|  |  |  |  |
| --- | --- | --- | --- |
| **GREAT HIVCORE 006 DATA MANAGEMENT PLAN** | | | |
| Version Number and date:  Supersedes version: None | | | |
| **Author** | **Position** | **Signature** | **Date** |
| Paddy Kafeero L | Data Manager |  |  |
| **Reviewed by** |  |  |  |
|  |  |  |  |
|  |  |  |  |
| **Approved by** |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |
| --- | --- |
| **Protocol Short Title:** | HIV-CORE 006 |
| **Protocol Title:** | A Phase 1 Trial of ChAdOx1- and MVA-vectored Conserved Mosaic HIV-1 Vaccines in Healthy, Adult HIV-1-negative Volunteers in Eastern and Southern Africa**.** |
| **Study Design:** | A double-blinded, randomised, placebo-controlled, multi-centre trial assessing safety and tolerability of a prime boost vaccine regimen utilising non-replicating simian adenovirus (ChAdOx1) followed by non-replicating poxvirus modified vaccinia virus Ankara (MVA) in adults in Eastern and Southern Africa in healthy adults aged 18-50 |

|  |  |
| --- | --- |
| **Clinical Research Centers:** | 1. Centre for Family Health Research in Zambia (CFHRZ)   CFHRZ-Lusaka  B22/F737 Mwembelelo Road  Lusaka, Zambia     1. KAVI– Institute for Clinical Research (KAVI-ICR)   College of Health Sciences, University of Nairobi  PO Box 19676- 00202, Nairobi, Kenya     1. The KEMRI-Wellcome Trust Research Programme (KWTRP) Kenya Medical Research Institute (KEMRI) / Wellcome Trust Centre for Geographic Medicine Research – Coast (CGMRC) PO Box 230, Kilifi, Kenya      1. MRC/UVRI & LSHTM Uganda Research Unit (MRC/UVRI & LSHTM) Medical Research Council/Uganda Virus Research Institute Uganda Research Unit. Uganda Virus Research Institute Plot 51-59 Nakiwogo Road, Entebbe, Uganda P.O. BOX 49 Entebbe, Uganda |
| **Sponsor:**  Data  Management  Coordinating  Centre: | University of Oxford, United Kingdom  MRC/UVRI and LSHTM Uganda Research Unit, Entebbe, Uganda |

Table of Contents

[Definitions/abbreviations 6](#_Toc60669021)

[1. Introduction 7](#_Toc60669022)

[1.1 Purpose 7](#_Toc60669023)

[1.2 Updating the Data Management Plan 7](#_Toc60669024)

[1.3 Data Management Plan review 8](#_Toc60669025)

[1.4 Trial information 8](#_Toc60669026)

[2. Description of the data 9](#_Toc60669027)

[2.1 Types of data 9](#_Toc60669028)

[2.2 Format and scale of the data 9](#_Toc60669029)

[3. Data collection/generation 9](#_Toc60669030)

[3.1 Methodologies for data collection/generation 10](#_Toc60669031)

[3.2 Data quality and standards 10](#_Toc60669032)

[4. Responsibilities and timelines 10](#_Toc60669033)

[4.1 Responsibilities 10](#_Toc60669034)

[4.2 Organisational responsibilities 11](#_Toc60669035)

[4.3 Data management activities 11](#_Toc60669036)

[5. eCRF design and data definition 11](#_Toc60669037)

[5.1 CRF design 12](#_Toc60669038)

[5.2 CRF review 12](#_Toc60669039)

[5.3 Trial CRFs/eCRFs 12](#_Toc60669040)

[5.4 Updating CRFs 13](#_Toc60669041)

[5.4.1 Making changes to existing CRFs 13](#_Toc60669042)

[5.4.2 Handling and distributing new versions of CRFs 13](#_Toc60669043)

[5.5 Database definition 14](#_Toc60669044)

[5.5.1 Database access for Oxford, MRC/UVRi & LSHTM Uganda Research Unit in Entebbe, OXUS Technologies and CRC staff 14](#_Toc60669045)

[5.5.2 Making changes to the database 14](#_Toc60669046)

[6. Participant management 15](#_Toc60669047)

[6.1 Trial Number/Study Number/Participant ID 15](#_Toc60669048)

[6.2 Randomisation 15](#_Toc60669049)

[6.3 Registration/randomisation procedures 15](#_Toc60669050)

[6.4 Participant tracking 16](#_Toc60669051)

[6.4.1 Missed visits 16](#_Toc60669052)

[7. Data receipt 16](#_Toc60669053)

[7.1 Data receipt at CRCs 16](#_Toc60669054)

[7.2 Data checking on entry 16](#_Toc60669055)

