**DATA MANAGEMENT PLAN**

**Protocol: *Specify full title for the study***

**Sponsor:** University of Technology Sydney

15 Broadway, Ultimo NSW 2007 Australia

*This document refers to the ITCC Standard Operating Procedures (*[*https://www.uts.edu.au/research-and-teaching/our-research/IMPACCT/paccsc/researcher-resources/standard-operating-procedures*](https://www.uts.edu.au/research-and-teaching/our-research/IMPACCT/paccsc/researcher-resources/standard-operating-procedures)*)*

# 

# **CONFIDENTIALITY STATEMENT**

The information contained in this document is the property of the IMPACCT Trials Coordination Centre (ITCC) on behalf of the University of Technology Sydney and may not be reproduced, published or disclosed to others without written authorisation from the ITCC on behalf of University of Technology Sydney*.*

# **DOCUMENT INFORMATION**

|  |  |  |
| --- | --- | --- |
| **Version Number** | **Version**  **Date** | **Summary of Revisions Made** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**APPROVAL SIGNATURES**

**Co-ordinating Principal Investigator**

|  |  |  |
| --- | --- | --- |
| **Email approval via email** |  |  |
| **Signature\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |  | **Date\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |

**PROJECT MANAGER**

|  |  |  |
| --- | --- | --- |
| **Signature\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |  | **Date\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |

# **TABLE OF CONTENTS**

[CONFIDENTIALITY STATEMENT 2](#_Toc99956743)

[DOCUMENT INFORMATION 2](#_Toc99956744)

[APPROVAL SIGNATURES 3](#_Toc99956745)

[TABLE OF CONTENTS 4](#_Toc99956746)

[ABBREVIATIONS 5](#_Toc99956747)

[DEFINITIONS 8](#_Toc99956748)

[DYNAMIC REFERENCES 10](#_Toc99956749)

[1. INTRODUCTION 11](#_Toc99956750)

[2. ORGANISATIONAL DATA GOVERNANCE OVERSIGHT 12](#_Toc99956751)

[3. STUDY STRUCTURE 13](#_Toc99956752)

[4. GENERAL STUDY INFORMATION 14](#_Toc99956753)

[5. DATA COLLECTION 14](#_Toc99956754)

[5.1 Consent for Data Collection and Use 14](#_Toc99956755)

[5.2 Data Collection 14](#_Toc99956756)

[5.3 Data Collection Flow 14](#_Toc99956757)

[5.4 Discrepancy Management 16](#_Toc99956758)

[6. PRIVACY AND CONFIDENTIALITY 17](#_Toc99956759)

[6.1 Breach of Privacy/Confidentiality 17](#_Toc99956760)

[7. FORMS OF DATA 17](#_Toc99956761)

[7.1 Identifiable Data 17](#_Toc99956762)

[7.2 De-identifiable Data 18](#_Toc99956763)

[8. ACCESS TO AND USE OF DATA 18](#_Toc99956764)

[8.1 Identifiable Data 18](#_Toc99956765)

[8.2 De-identified Data 19](#_Toc99956766)

[9. STORAGE AND DESTRUCTION OF DATA 19](#_Toc99956767)

[9.1 Identifiable Data and Source Documents 19](#_Toc99956768)

[9.2 De-identified Data 20](#_Toc99956769)

[10. CONSULTATION 20](#_Toc99956770)

[10.1 Aboriginal or Torres Strait Islander 20](#_Toc99956771)

[11. Return of Results 21](#_Toc99956772)

[11.1 Incidental Findings 21](#_Toc99956773)

[11.2 Results Arising from Future Research 21](#_Toc99956774)

[12. WITHDRAWAL OF DATA 21](#_Toc99956775)

[13. DATABASE ARCHITECTURE 21](#_Toc99956776)

[13.1 Database 21](#_Toc99956777)

[13.2 REDCap EDC System Security 22](#_Toc99956778)

[13.2.1 REDCap EDC server’s physical plant 23](#_Toc99956779)

[13.2.2 The REDCap EDC system’s disaster-mitigation measures 23](#_Toc99956780)

[13.2.3 Website security technologies 23](#_Toc99956781)

[13.3 IWRS 23](#_Toc99956782)

[14. DATA SECURITY 23](#_Toc99956783)

[15. DATA VALIDATION PROCESS 25](#_Toc99956784)

[16. VERIFICATION OF REDCap EDC SETUP AND IMPLEMENTATION 26](#_Toc99956785)

[17. SYSTEM SPECIFICATIONS 27](#_Toc99956786)

[17.1 REDCap EDC User Acceptance Testing (UAT) 28](#_Toc99956787)

[18. DATA ENTRY 28](#_Toc99956788)

[18.1 Pre-requisites for Data Entry 31](#_Toc99956789)

[18.2 Granting Access to the Production Mode of the REDCap EDC 33](#_Toc99956790)

[18.3 Entering Data 33](#_Toc99956791)

[19. DATA CLEANING 34](#_Toc99956792)

[19.1 Automatic Edit Checks 35](#_Toc99956793)

[19.2 Manual Edit Checks 36](#_Toc99956794)

[19.3 Detecting Outlier 36](#_Toc99956795)

[19.4 Missing Data Identification 36](#_Toc99956796)

[20. MONITORING 37](#_Toc99956797)

[21. CHANGES TO A REDCap EDC PRODUCTION MODE 38](#_Toc99956798)

[22. DATA DOWNLOAD 39](#_Toc99956799)

[23. DATA BASE CLOSURE 39](#_Toc99956800)

[23.1 Closure Checks 40](#_Toc99956801)

[23.2 Database Lock and Preparation for Final Analysis 41](#_Toc99956802)

[23.3 Updating the Database after Database Lock 41](#_Toc99956803)

[24. DATA BASE TRANSFER TO CPI/ study statistician 42](#_Toc99956804)

[25. REFERENCES 44](#_Toc99956805)

[26. APPENDICES 45](#_Toc99956806)

**ABBREVIATIONS**

AAF Australian Access Federation

AE Adverse Event

API Application Programming Interface

AR Adverse Reaction

CDISC Clinical Data Interchange Standards Consortium

CONSORT Consolidated Standards of Reporting Trials

CPI Coordinating Principal Investigator

CSRI Client Services Receipt Inventory

DB Database

DMP Data Management Plan

EDC Electronic Data Capture

eCRF electronic Case Report Form

GCP Good Clinical Practice

HREC Human Research Ethics Committee

ICH International Conference on Harmonisation of Technical Requirements for Registration   
 of Pharmaceuticals for Human Use

IMPACCT Improving Palliative, Aged and Chronic Care through Clinical Research and Translation

ITCC IMPACCT Trials Coordination Centre

IWRS Interactive Web Response System

ODM Operational Data model

NHMRC National Health and Medical Research Council

PI Principal Investigator

PICF Participant Information and Consent Form

REDCap Research Electronic Data Capture

RGO Research Governance Office

SAE Serious Adverse Event

SAR Serious Adverse Reaction

SDV Source Data Verification

SEC Self-Evident Correction

SOP Standard Operating Procedure

SUSAR Suspected Unexpected Serious Adverse Reaction

TGA Therapeutic Goods Administration

TMF Trial Master File

**DEFINITIONS**

**Audit trail** Documentation that allows reconstruction of the course of events.

**Cross-check** An edit check that compares variables from different CRF/worksheet pages or participants’ variables from different CRF/worksheet pages.

**Database lock** The procedure which prevents data from being changed after validation to prevent corrupting or invalidating when multiple users try to write in the database.

**Data collection** Data collection worksheets are designed to capture required data and to   
 enable smooth data entry in electronic case report forms (eCRFs) through   
 similar question and response structures to minimise data entry errors.

**Worksheet** These worksheets may be considered source documentation, where the worksheet is the first recording of study data.

