****  SOP 5.0 Data Management

**KEMRI/LBBS**

**Research and Public Health Collaboration**

|  |  |  |  |
| --- | --- | --- | --- |
| **Study Specific Procedure**  **The Longitudinal Bio-behavioral Survey** | | | |
| **Standard Operating Procedure on Data management** | | | |
| HISS Branch | **LBBS SOP No: 03**  **Version:** 1.0  **Effective Date:** 23-JAN-2018  **Supersedes:** N/A  **Pages:** 1 to 9 | | HISS Research |
| **Written By:**  Jully Odhiambo  Date: | | **Reviewing Authority:**  Vivienne Kamire  Date: | **Branch Chief Approval:**  Daniel K. Kwaro  Date: |

1. purpose and applicability

Purpose

The purpose of this Standard Operating Procedure (SOP) is to provide guidelines for all the key aspects involved in Data Management for the Longitudinal Bio-behavioural survey (LBBS) study. It describes the process followed for the collection, transmission, storage and management of the study data.

Applicability

This SOP is applicable to all LBBS study staff involved in the collection, cleaning and analysis of data

1. abbreviations
   * + LBBS – Longitudinal Bio-behavioural survey
     + KEMRI – Kenya Medical Research Institute
     + CDC – Center for Disease Control and Prevention
     + SOP – Standard Operating Procedure
     + PI – Principal Investigator
     + DM – Data Management
     + CRF – Clinical Report Form
     + MO – Medical Officer
     + CO – Clinical Officer
     + DB – Database
     + ODK – Open Data Kit
2. definition of terms
   1. **Source Data** – All information in original records of clinical findings, observations, or any related study activities. It is data in its initial form from the source without any form of modification.
   2. **Source Documents** – Original documents, data, and records (e.g., clinical records, laboratory notes, evaluation checklists, pharmacy/dispensing records, recorded data from automated instruments, copies or transcriptions certified after verification as being accurate and complete, and records kept at the pharmacy, at the laboratories involved in the clinical trial).
   3. **Confidentiality** – prevention of disclosure to unauthorised individuals, of a sponsor’s proprietary information or of a subject’s identity.
   4. **Cleaned Data** – Data that has been validated and checked for errors and inconsistencies
   5. **Electronic Data** – Data in a digital format i.e. machine readable data like images.
   6. **Validation**- Checking to confirm that all the variables have been captured and the data is within the expected range.
   7. **Data Cleaning** – The process of checking for inconsistencies in the data.
   8. **Validation/Cleaning Codes** – These are programs written to help in the processes of checking the data to ensure it meets the required standards.
   9. **Quality Assurance (QA):** All those planned and systematic actions that are established to ensure that the trial is performed and the data are generated, documented (recorded), and reported in compliance with GCP and the applicable regulatory requirement(s).
3. EQUIPMENT AND MATERIALS
   * 1. Computers installed with MS SQL Server or MySQL database systems
     2. Tablets for Data Collection
     3. External Hard drives/Flash Disk (for data transfer or backups)
     4. Data Tracking Forms
4. attachments
5. Data Clarification form
6. responsibilities

This SOP applies to the LBBS research team members involved in data collection, transcription to CRFs, and management of the data.

| Role | Responsibility |
| --- | --- |
| *Data Manager* | * Coordinate and supervise the activities involving the project data. * The data manager will perform specific data analyses, but must be aware of the data management tasks that need to be performed to develop and maintain a high-quality data resource. |
| *Principal Investigator* | * Approval of the SOPs and ensure implementation of the procedures listed in the SOPs |
| *Data Analyst/Specialist* | * Handling and storing the data, developing quality check measures and ensuring the safety and integrity of the data. * Responsible for doing regular downloads and responding to data capturing challenges. |
| *HTS Counsellors* | * Recruiting participants into the study and assigning them study IDs. * Performing tests and filling information in tablets. |