[7.3 Priority data checking 17](#_Toc60669056)

[8. Data entry 17](#_Toc60669057)

[8.1 Data Entry Procedures 17](#_Toc60669058)

[8.2 Unavailable Data Items 18](#_Toc60669059)

[9. Quality management 18](#_Toc60669060)

[9.1 eCRF validation 18](#_Toc60669061)

[9.2 Data entry 19](#_Toc60669062)

[9.2.1 New staff 19](#_Toc60669063)

[9.3 On-site quality control of CRFs 19](#_Toc60669064)

[10. CRF tracking 19](#_Toc60669065)

[10.1 Generating/identifying overdue CRFs and/or missed visits 19](#_Toc60669066)

[10.2 Sending reports of overdue CRFs 19](#_Toc60669067)

[10.3 Managing missed CRFs 19](#_Toc60669068)

[11. Query Handling 20](#_Toc60669069)

[11.1 Generating/identifying queries 20](#_Toc60669070)

[11.1.1 Internal to the trial database 20](#_Toc60669071)

[11.1.2 External to the trial database 20](#_Toc60669072)

[11.1.3 Queries raised by the trial lead Data Manager 21](#_Toc60669073)

[11.2 Sending/receiving queries 21](#_Toc60669074)

[11.3 Handling irregular query responses 21](#_Toc60669075)

[12. Data management, documentation and curation 21](#_Toc60669076)

[12.1 Managing, storing and curating data. 21](#_Toc60669077)

[12.2 Metadata standards and data documentation 21](#_Toc60669078)

[12.3 Data preservation strategy and standards 22](#_Toc60669079)

[13. Storage of participant related documents 22](#_Toc60669080)

[13.1 Prior to data entry 22](#_Toc60669081)

[13.2 After data entry 22](#_Toc60669082)

[13.3 Unblinding 22](#_Toc60669083)

[Scheduled unblinding 22](#_Toc60669084)

[Unscheduled unblinding 22](#_Toc60669085)

[14. Data extraction 22](#_Toc60669086)

[14.1 Analyses 23](#_Toc60669087)

[14.2 Final analysis 23](#_Toc60669088)

[14.3 Data quality checks prior to analysis 23](#_Toc60669089)

[14.4 Data extraction procedure 23](#_Toc60669090)

[15. Monitoring 23](#_Toc60669091)

[15.1 Regular Reports 23](#_Toc60669092)

[15.2 Ad-hoc Reports 24](#_Toc60669093)

[16. Data freeze and database lock 24](#_Toc60669094)

[16.1 Data freeze 24](#_Toc60669095)

[16.2 Data quality checks prior to lock 24](#_Toc60669096)

[16.3 Database lock 24](#_Toc60669097)

[16.4 Database lock procedure 25](#_Toc60669098)

[17. Archiving 25](#_Toc60669099)

[18. Data security and confidentiality of potentially disclosive information 25](#_Toc60669100)

[18.1 Formal information/data security standards 25](#_Toc60669101)

[18.2 Main risks to data security 25](#_Toc60669102)

[19. Data sharing and access 26](#_Toc60669103)

[APPENDICES 27](#_Toc60669104)

[APPENDIX 1: Data Management Plan Review History 27](#_Toc60669105)

[APPENDIX 2: Priority forms 28](#_Toc60669106)

# Definitions/abbreviations

**REDCap:** Electronic Data Capture (EDC) System to be used by the trial to collect and store the trial data.

**Case Report Form (CRF):** CRFs are called Data Collection Instruments in REDCap. A printed (pCRF) or electronic (eCRF) document designed to record all the protocol required information to be reported to the Sponsor on each trial subject.

**PI:** Principle Investigator

**CRC:** Clinical Research Centre

**DM:** Data Manager

**CRF data:** Data that will be entered in the trial specific CRFs. This usually (but not always) refers to all the data that is generated at the trial site during the process of undertaking all the trial procedures as defined in the trial protocol. Details of procedures, examinations, decisions, laboratory results etc. will be available in the trial site data systems and or the patient medical records. This provides the source data for CRF completion.

**Non-CRF data:** Data derived in the trial other than the CRF data, which is not recorded in the CRFs but recorded in other approved and validated systems. There could potentially be multiple sources of non-CRF data.

**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 data are contained in source documents (original records or certified copies).

**Source documents:** Original documents, and records (e.g. hospital records, clinical and office charts, laboratory notes, memoranda, subjects' 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, subject files, and records kept at the pharmacy, at the laboratories and at medico-technical departments involved in the clinical trial).

**Critical Data:** Data such as safety data, data items contributing to the derivation of primary and secondary end-point assessments and eligibility verification that if missed will compromise data integrity and will have a significant impact on the conclusion of the trial. This data is subject to monitoring (either through source data verification of central monitoring) to ensure reliability.