**Project Data Manager** Individual who manages all the data-related aspects of the project at the ITCC.

**Principal Investigator** Individual who is responsible for the conduct of the trial at each participating site including data entry and validation.

**Project Manager** Individual who has oversight over all activities regarding the collection, storage, access, securityand primary and secondary use of all data created or held by the ITCC.

**Central Monitor** Individual who has completed training to conduct monitoring activities at the ITCC and is responsible for the resolution of outstanding data queries.

**Source data** All information in original records and certified copies of original records of clinical findings, observations or other activities in a clinical trial necessary for the reconstruction and evaluation of the trial.

**Source documents** Original documents, data and records (e.g., hospital records, clinical and office charts, trial worksheets, laboratory notes, memoranda, participants’ diaries or evaluation checklists, pharmacy dispensing records, recorded data from automated instruments, copies or transcriptions certified after verification as being accurate copies, microfiches, photographic negatives, microfilm or magnetic media, x-rays, participants’ files and records kept at the pharmacy, laboratories and medico-technical departments involved in the clinical trial).

# **DYNAMIC REFERENCES**

| **ITEM** | **CONTENT** | **DOCUMENT OWNER** | **CURRENT VERSION STORAGE LOCATION** |
| --- | --- | --- | --- |
| Annotated case report form | The full set of CRFs, including SAS variable names, sorted in protocol schedule order. | Project data manager |  |
| Data dictionary | List of dataset names. List of variable names, with corresponding valid values, data types, labels. | Project data manager |  |
| Monitoring plan | Description of clinical monitoring activities. | Central monitor |  |
| Redcap EDC instrument-event mappings | Describes and tracks the process for review and approval of the study-specific REDCap EDC data collection instruments that are utilised for each event in a longitudinal data collection project. | Project data manager |  |
| Statistical analysis plan | Containing a detailed elaboration of the principal features of the analysis described in the clinical trial protocol and includes procedures for statistical analysis of the primary and secondary variables. | Statistician |  |
| Study contact list | Names, roles, and contact information for key staff and ITCC study team members. | ITCC team |  |
| Study Start-up Checklist | List of items to be completed before study start-up. | Project data manager |  |
| Study timeline | Projected timeline of study events and deliverables. | Trial coordinators |  |

# **1. INTRODUCTION**

This Data Management Guide outlines how data will be handled during the study [study code] and after its completion.

The quality of data generated during a clinical trial plays an important role in the outcome of the study itself. With the aim to generate high-quality, reliable, and statistically sound data, clinical data management plays a critical phase in clinical research (Krishnankutty et al. 2012). Indeed, various procedures in clinical data management including electronic case report form (eCRF) design, CRF annotation, database design, data-entry, data validation, discrepancy management, medical coding, data extraction, and database locking are assessed for quality at regular intervals during a trial (Krishnankutty et al. 2012). High-quality data should be complete, accurate, verifiable, and suitable for statistical analysis. These should meet the protocol-specified parameters and comply with the protocol requirements. This also means that data need to be trustworthy, i.e., must originate from observations of study participants as reported in source documents and are not altered or falsified.

The purpose of this data management plan (DMP) is to guarantee the authenticity, integrity, and confidentiality of data collected during the [abbreviated study title] and to provide guidance in all the study phases, from preliminary e-CRF design and data collection to data cleaning and analysis. In particular in this study, data collection will be supported by a REDCap electronic data capture (EDC) system. The present DMP covers the design and production of the data capture tool for the collection of the study participants’ data at the Investigator site, the design and construction of the database to maintain the data electronically, the processing of the data (entry, cleaning, and query management) and the production of the final dataset ready for analysis and archiving.

Therefore, the ITCC will use REDCap to 1) design questions and items to capture the study required data; 2) build the worksheets/CRFs to collect the data; 3) add in check points and instructions to assist with the collection of the data; and 4) monitor the changes made to the data (audit trail) as required by Good Clinical Practice (GCP).

# **2. ORGANISATIONAL DATA GOVERNANCE OVERSIGHT**

All processes detailed in this data management plan will be carried out in accordance with applicable ICH Good Clinical Practice (GCP) requirements and applicable Australian Directives.

The following constitute the most important sources for GCP-compliant data management in Australia:

* ICH E6 (R2) Good Clinical Practice
* NHMRC National Statement (2007) -Updated in 2018

Additional sources refer to Australian regulations or to other recommendations, such as:

* Therapeutic Goods Administration: Australian Clinical Trials Handbook (2018)
* Standard operating procedures for Human Research Ethics Committee (HREC) and Research Governance Office (RGO)
* Privacy Act 1988 and Australian Privacy Principles
* Good Clinical Data Management Practices, version 4
* CDISC Clinical Data Interchange Standards Consortium, Operational Data model (ODM)
* CONSORT documentation for reporting of clinical trials.

Any modification (minor or substantial) needed during the project, is governed by the ITCC standard operating procedures (SOPs).

# **3. STUDY STRUCTURE**

**TABLE 3.1. STUDY STRUCTURE**

|  |  |
| --- | --- |
| **Sponsor** | University of Technology Sydney  15 Broadway  Ultimo NSW 2007 Australia |
| **Co-ordinating Principal Investigator** | *Name*  *Position*  *Organisation*  *Address*  *Phone*  *Fax*  *Email* |
| **Lead Site (Australia)** | *Name*  *Address* |
| **Lead Investigator (Australia)** | *Name*  *Position*  *Organisation*  *Address*  *Phone*  *Fax*  *Email* |

# **4. GENERAL STUDY INFORMATION**

Study drug

Study population

Study design

Estimated enrolment

Primary objective

Secondary objectives

Number of e-CRF Instruments

# **5. DATA COLLECTION**

## 5.1 Consent for Data Collection and Use

All participants will be informed of, and provide consent for, the collection and use of their data for the purposes of this study.

## 5.2 Data Collection

Data will be collected from the following sources:

* Direct communication with the participant
* Study assessments, including laboratory test results, imaging, biomedical monitoring, questionnaires, interviews, and data downloaded from apps
* Participant medical records
* Communications with participant’s clinical care team

Data will be collected primarily by the Investigator or designated study staff. All study personnel involved in data collection will be trained in GCP, the study protocol, and collection requirements.

Collection of data will be limited to that necessary for the specified purposes of the study, or for additional purposes that the participant has explicitly consented to.

## 5.3 Data Collection Flow

The Investigator or delegate will be responsible for the accuracy of the participant data entered in the REDCap EDC system. Data from source documents, relevant to the protocol, should be reported under the responsibility of the Investigator, as soon as possible, on the provided electronic data capture system. All entries must be completed in English.

The REDCap EDC system, provided with audit trail enabled, will allow identified and authorised users to remotely store data in electronic forms so that all data transactions between sites and central database are automatically and chronologically recorded.

The Principal Investigator should ensure the accuracy, completeness, legibility, and timeliness of the data reported to the Sponsor in the eCRFs from the REDCap EDC system and in all required reports.

Study data reported in the REDCap EDC system will be checked for consistency at UTS and any discrepancies found will be notified to the Investigator who is responsible for resolving the data discrepancies, reporting data changes, and maintaining accurate and consistent data in source documents and REDCap EDC system.

As specified in the study protocol, eCRFs may be considered source documentation, where the eCRF is the first recording of study data, such as during study visits where there are no other traditional source documents. To assist sites in ensuring that source documentation and data reported in the eCRFs are complete and consistent, ITCC develops and provides study specific data collection worksheets for use in the [abbreviated study title]. The worksheets are designed to capture the required data and to enable smooth data entry in the eCRFs through similar question and response structures to minimise data entry errors. These worksheets may be considered source documentation, where the worksheet is the first recording of study data. All data entered in the eCRF must be verifiable with source data from the participant’s data collection worksheets, medical records, and other relevant documents as applicable. Any discrepancies between the eCRF and the source data (including study worksheets) should be explained and documented.

