investigator
CRF	Case Report Form
CTIMP	Clinical Trial of Investigational Medicinal Product
DMP	Data Management Plan
DMS	Data Management System
EU	European Union
ePRO	Electronic Participant Recorded Outcome
GCP	Good Clinical Practice
GDPR	General Data Protection Regulation
NHST	NHS Tayside
pCRF	Paper Case Report Form
SDV	Source Data Verification
SOP	Standard Operating Procedure
TASC	Tayside Medical Science Centre
TCTU	Tayside Clinical Trials Unit
TMF	Trial Master File
UoD	University of Dundee

6. ASSOCIATED DOCUMENTS & REFERENCES

None.

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:
5	Emma McKenzie (Clinical Trial Information Systems Manager)	22/09/2021	Additional information on use of eCRFs, ePROs and data reconciliation/import between systems. Section 4.3 updated to comply with GDPR with regard to Date of Birth.
6	Marcus Achison (Database Manager)	25/08/2023	Visual Verification no longer mandatory. Updates to text.

8. APPROVALS

Approved by:	Date:
Dr Valerie Godfrey, TASC Quality Assurance Manager, on behalf of TASC Clinical Research Guidelines Committee	24 Aug 2023