1. procedures
   1. data collection
      * Data will be captured using Tablets (using ODK) at the point of recruitment or subsequent visits.
      * The HTS counsellor/study personnel will be trained to complete the questionnaires in accordance with the following requirements.
        1. Complete all entries on questionnaires
        2. All questions should be answered according to the specific instructions on the tablet
      * All the questions shall be answered according to the instruction on the respective form completion SOP (ODK Form Completion).
      * Data entry will be done at the point of contact with study participants during home visitations.
      * The interviewer will capture the demographic, sexual behaviours, general observations among others
      * An enrolment log will be filled for all study participants
      * The data entry process will be guided by skip patterns and validation rules as specified and/or as pre-programmed in the tablets.
      * The data manager and study personnel will be trained to review, correct and on handling tablets during data capture.
      * Data errors and corrections shall be done in accordance to the data correction and discrepancy resolution section (####) of this SOP.
   2. participant identification

a) Upon enrolment of the participants, a study identification number will be assigned to each participant sequentially; this will be the unique identifier of the participant.

b) The study ID will be used in the tablet questionnaire relating to the participant.

* 1. data transmission, storage and security

Data collected shall be in Electronic format. Data collected using tablets shall be transmitted securely to the study cloud server.

* + 1. **Verification**

1. The database shall synch automatically using the wireless network set up at the facility. This is to allow the data manager access the raw data remotely for reports and queries.
2. The data manager shall keep track of the database changes after verification ensuring that the raw data and the cleaned data are properly tracked.
   * 1. **Transmission**
3. Data shall be transmitted electronically via internet to the KEMRI-CGHR Centers in Kisumu (Kisian). This shall be an automated process.
4. Data collected using tablets shall be transmitted automatically from the field once they have been saved. The transmission will be done using the EDGE/2G/3G/4G technologies depending on availability and the mobile service providers in those regions.
   * 1. **Data Cleaning**
5. Data cleaning and validation codes will be written by the data manager and stored at the central project folder/link. The cleaning codes would be done in STATA, R or SAS.
6. Periodically, the cleaning/validation codes will be run on a copy of the raw data generating reports and queries for resolutions.
7. After cleaning and validation, the cleaned copy shall be stored in the cleaned data folder with the filename indicating the date of the cleaning.
8. The cleaning may involve physical reference to the source documents and the data tracking forms. In some cases it may involve calling the field based staff to confirm/verify specific issues/queries on the data.
9. Once the separate tables have been cleaned and validated, the tables will be merged to come up with complete dataset. The merging shall be done by linking the relevant tables using the appropriate linking keys.
   * 1. **Storage**
10. All study data from all the sites shall be stored in Google Drive online storage facility setup for the study and also backed up in the drop box and the KEMRI-CGHR central file servers in Kisian. Backups shall be stored in the respective laptops and backup copies stored by the data manager on external drives.
11. The study data will be stored in different categories i.e. raw data, cleaned data, and final data.
12. Full data backups shall be periodically done by the study data staff under the supervision of the study data manager.
13. Incremental backups for the scanned patient data shall be done on the respective site laptops and periodically picked by the data manager.
    * 1. All study data shall be triple backed up in different media including cloud, local servers/link folders and external hard drives.
      2. **Data Access and Security**
      3. Only authorized study personnel and any other staff approved by the study PI will have access to the study data.
      4. **Raw** data will be the electronic source data from the CRFs or downloaded from the Aggregate Server, **cleaned** data will be the raw data that has been checked for inconsistencies, corrected, and validated. **Final** data will be the clean data that has been merged from the data sets.
      5. Tablets used for data collection process shall be installed with antivirus software to ensure the devices are safe from virus/malware attacks. The antivirus software shall be updated regularly by the data assistants and recorded on the data security log.
      6. The tablets will be customised /kiosked so as to allow only the data entry application to run on the device to avoid potential failures.
      7. Tablets and laptops designated for data management will be password protected and the data therein encrypted, and used exclusively for data management. Each user will be assigned user access accounts and roles depending on the roles in the study.
      8. Data tracking and correction logs will be used to track corrections and locations of all study forms and logs by the relevant study personnel.
      9. **Physical security** – the study rooms and offices where study files, essential documentation and study equipment used in data collection and management shall be protected from possible disasters such as water or fire damage and installed with an access control system to prevent unauthorised access.
      10. **Logical security** – All data shall be encrypted and access limited to specific study personnel with the decryption keys installed on their laptops ensuring only project specific laptops have access to the data. Together with the encryption, there shall be a three tier authentication (identification, authentication and authorization) process subjected to all authorized personnel in accessing the data.
      11. Privileges for physical or electronic access to data shall be granted to personnel and updated according to the roles and responsibilities as defined by the PIs.
          + The Data Manager shall keep a log of the data backups.
      12. **Data Analysis and reporting**
      13. There shall be periodic preliminary analyses and reports for evaluating certain aspects of the study. These analyses will be coordinated and documented to ensure that there is timely delivery and reporting. Requests will be done through the Data manager of the study who may then delegate.
    1. QUALITY CONTROL
       1. **Quality Checks**
       2. The interviewer supervisor shall ensure that all the questions are duly filled, completed and verified before sending the data.
       3. Errors found in the questionnaires shall be recorded in the data clarification log and appropriate communication and action taken.
       4. There will be a backup copy of the study data in more than one location.
       5. Access to the stored data shall be restricted and only authorized personnel will be able to view and/or modify the data. A database correction log will be kept by the data management team that will document all the modifications/corrections on the data.
       6. The Data Manager will run validation codes to check for inconsistencies in the data transmitted from the health facilities.
       7. **Query Resolution**
       8. The data manager and the study personnel will be trained on query resolution in accordance with the practices. This involves checking communication systems used for making queries such as email and memos, within a period of two weeks.
       9. Once the data manager and the study personnel have resolved the queries, the relevant information in the database will be updated by the data manager and recorded in the data correction log.