The REDCap EDC system for the [abbreviated study title] is provided by UTS, supported by Australian Access Federation (AAF) Virtual Home, including eCRFs and data management services. REDCap is a secure web platform for building and managing online databases and surveys. Computer workstations do not require any specific client-side software installation. REDCap's streamlined process for rapidly creating and designing projects offers a vast array of tools that can be tailored to virtually any data collection strategy. REDCap provides automated export procedures for seamless data downloads to Excel and common statistical packages (SPSS, SAS, Stata, R), as well as a built-in project calendar, a scheduling module, ad hoc reporting tools, and advanced features, such as branching logic, file uploading, and calculated fields.

## 5.4 Discrepancy Management

Discrepancy management is a process of cleaning participants’ data in the clinical data management system; it includes reviewing discrepancies, investigating the reason, and resolving them with documentary proof or declaring them as irresolvable (Figure 1). Discrepancy management helps in cleaning the data and gathers enough evidence for the deviations observed in data.

The [abbreviated study title] REDCap EDC system has a discrepancy database where all discrepancies will be recorded and stored with audit trail. Based on the types identified, discrepancies are either flagged to the Investigator for clarification or closed in-house as Self-Evident Corrections (SEC) without sending the discrepancy to the site. For those discrepancies that require clarification from the Investigator, a data query will be raised in REDCap and assigned to the study coordinator/investigator. The study coordinator/investigators will check the query, and either amend the data or explain the circumstances that led to the discrepancy in the data. The Data Management team reviews all discrepancies at regular intervals to ensure that they have been resolved. All resolved data discrepancies will be checked and recorded as ‘closed’. All discrepancies will be assessed against critical data elements and assigned for resolution as below (Table 5.4.1).

**Table 5.4.1. Flow of discrepancy management**

|  |  |  |
| --- | --- | --- |
| **Data Element** | **Source of error** | **Resolution site** |
| Consent | Data entry  CRF/Worksheet | ITCC  Study site |
| Eligibility | Data entry  Worksheet | ITCC  Study site |
| Primary outcome | Data entry  Worksheet | Study site  Study site |
| Safety | Data entry  Worksheet | ITCC  Study site |
| Secondary outcome | Data entry  Worksheet | ITCC  Study site |
| Medication compliance | Data entry  Worksheet | Study site  Study site |

# **6. PRIVACY AND CONFIDENTIALITY**

Participants’ privacy and confidentiality will be respected through the protection of their data as outlined in this plan. The Investigator will comply with legal and regulatory requirements regarding the privacy and confidentiality of participants’ data.

Participants have the right to access and correct personal data held by the site.

## 6.1 Breach of Privacy/Confidentiality

A breach of privacy means unauthorised or accidental access to, or disclosure, alteration, loss, or destruction of a participant’s information.

In the event participant privacy and confidentiality is breached during the study, the following steps will be taken:

* Action will be taken to reduce the risk of harm following the breach. Where possible, the recipient will be contacted and asked to destroy or return any disclosed electronic material.
* The participant will be informed of the breach as soon as practicable and provided with support as required.
* A UTS ITCC quality review will be conducted to ascertain factors contributing to the breach, and any corrective action required to prevent future breaches.
* The approving HREC/HDEC will be informed.
* For notifiable privacy breaches of privacy under the relevant and applicable acts
  + Australian Privacy Act 1988, the Office of the Australian Information Commissioner will be notified
  + New Zealand Privacy Act 2020, the New Zealand Privacy Commissioner will be notified in accordance with that Act

# **7. FORMS OF DATA**

Data collected from participants do not include any sensitive personal information other than health information.

## 7.1 Identifiable Data

Some study data are collected in identifiable form.

Source documents refer to identifiable data collected for the purposes of this study. For the purposes of this data management plan, identifiable data includes the participant’s existing medical / clinical records.

Source documents are held at the site in identifiable form.

## 7.2 De-identifiable Data

De-identified data in this study includes but is not limited to:

* Case Report Forms/Data collection worksheets.
* Safety and screening results entered into the analysis data set.
* Communications from the site to the Sponsor (UTS-ITCC).
* Data sent to and generated by the imaging vendor.

De-identified data will carry the participant’s unique study code as follows:

Study code/Site code/ID number: \_ \_/ \_ \_ \_/ \_ \_ \_.

The Investigator will retain a log linking the participant code with personal identifiers such as name, date of birth, contact details including address and telephone number. This log will not be made available to the Sponsor (UTS-ITCC).

All data sent to the Sponsor will be de-identified. All data generated by these parties will be in de-identified form. No attempt will be made to re-identify participants.

# **8. ACCESS TO AND USE OF DATA**

Collected data will be used to answer the research questions and fulfil the study requirements described in the study protocol.

## Identifiable Data

Identifiable data may be accessed by the following groups:

* The Investigator and designated study staff, to fulfil protocol requirements.
* Local radiology staff, to process, analyse and report images
* Sponsor study monitor(s), for eligibility confirmation and source data verification purposes.
* The Sponsor, for audit purposes.
* The Sponsor and its authorised representatives, in the event of a compensation claim by a participant.
* The approving Ethics Committee, for legal and regulatory purposes.
* Health, regulatory, or government agencies, for legal and regulatory purposes.
* The participant’s GP or appropriate specialist, to inform them of study participation, and in the event of an incidental finding of potential clinical significance.

## De-identified Data

De-identified data may be accessed and used by the following groups:

* The Investigator and suitably trained and experienced study staff, to conduct the study.
* Sponsor study monitor(s), for source data verification purposes.
* The imaging vendor, for analysis and reporting purposes.
* The Sponsor, for study conduct, data analysis and pharmacovigilance purposes, product registration and marketing, or as otherwise permitted by applicable local and international laws and regulations.
* The approving Ethics Committee, to comply with legal and regulatory duties.
* Health, regulatory, or government authorities, to comply with legal and regulatory duties.

De-identified data may be included in published study results including, but not limited to, peer-reviewed publications, clinical trial registry websites, scientific meetings, and regulatory / marketing submissions.

# **STORAGE AND DESTRUCTION OF DATA**

## Identifiable Data and Source Documents

During the study, study-specific source documents are maintained in locked filing cabinets in locked rooms, and password protected databases via password protected computers.

Post-study, study-specific source documents are archived on-site or by an external archiving company in locked filing cabinets in a locked room. Archiving of source documents at some sites may be outsourced to an external document archiving company if required per local institutional processes. Archival conditions must have appropriate environmental controls and adequate protection from physical damage (e.g flood, fire and theft security).

Source documents are retained for at least 15 years, then are shredded, pulped, or burnt, within the storage facility. There may be individual country, state and territory requirements that apply to the retention of research materials so the longer period specified will always be applied.

* 1. De-identified Data

Identifiable data are converted to a de-identified form at the study site, at which point it is entered into electronic case report forms using a secure data platform - REDCap. The REDCap EDC platform complies with international and national regulatory requirements for electronic data capture systems in the countries where it is used. Data entry is limited to designated study staff trained and experienced in transcribing data for this purpose.

De-identified data will be carried without reference to health information by using a participant code. The Site Investigator retains a log linking participant code with personal identifiers. This log will not be made available to the Sponsor (UTS).

The de-identified database will remain on REDCap and archived on UTS eResearch servers for up to 15 years from the date of the last publication of the research in line with the state records act of NSW. Electronic records are reformatted, overwritten or shredded, then disposed of through normal channels. Destruction of digital data will be carried out by the UTS Information Technology Division according to International Organisation for Standardisation (ISO) best practice at that time.

# **CONSULTATION**

Consultation regarding data management will be undertaken with the following relevant communities/stakeholders:

## Aboriginal or Torres Strait Islander

Potential participants are asked for their Aboriginal or Torres Strait Islander status. This study has not been reviewed and approved by a culturally specific HREC or committee, but if any person identifies as Aboriginal or Torres Strait Islander the requirements for consultation will be applied, such as consent and information sharing, in order to obtain and meet appropriate requirements. This may require simplification of the health research language, consultation with key community representatives, and providing other appropriate supports. No identifiable data will be generated for publication by Aboriginal status.

# **RETURN OF RESULTS**

Screening and safety results will be provided to participants on request. Participants have the right to request a lay summary of the study results.

## Incidental Findings

In the event that a study assessment returns a result of potential clinical significance, the participant will be informed. The participant’s usual doctor and / or an appropriate specialist will be notified, and follow-up will be arranged.

## Results Arising from Future Research

No future unspecified research is planned for data collected in this study.

# **12. WITHDRAWAL OF DATA**

Participants may withdraw consent for the collection of data at any time, without providing a reason.

Should a participant withdraw consent, no further data will be collected by study staff.

Data collected prior to the participant’s withdrawal will continue to be used and analysed.

# **13. DATABASE ARCHITECTURE**

## 13.1 Database

This web-based REDCap EDC system is developed to support research work by providing access to eCRF, data management and data extraction. The REDCap EDC Database is handled on different levels by these three tools as described in Table 13.1.1**.**

**Table 13.1.1. Database levels and functionalities**

| **TOOL** | **FUNCTIONALITY** |
| --- | --- |
| electronic Case Report Forms (eCRFs) | * Online design of eCRFs and questionnaires * Online data entry (create and edit) from multiple sites with a single coordinating site * Online data checking * Conditioning and masking of unnecessary items * Automatic email alerts (on inclusion, randomisation, adverse events, serious adverse events) * Web-based and email-based form completion * Scheduling (Utilise a built-in project calendar and scheduling module for organising trial events and appointments). * Data queries - Document the process of resolving data issues using the Data Resolution Workflow module. |
| Data Management Tool | * Data monitoring * Online pre-set reports and statistics * Documentation and management * Study progress tracking |
| Data Extraction Tool | * Basic reporting of results with features such as percentages, graphs, and tables * Export data to common data analysis packages - Export your data to Microsoft Excel, PDF, SAS, Stata, R, or SPSS for analysis. * Generate a PDF version of trial forms and surveys for printing to collect data offline. |

## 

## 13.2 REDCap EDC System Security

The system security is granted at several levels:

* REDCap EDC server’s physical plant;
* REDCap EDC system’s disaster-mitigation measures;
* Website security technologies.

### 13.2.1 REDCap EDC server’s physical plant

The REDCap EDC system has been configured for UTS. The REDCap Database server is housed in the Intersect Data Centre at UTS. The physical servers are located in protected premises, with remote control surveillance of local alarms. Physical intrusion, smoke, fire, temperature control, and electrical failure are permanently monitored (24h).

### 13.2.2 The REDCap EDC system’s disaster-mitigation measures

There is a Nessus Professional Security scanner (from Tenable) in the Intersect Data Centre at UTS. It performs a monthly automated scan of the UTS REDCap server. The UTS REDCap instance has the following backup policy in place:

* Instance snapshots via the OpenStack API: The daily backups have a retention period of 7 days, weekly backups have a retention period of 4 weeks and the monthly backups have a retention period of 6 months. Therefore, at any one point there are up to 17 snapshot backups available.
* REDCap data directory snapshots: The REDCap data directory is a Network File System (NFS) mount that is also mounted on another instance that takes daily, weekly, and monthly backups.
* Nightly database dumps using MySQL dump onto the mounted NFS storage prior to the backup.

### 13.2.3 Website security technologies

All services are accessible through a secure internet connection (https://ds.aaf.edu.au/), provided by [AAF Discovery Service](https://ds.aaf.edu.au/discovery/aaf/c32vaiHBrboDWKjKQb0C9g?entityID=https%3A%2F%2Fredcap.research.uts.edu.au%2Fshibboleth&return=https%3A%2F%2Fredcap.research.uts.edu.au%2FShibboleth.sso%2FLogin%3FSAMLDS%3D1%26target%3Dss%253Amem%253Ab706c85948e9813d).

## 13.3 IWRS

An interactive web response system (IWRS) dedicated to randomisation and treatment allocation is integrated in the international collaborator’s EDC system, which is hosted in a facility based in xxx. The Secure Sockets Layer (SSL) certificate is provided by xxx (website).

# **14. DATA SECURITY**

Several measures are in place in order to guarantee the confidentiality and integrity of the trial data.

**Access controls:**

REDCap users can use AAF (Australian Access Federation) to login. Researchers at UTS can log in using their university credentials. External users from non-AAF institutions (e.g., data collectors in hospitals and allied health facilities, health departments and LHDs, etc) need an AAF Virtual Home account to get access.

The Data manager from UTS-ITCC needs to submit a ticket to request access for external users. The decision to approve or deny access to this database is recorded and communicated via email in REDCap to the applicant. If the access is approved, the data manager will apply via UTS Service Connect online to the UTS eResearch Team to organise access to REDCap for the applicant. The applicant receives notification to activate their account via email. Upon account activation, the Data manager provides access to the relevant project as per the details on the approved Access Request Form. Access for third parties is valid for up to a maximum of one year.

The Data Manager maintains a list of all users who have access to the database. External researchers are required to store data on firewall protected file shares with access restricted to relevant project personnel only. This evidence is asked as part of the application for access process.

Each user is identified by a personal, anonymous pseudonym (user code) which is supplemented by an individual password. After five unsuccessful attempts, the account will be automatically locked by the system for 15 mins. Every 3 months the operator is required to change their password; user-chosen passwords cannot include previously used passwords. A system time out will automatically disconnect the user after 30 min of inactivity (this time is calculated by two separate send/receive data from/to the REDCap EDC system) with the scope to limit unauthorised access from a computer left unattended while accessing the REDCap EDC system. Moreover, specific profiles are drawn to users, resulting in different rights (access blocked, consultation only, reading-writing) on each of the server’s functions.

There are several levels of user access to REDCap:

* *Administrator access* – a very limited number of people have password access to enable website design and maintenance.
* *Manager access* – users can check and query data entered by others, with access to the forms provided by the administrators. This level of access enables data correction and query generation, reporting and download functions.
* *Officer access* – users can access a restricted number of forms (allocated by the manager) and are able to view the eCRFs for printing and direct data entry only. Access can be restricted to specific people for specific eCRFs, such as pharmacists to access the randomisation eCRFs.

Nominated and delegated study staff are granted access to enter data only. This requires each user to access REDCap through their unique username and password. This access must not be given to anyone else through the sharing of username and passwords. All users must complete training before being granted access to the REDCap for each study. The Data Manager keeps a record of the access details provided to all levels of users.

# **15. DATA VALIDATION PROCESS**

Process validation is defined as the collection and evaluation of data from the process design stage throughout production, in accordance with the study protocol specifications. The purpose is to ensure that all data are valid for their intended data types and stay valid throughout the application that is driving these data. What this means is that in order to be as successful as it can be, data validation must be implemented in specific parts that get the data, process them and save or print the results, detecting any discrepancy (i.e., data points that fail to pass a validation check).

The major area where the data validation is applied appropriately is the **user interactivity screens and forms**: any section of the e-CRF that requires the user to enter the data is considered as a prime candidate for data validation. Human error always ends up as the prime suspect for invalid data, no matter how well intended the users are.

Different types of data validation are applied to the e-CRF.