**DOCUMENTN CONTROL SECTION**

**SOP tracking review log**

**Purpose:** The log records this SOP’s review dates and the status of the review. The Tracking Changes and Version Control Log are completed to detail status of the review.

**When:** The SOP is reviewed every two years or more often when necessary.

**By whom:** The SOP is reviewed by staff directly following the SOP (e.g. Data manager, Interviewer, etc). The review process is overseen by site study coordinator/designee and reviewed by QA staff.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Supersede Version number / version date** (DD/Mmm/YYYY) | **Review date**  DD/Mmm/YYYY) | **Review status:**  **Changes made: Y/N**  (Complete Tracking Changes Log below) | **Reviewer’s name and initials** | **QA staff name** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**SOP Tracking Changes and Version Control Log**

**Purpose:** The log records this SOP’s changes made to the SOP and rationale for changes. If no changes are made to the SOP this is also recorded as: ‘no change was made’. If no change to version is made the supersede version remains current version.

**When:** Every two years or more often when necessary.

**By whom:** By staff directly following the SOP.

|  |  |  |  |
| --- | --- | --- | --- |
| **Supersede Version / Issue date** | **Revisions/Reason for Change/Rationale** | **Current**  **Version / version date** | **Reviewer’s name** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**SOP Copy Control Log**

**Purpose:** The log records the number of certified copies of this SOP printed and where they were distributed.

**When:** Whenever the SOP is reviewed: Two-annually or more often when necessary.

**By whom:** By QA staff / designee

|  |  |  |  |
| --- | --- | --- | --- |
| **Distribution Date:** | | **Total number of certified copies** (including Master Copy): | |
| **SOP Distribution (location and number of certified copies)** | | | |
|  | | | |
|  |  |  |  |
|  |  |  |  |