* **Field-level validation**: to avoid errors caused by human interaction, specific controls at field levels are put in place for each value to be entered. Controls involve automatic calculations, values to be entered before others and identification of mandatory fields whose failure to compile can prevent the validation of the form.
* **Form-level validation**: when applicable, fields to be validated are clustered into forms/sections (e.g., laboratory, uranalysis, etc.) and no specific field needs to be entered before the others in the form. Thus, the user will enter the information as preferred and validation is performed once, for the whole form.
* **Range validation**: this applies to numeric values and dates. A control is established to make sure that a value entered is within a range of specific values and/or respects the format required as per type of field identification (e.g., date, time, date-time, email, integer, letters only, number, etc.).
* **Masked input validation**: fields, which envisage a specific input pattern that should be followed when data are entered to assure that the right information is included, are entered using a masked or filtered input system (e.g., dropdown menu, calendar pop-ups). Such control is put in place to facilitate the correct data entry.

In addition, to the above-mentioned data validation systems, the following controls are activated on the e-CRF:

* + - Valid value alerts: only “Yes” or only “No” answers, and
    - Multiple-choice fields: in which the user is required to choose at least one checkbox from a series that represents different choices.

Some of these controls, when alarmed, trigger an automatic query that can be either confirmed by the users as true and accurate or corrected (e.g., “One of these fields have been answered as 'NO'. Please check this form as this would mean that this participant is NOT SUITABLE to participate in this trial”).

# **16. VERIFICATION OF REDCap EDC SETUP AND IMPLEMENTATION**

The REDCap EDC system of the [ABBREVIATED STUDY TITLE] is protocol-specific; the steps followed and related competence roles for its setting up and implementation are shown in Table 16.1.

**Table 16.1. [ABBREVIATED STUDY TITLE] REDCap EDC system set up: steps and roles**

| **STEP** | **Roles Of Competence** |
| --- | --- |
| Drafting of the paper CRF corresponding to the protocol requirements and visits | Data manager, Trial coordinators, Chief investigator, Country-specific coordinating investigators |
| Database design | Redcap EDC provider |
| Specification recording and update | Redcap EDC provider, Data manager, Trial Coordinators, ITCC |
| e-CRF design | ITCC |
| e-CRF testing | Country-specific coordinating investigators, Data manager, ITCC trial coordinators |
| e-CRF release | ITCC |
| e-CRF maintenance | ITCC |

# **17. SYSTEM SPECIFICATIONS**

The output from the design process is the specification details of the database to be setup in the REDCap EDC. The data dictionary is a specifically formatted spreadsheet in CSV (comma separated values format) containing the metadata used to construct data collection instruments and fields. It consists of an Excel file containing all the characteristics of the eCRF, including the following details:

* Variable / Field Name
* Form Name
* Section Header
* Field Type (i.e. text, notes, drop-down menu, radio buttons, checkboxes, file, calculations, descriptive, slider, yes/no, true/false)
* Field Label
* Choices, Calculations, or Slider Labels
* Field Note
* Text Validation Type (i.e. date, time, number, integer etc.) OR Show Slider Number
* Text Validation Min
* Text Validation Max
* Identifier (i.e. Does the field contain identifying information (e.g., name, SSN, address)?)
* Branching Logic (Show field only if...)
* Required Field? (i.e. Prompt if field is blank)
* Custom Alignment
* Question Number (Surveys only)
* Matrix Group Name
* Matrix Ranking (i.e. Only be used for a radio button matrix)
* Field Annotation

## 17.1 REDCap EDC User Acceptance Testing (UAT)

Validation involves testing each element of REDCap to reduce the impact on data storage and possibly analysis, due to poor design choices or mistakes in the database creation. It is important to test the essential components of the REDCap project before moving it into production mode. A few test records are created, and some data are entered for each to ensure that the data collection instruments appear and behave as expected, especially branching logic and automatic calculations. Entering and saving data is the only way to test that the branching logic and calculated fields are working properly. An example of testing branching logic is to confirm that each <or a specified subset of> is checked via test data designed to trigger a warning (e.g., not matching the Inclusion/Exclusion criteria warns the investigator to proceed with this participant). Testing automatic calculations is to confirm each automatic calculation set-up works correctly (e.g., age at enrolment and other formula for variables computation, etc.). Test data are reviewed by creating reports and exporting the data to view in Excel or a statistical analysis package.

The [abbreviated study title] REDCap will be designed in a development environment where mock participants can be added. Such an environment can be used during the study for training purposes and modification requests. It will frequently be tested during its development, and it will be approved by selected Principal Investigators or Project Managers prior to moving the project to production mode. The procedure used to test the e-CRF consists of data entry for complete mock participants both reflecting the study characteristics from the inclusion (e.g., screening visit) through the follow-up (e.g., last visit/contact) or early termination to study completion.

# **18. DATA ENTRY**

The access rights of the personnel assigned to the study project should be identified before starting the data collection phase. A training schedule is drafted containing a list of attendees and a potential timeline. The REDCap User Access Dashboard is a reporting tool designed to assist administrators in the management of users that have been granted access to the REDCap projects. The User Access List can be filtered by project status and project purpose. The administrator is responsible for reviewing the User Access List regularly to ensure that each person listed still requires access to each project.

The REDCap user rights can be broken down to the tasks involved in each different role within the project (role assignment). The study owner will assign/re-assign individuals to six different types of Data Access Groups (Project Manager (including Project Manager and Data Manager), Central Monitor, Principal Investigator, Site, Pharmacist, Invoice). Data Access Groups provide the ability to place barriers between the sites' data (i.e., group A cannot see, export, or edit group B's data). The Database access is set up on different levels by the roles and user rights as described in Table 18.1**.**

**Project Manager** Allows read-only access to all the eCRF forms and to the Data Management and Extraction tools. Allows the opening of queries in each editable field of the eCRF forms.

**Central Monitor**  Allows the opening of queries in each editable field of the eCRF forms.

**Principal Investigator** Allows the user to enter data and validate. Typically, it represents the account for Principal Investigators and close deputies. The user is allowed access to forms only concerning its site.

**Site** Allows the data entry without the possibility to validate the data entered. The user is allowed the access only the forms concerning its site of belonging.

**Pharmacist** Allows the read-only access to the Pharmacovigilance section.

**Invoice** Allows read-only access to Invoicing section.

**Table 18.1. Set-up roles and user rights**

|  |  |  |
| --- | --- | --- |
| **User Right** | **Definition** | **Access** |
| Data entry rights | Grants user “No Access”, “Read Only”, “View&Edit”, “Edit Survey Responses” rights to the project’s data collection instruments. | Project manager |
| Expiration date | Automatically terminates project access for the user on date entered. | Project manager |
| Project design and setup | Grants user access to add, update or delete any forms within the project. Also allows user to enable and disable project features and modules. | Project manager, Principal investigator |
| User rights | Grants user access to change the rights and privileges of all users on a particular project, including themselves. | Project manager |
| Data access groups | Grants user access to create and add users to data access groups. | Project manager |
| Data exports | Grants user “No Access”, “De-identified Only”, “Remove all tagged Identifier fields” and “Full Data Set” access to export all or selected data fields to one of the 5 default programs in REDCap (SAS, SPSS, R, Stata, Excel). | Project manager |
| Add / edit reports | Grants user access to build reports within the project. | Project manager |
| Stats and charts | Grants user access to view simple statistics on each field in the project in real time. | Project manager, Principal investigator, Site, Central monitor, Pharmacist, Invoice |
| Survey distribution tools | Grants user access to manage the public survey URLs, participant contact lists, and survey invitation log. | Project manager, Principal investigator, Site |
| Calendar | Grants user access to track study progress and allows user to update calendar events, such as mark milestones, enter ad hoc meetings. | Project manager, Central monitor, Principal investigator, Site, Pharmacist, Invoice |
| Data import tool | Grants user access to download and modify import templates for uploading data directly into the project bypassing data entry forms. | Project manager, Principal investigator |
| Data comparison tool | Grants user access to see two selected records side by side for comparison. This is helpful when using double data entry. | Project manager, Central monitor, Principal investigator, Site |
| Logging | Grants user access to view log of all occurrences of data exports, design changes, record creation, updating and deletion, user creation, record locking, and page views. This is the audit trail for the project. | Project manager, Central monitor |
| File repository | Grants user access to upload, view, and retrieve project files and documents (e.g., protocols, instructions, announcements). In addition, it stores all data and syntax files when data is exported using the Data Export Tool. | Project manager |
| Data quality | Grants user access to find data discrepancies or errors in project data by allowing user to create & edit rules; and execute data quality rules. If user does not have access to a data collection instrument that the query is referencing, access will be denied for query results. | Project manager, Central monitor |
| Create records | Grants user access to add record and data to database. | Principal investigator, Site |
| Rename records | Grants user access to change key id of record. | Principal investigator, Site |
| Delete records | Grants user access to remove an entire record. | Principal investigator |
| Record locking customization | Grants user access to customize record locking text. | Project manager, Principal investigator |
| Lock/unlock records | Grants user access to lock/unlock a record from editing. Users without this right will not be able to edit a locked record. User will need “Read Only” or “View&Edit” to lock/unlock a data collection instrument. | Project manager, Principal investigator |
| Lock/unlock \*entire\* records (record level) | Grants user access to lock/unlock an entire record while one or more instruments are currently locked. | Project manager |

## 18.1 Pre-requisites for Data Entry

Each user must be trained on the REDCap EDC system prior to being granted permission to work in the production mode of the eCRF under ‘[abbreviated study title] Forms and Documents’ REDCap project. A training record will be created and maintained. Training records will be stored in the Trial Master File. Training materials (i.e., slides) can be accessed from the ‘[abbreviated study title] Forms and Documents’ REDCap project. This project is to be used for storing and retrieving files and documents used for the [abbreviated study title]. The UTS-ITCC team upload files to ‘[abbreviated study title] Forms and Documents’ REDCap project for retrieval, so that the site staff can download documents in the file list. In general, the ‘Forms and Documents’ is set up with different instruments as described in Table 18.1.1.

**Table 18.1.1. REDCap project ‘Forms and Documents’ Set-up**

|  |  |  |
| --- | --- | --- |
| **Instrument** | **Definition** | **Access** |
| Study protocol | All versions of protocol, protocol amendments, product information | Project manager, Central monitor, Principal investigator, Site |
| PICFs | All versions of Participant/Caregiver Information Sheet and Participant/Caregiver Consent Form (e.g., Master, Withdrawal, Sub-study) | Project manager, Central monitor, Principal investigator, Site |
| Worksheets | Also called CRFs, including concurrent medication log, AE log, SAE template. It is recommended to be maintained in a folder for each participant. | Project manager, Central monitor, Principal investigator, Site |
| Ethics and regulatory | ANZCTR Trial registration, CTN listing, Insurance, HREC approval | Project manager, Central monitor, Principal investigator, Site |
| Advertising and marketing | All versions of advertising and marketing materials (e.g., staff poster, public poster, clinic card, referral form) | Project manager, Central monitor, Principal investigator, Site |
| Guides and instructions | All types of guides and instructions related to the study (e.g., Instrument scoring manual, COVID script, face-to-face interaction record (COVID), sub-study procedure manual, email template) | Project manager, Central monitor, Principal investigator, Site |
| Assessment Tools –(Participants/Caregiver) | Individual booklets and/or forms of assessments and questionnaires. | Project manager, Central monitor, Principal investigator, Site |
| Other Patient Facing Documents | All types of patient-facing documents (e.g., emergency card, patient ID card, participant study summary, safety poster, drug dosing instructions, patient diary instruction, patient pack instructions and coversheets) | Project manager, Central monitor, Principal investigator, Site |
| Site Initiation Training Slides | The training sessions cover pharmacy, regulations, monitoring and management, protocol and implementation, and data management. | Project manager, Central monitor, Principal investigator, Site,Pharmacist |

Access to the eCRF is not granted until the identified users have completed the training on the eCRF. Such training will be arranged through online site-training workshops, review of the study specific slides, and general REDCap self-directed training slides. REDCap EDC system training and instructions are provided during site initiation for each study. There are also additional training modules available in REDCap. This training ensures that all users are up to date with the use of REDCap and the restrictions and capabilities of the system. Training completion must be documented on the REDCap Training Compliance Form and submitted to the data management team.

Access to the development eCRF is allowed to all investigators as they may be requested to include and randomise at least a test patient by the site initiation visit, under the supervision of UTS-ITCC, if needed.

Subsequent changes and updates to the REDCap eCRFs and database are implemented in response to an identified issue, to correct an error or discrepancy, following a protocol amendment or upon request from the Coordinating Principal Investigator. If the REDCap EDC system is modified in any way, training should be considered depending on the complexity of the modification. This will be a simple user guide or may require more detailed training.

## 18.2 Granting Access to the Production Mode of the REDCap EDC

Once the training is completed, login details are sent via email to the user with the Principal Investigator in copy for information and archiving. To avoid account scam, at the first access every user is asked to change the system generated password received.

## 18.3 Entering Data

The longitudinal model setting in REDCap allows any data entry page to be repeated, any given number of times, across pre-defined time-points, which are specified by the user before data is collected. Data entered into the REDCap EDC system are based on source documentation maintained at the clinical site (Site Investigator File) and should be recorded in the REDCap EDC system only by trained and delegated site personnel with the appropriate level of access.

Data is entered into REDCap, including those collected in worksheets, diaries, and questionnaires. The data is collected by the health services involved in the care of the study participants. Data collection worksheets are provided by the ITCC team to assist sites in ensuring that source documentation and data reported in the eCRFs are complete and consistent. The data collection worksheets are designed to capture the required data and to enable smooth data entry in the eCRFs through similar question and response structures to minimise data entry errors. Data collection worksheets do not contain any personally identifiable information such as name or hospital record number. These data collection worksheets may be considered source documentation (e.g., vitals signs, AKPS etc.), where the worksheet is the first recording of study data. Data in the worksheets and eCRF needs to be supported by other corroborating information, such as admission records, recording of adverse events, clinical visits as applicable.

Data recorded on the worksheets is entered to the REDCap eCRFs. No personally identifying information will be entered into REDCap at any time. Specific care is to be taken when recording clinical information such as notes within Serious Adverse Events (SAE). Each data file is protected within Excel using encryption and password through the ‘Protect Worksheet’ function.

On completion of data entry for each form, the study site submits the data and is required to change the data status from ‘Incomplete’ to ‘Complete’. Any further changes to the data initiate a log of changes in the form of an audit trail. Source Data Verification (SDV) is a process to verify that the data collected in the eCRFs is correct and has been transcribed accurately from the source document. A central monitor at ITCC is provided with a REDCap account with monitoring authorisation which allows the monitor to verify and record in the eCRF whether a datum corresponds or not to the source document checked during the SDV. The monitor is allowed to open manual queries on specific fields in REDCap eCRFs. Changes or queries raised by the ITCC will result in the data status being changed to ‘Unverified’. A copy of the original data collection worksheet is uploaded to REDCap to the ITCC to enable verification of data entry and for filing.

# **19.** **DATA CLEANING**

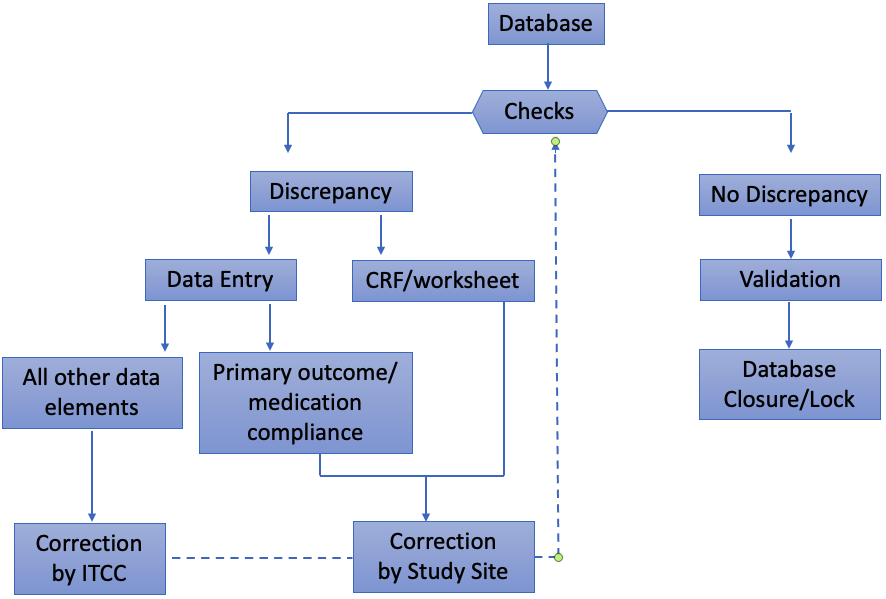
It is essential that quality control is applied to each stage of data handling to ensure that all recorded data are reliable and correctly processed.

Good practice guidelines for data management require transparency and proper documentation of all procedures.

Data cleaning intends to identify and correct possible errors or at least to minimize their impact on study results.

Data is considered clean when the following has occurred:

* All expected data is confirmed by nominated site personnel
* All expected eCRF forms have been completed
* No outstanding discrepancies
* All data review activities completed, and reviewers have no further queries regarding the data



**Figure 1. Data Cleaning Flow**

## 19.1 Automatic Edit Checks

A list of all automatic edit checks has been programmed in the REDCap EDC system. One strategy is to request if the data was collected, with an entry required for either ‘Yes’, in which case the data entry fields are revealed, or ‘No’, in which case no data can be entered. Other types of checks in this list include, but are not limited to:

* protocol violations (e.g., inconsistency in inclusion criteria, out of range visit, etc.);
* missing values, most fields are ‘Required’, if data is missing, the user can enter a reason for the missing data from a range of options;
* missing patient records;
* range checks (e.g., height between 150 and 200 cm); or
* logical inconsistencies (e.g., lab values are confirmed as not done, while lab values are entered).

Such edit checks are always enabled in the eCRF and trigger error detections that can prevent a specific form to be validated, based on the confirmability of the field.

## 19.2 Manual Edit Checks

Manual queries are used in cases where the data issue is sufficiently complex as to be impractical to program as an automatic system check and/or requires human judgment.

A list of all manual checks to be performed on the study data will be created. The type of checks in this list should be amongst others:

* free text;
* cross-check concomitant medication table versus adverse event table;
* SAE reconciliation;
* crosschecks between modules or visits (e.g., order of visit dates).

Manual queries are generated and tracked up to resolution in the REDCap EDC system.

## 19.3 Detecting Outlier

An outlier is an observation that is numerically distant from the rest of the data.

Methods of outlier detection are manual inspections of graphically represented data. This inspection will be conducted at study sample size, or earlier if repeated errors become evident.

The identification of possible outliers is important both for improving of the data quality, following GCP guidelines, and reducing the impact of outlying values in the process of data analysis and data mining.

## 19.4 Missing Data Identification

Missing values may be due to interruptions of the data flow or the unavailability of the target information. It is essential to understand the difference between "handling missing data" for data cleaning purposes.

Handling missing data for data cleaning purposes must answer to the following two questions:

1. Were missing data collected on eCRFs?
2. Were missing data lost during database manipulations?

If the missing data were not collected on the eCRF, it is possible to create a query in order to replace the missing value. The queries that arise during the data reconciliation should be handled in the same manner in which clinical queries are handled.

If the missing data were lost during database manipulations, it is necessary to replace the missing value with the original value (source data).

In the REDCap EDC system, fields that have a blank/missing value may be marked with a custom 'Missing Data Code' to note why the value is blank. These missing codes may be used to aid in data analysis by specifying why a field lacks a value. The REDCap EDC system missing data codes for its setting up and implementation are shown in Table 19.4.1.

**Table 19.4.1. Missing Data Codes**

|  |  |
| --- | --- |
| **Missing Data Codes** | **Definition** |
| DTH | Missing data due to death |
| DPRO | Missing data due to disease progression unrelated to the trial (i.e. the person became more unwell due to their underlying condition but it is specifically not thought to be related to the trial itself - This should ideally be signed off by the local PI) |
| AEI | Missing data due to adverse events/reactions related to the intervention |
| RPO | Missing data related to the primary outcome (i.e. due to the participant being in more pain, less pain or they don't feel the pain has improved so chose to leave fields blank/leave trial) |
| PMQ | Participant missed question by accident |
| RFQ | Research staff forgot to ask/fill out field |
| QUB | Missing data due to question burden (i.e. Participant thought there were too many questions to fill out and did not want to complete question) |
| MEM | Missing data due to participants' mental/emotional health |
| NA | Not Applicable |

# **20. MONITORING**

Monitoring activities are the responsibility of the ITCC. Given the pragmatic nature of the [abbreviated study title] trial, monitoring will be carried out remotely on the eCRFs generated in the REDCap EDC database. It follows that the procedures related to SDV will be conducted by the local PI and by the local study team.

The central monitor at the ITCC checks that eCRF data entry is up to date and monitors completion of participant data collection worksheets. The central monitor discusses with the site PI and local study team the issues identified during data verification activities, by following the REDCap data resolution workflow.

The REDCap Data Resolution Workflow, sometimes called a data query, is a process for managing and documenting resolution of data entry issues. A data query can be initiated on a data entry form or in the Data Quality module when discrepancies are found. Individual users must be granted appropriate user rights to open, respond to, or close data queries. The central monitor is allowed to open a workflow for documenting the process of resolving issues with data in the project (i.e. opening, responding to, and closing data queries).

During the monitoring process, the central monitor documents any issues and deficiencies on a Corrective Actions Sheet. The Corrective Actions Sheet is used to summarise all findings identified during the monitoring visit including any raised or resolved REDCap query. The central monitor is responsible for entering possible solutions for all findings.

# **21. CHANGES TO A REDCap EDC PRODUCTION MODE**

Amendments to the [abbreviated study title] protocol that have an impact on the eCRF or require modifications to the eCRF (e.g., addition of a unit of measure for a specific site, etc.), follow a precise process of drafting, communication, and testing, as summarised in Figure 21.1.

If deemed necessary, the worksheet should be discussed with the Trial Steering Committee during a teleconference for review and final approval. Since the project is in production, changes will not be made in real time. The ITCC team may make changes while entering the project in Draft Mode, which will not affect the existing structure of the project, but all edits will need to be reviewed and approved by a UTS Research IT administrator after finalising the changes. Once all the required changes have been made, the project manager will submit the draft for administrator’s approval. The REDCap administrator will review all drafted changes before they are committed permanently. Once ascertained that the modifications work properly, the administrator confirms the possibility to move the modification on the production environment, so as to be implemented on the real eCRF. The project manager will be notified by email when the changes have been committed to the project.

The ‘project revision history’ on REDCap provides a table that list information about when major changes and revisions are made to the project. The table displays when the project is created, and if the project is in production, lists the time it is moved to production, as well as any revisions made to the project fields while in production. After each revision, the ITCC project data manager is responsible for downloading all the Data Dictionaries from past revisions from the ‘project revision history’ functionality on REDCap.

**Figure 21.1. eCRF modification flowchart**

# **22. DATA DOWNLOAD**

This procedure is undertaken by the project data manager at ITCC, and downloaded onto a password protected network drive, with clear and secure backup and retrieval procedures (refer to section 13.2.2).

On each occurrence, the following procedure is followed:

* Data download is saved to a password protected computer as an excel workbook or CSV text file with file name indicating Case Report Form (or other data file) name and date.
* The downloaded file is viewed for completeness and gross validation. This is not data checking but a check to see that all data fields have been downloaded in a complete and consistent manner (i.e., after five records of randomisations, then at 50% recruitment goal and then upon reaching sample size, unless there is reason to do so earlier).

# **23. DATA BASE CLOSURE**

Database closure occurs when:

* Study recruitment has been stopped
* All data has been entered and checked
* The data meets the definition of ‘clean data’

The database for each study is checked by the ITCC at the time of download for the following elements:

* To identify and remove any duplicate or blank entries
* Correct any participant ID numbers that have been incorrectly entered, and where the ID number can be verified from another source (for e.g., email confirmation or Randomisation Registration Notification)

Changes to the database are recorded within the system audit trail.

After a proper quality check and assurance, the final data validation will be run. If there are no discrepancies, the datasets will be finalised in consultation with the study statisticians.

All data management activities should have been completed prior to database lock. To ensure this, a pre-lock checklist (Data Closure Checklist) will be used, and the completion of all activities will be confirmed. This is done as the database cannot be changed in any manner after locking. Once the approval for locking is obtained from all stakeholders, the database will be locked, and clean data will be extracted for statistical analysis. Generally, no modification in the database is possible. But in case of a critical issue or for other important operational reasons, privileged users will be able modify the data even after the database is locked. This, however, requires proper documentation and an audit trail must be maintained with sufficient justification for updating the locked database.

Data extraction will be done from the final database after locking. This will be followed by its archival.

The main roles acting in this phase are:

* + - **Data Entry staff**: responsible for updating of the database before the final closure analysis.
    - **Project Data Manager**: responsible for completing database closure checks and locking of data in REDCap EDC system and suspending of the access rights to the database.
    - **Project Manager**: responsible for approving the database lock and filing the completed approval form in Trial Master File (TMF).

## 23.1 Closure Checks

Closure checks refer to the verifications that should be performed prior to database lock to verify the integrity and completion of the study database. They will include:

* Check that all expected eCRFs have been entered;
* Check that all the queries are resolved;
* Check that the database is consistent;
* Determine the status of each participant entered (i.e., excluded, ongoing, completed, withdrawn, lost to follow-up, etc.);
* Check for value formatting problems in database exports;
* Confirm that all expected site signatures have been applied.

The data closure checklist is signed and filed in the TMF.

## 23.2 Database Lock and Preparation for Final Analysis

After the date of database closure, all the data are downloaded and filed in a password protected network drive, with clear and secure backup and retrieval procedures (refer to section 13.2.2). All the records are locked in the REDCap EDC system. This point is considered Database Lock. Database lock is the time point, for a clinical trial, at which a database is expected to be clean, all data are completed and consistent, all queries resolved, and a final quality control has been performed, so that the database is ready for final analysis. Once all data has been transferred to the study statistician and any resultant queries have been resolved, the database is locked, all continuing access is ceased, except for the data manager or coordinator.

This process will follow these steps:

* Authorisation for removal of the access rights to the REDCap EDC system using the Data Closure Checklist and Lock Approval Form.
* Removal of access rights to the database done by the database administrator or data manager. Export of data for the analyses. This ‘final’ database should be clearly indicated with the date (and time if applicable). In addition, exported data files are preferable saved as read-only files.

The data manager should document the status of the database on the ‘Data Closure Checklist and Lock Approval Form’. The project is responsible for the approval of the database lock recorded and documented the approval form in TMF.

An audit should be performed of a sample of completed forms in the REDCap EDC system against exported datasets to ensure the integrity of the final study data.

## 23.3 Updating the Database after Database Lock

Updates to a locked database should be limited to important corrections, for instance if the data to be changed have a significant impact on the reliability of the results.

If the statistician has data queries following database lock, they will need to provide justification for requesting the unlocking to the REDCap EDC system provider and detailed reason for changing the data.

Both request and approval will need to be documented on a Database Unlock Request Form.

Only the designated data entry staff member (as detailed on the study delegation log) will be granted access to the database again to implement the required changes.

The database will be re-locked as soon as the corrections have been made to prevent other data changes. Once the database is locked again, a new final database file (not overwriting the original database lock) will be created.

This process will not only be recorded on a new ‘Data Closure Checklist and Lock Approval Form’ by the project data manager, but also approved and filed in TMF by the project manager. Re-locking of the database must be performed according to procedures set out in section 23.2.

An audit trail will be implemented, listing the database lock date, and keeping details of what has been changed and when.

# **24. DATA BASE TRANSFER TO CPI/ STUDY STATISTICIAN**

Dataset Specifications: File Transfer Format

UTS-ITCC will provide a final locked database to the CPI/ study statistician upon request. All files “zipped” into a WinZip file are sent via OneDrive (for business) to CPI/ study statistician. "OneDrive for Business" with its file-sharing security features and encryption (at rest and in-transit), is suitable for transferring the UTS highest classification - "UTS Confidential" data. The UTS-ITCC team sign in with their Microsoft account or staff account to access OneDrive (for business). UTS-ITCC team ‘share’ the files which will insert a link to the OneDrive file in an email message sent to the CPI/ study statistician. Access to the link is granted to the appropriate individuals only. The link will only work until the date set. The expiration date will be set for 24 hours after sending the link. Once it’s expired, the link will be invalid. The password to open the zipped data file is forwarded to the CPI/ study statistician separately via telephone or by email where the email subject line does not link the two emails. When the CPI/ study statistician clicks the link, they will be prompted to enter a password before they can access the file. Once the encrypted data file is opened using the password, the CPI/ study statistician imports the data into an appropriate statistics programme using the import function.

The format for data transfer will be using Excel files. CSV files can be converted to various other formats using commercially available off the shelf software. Each dataset will be provided in a single Excel file, or those datasets which are divided will be clearly named to aid the reviewer in reconstructing the original dataset. The transfer document will identify the range of participant numbers (or other criteria used for division) in the label for each of the divided datasets.

For all datasets, in order to significantly reduce dataset file sizes, the allotted character column length/size for each column will be the maximum length used. All dataset names and dataset labels are unique across the tabulation datasets submitted for the study. The internal name for the locked dataset on these CSV files will be the same as the variable/field name shown in the data dictionary codebook. The key variables will appear first in the datasets. Each participant will be identified by a single and unique participant identifier code.

# **25. REFERENCES**

Krishnankutty B, Bellary S, Kumar NB, Moodahadu LS. Data management in clinical research: An overview. Indian J Pharmacol. 2012 Mar;44(2):168-72.

Paul A. Harris, et al. Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics. 2009;42 (2): 377–381.

Rousseeuw PJ and Hubert M. Robust statistics for outlier detection. WIREs Data Mining KnowlDiscov 2011;1: 73–79. doi:10.1002/widm.2

Van den Broeck J, Argeseanu Cunningham S, Eeckels R, Herbst K. Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities. PLoS Med 2005;2(10): e267.

# **26. APPENDICES**

Appendix 1. Study Protocol

Appendix 2. Annotated CRFs/Data Collection Worksheets

Appendix 3. Corrective Action Sheet

Appendix 4. REDCap EDC System Modifications Sheet

Appendix 5. REDCap Training Compliance Form

Appendix 6. Data Closure Checklist and Lock Approval Form (all sites)

Appendix 7. Database Unlock Request Form