**DMEC:** Data Monitoring and Ethics Committee. Responsible for continual review of safety data and assessment of safety events that result in a study pause.

**MUL:** MRC/UVRI & LSHTM Uganda Research Unit

# Introduction

* 1. **Purpose**

This data management plan describes the processes that will be undertaken to assure the integrity of the data collected and processed as part of the HIV-CORE 006 Trial Study. The plan applies to the above-named study only and will be applied up to the point of release of data sets for statistical analysis. This excludes any procedures relating to the handling and processing of serious adverse events (SAEs), which should be documented in a Safety Management Plan (SMP), or equivalent.

The purposes of this plan is to:

* Ensure that all personnel to partake in the Data procedures know their responsibilities clearly
* Ensure the validity and integrity of the data
* Ensure that the data collected will be usable to monitor safety and to address the objectives of the trial
* Ensure the data are as complete and accurate as possible, and satisfy the requirements for audit and inspection by regulatory bodies

All trial study data will be stored in a single study database, having been entered in real-time on the HIVCORE006 trial eCRFs used in the clinical research centers (CRC). Details of the data external to this database mentioned here are provided in this document as well.

The DMP has been prepared by the data manager and has been reviewed and agreed upon by the trial clinical team. The plan specifies the procedures that will be carried out in order to ensure an accurate and validated study database and specifies the quality control tasks that will be carried out in order to achieve the goals of the plan.

All trial staff involved in data management must receive training with particular emphasis on the data management procedures and the production of the study progress and Protocol Trial Safety reports. They should also be familiar with the relevant sections of the Study Protocol, Study Operations Manual (SOM), CRF completion guidelines, Safety Management Plan, DMP and other Standard Operating Procedures (SOPs) relating to the study.

A Statistical Analysis Plan (SAP) for interim and final analyses will also be developed, prior to the analysis of the data, to ensure planning and discussion of the key concepts and questions of the study amongst members of the study team. The SAP is contained in a separate document.

* 1. **Updating the Data Management Plan**

The DMP should be reviewed at least annually, as well as following any changes to data management working practices. When a new version of the DMP is created, it must be reviewed and approved by the required personnel and a reason for the revision(s) must be listed in the Data Management Plan Review History table.

When a Data Manager (DM) leaves MRC/UVRI & LSHTM Uganda Research Unit (MUL), moves between trials or changes roles, they must ensure that the DMP is up to date before they leave.

* 1. **Data Management Plan review**

The DMP is submitted for review to the trial clinical team, along with all other quality management documentation.

This review should be carried out prior to project start and when there has been a significant change to the documentation, for example following a DMP annual review or a substantial protocol amendment. The purpose of the review is to ensure all issues identified are added to the DMP and other relevant quality documents (where applicable). The review will also ensure compliance with MUL SOPs and legibility.

* 1. **Trial information**

The Primary, secondary and exploratory outcome measures, and the associated electronic Case Record Forms (eCRFs) are listed in Table 1. The study will be a double-blinded, randomised, placebo-controlled, multi-centre trial assessing safety and tolerability of a prime boost vaccine regimen utilising non-replicating simian adenovirus (ChAdOx1) followed by non-replicating poxvirus modified vaccinia virus Ankara (MVA) in adults in Eastern and Southern Africa in healthy adults aged 18-50. Participants will be randomised into one of 2 groups, receiving vaccine regimen or placebo in a ratio of 72:16:

* Vaccine Arm (ChAdOx1.tHIVconsv1 prime followed by MVA.tHIVconsv3 and MVA.tHIVconsv4 boost at 4 weeks after enrolment); 72 vaccine recipients.
* Placebo Arm; 16 recipients

The primary endpoint isto find the:

* Proportion of volunteer who experience local and systemic reactogenicity events.
* Proportion of volunteers with Grade 3 or 4 unsolicited adverse events.
* Proportion of volunteers with vaccine related serious adverse events.
* Proportion of vaccine recipients developing HIV-1 specific T-cell responses.

# Description of the data

## Types of data

i) Identifiable data: these data will include participant study number, full names, date of birth and detailed locator data (e.g. contact phone numbers, physical address including a detailed description of home location and surroundings). Identifiable data will not be captured on the study database along with other research data; it will instead be kept separately under double lock and key at the CRCs.

ii) Other research data: These data will include socio-demographic characteristics, medical history, concomitant medications, attendance of study visits (i.e. retention), HIV serology, vaccination details, risk assessment forms, safety information (e.g. vaccine reactogenicity), physical examination data (e.g. height, weight etc.) and lab data(hematology, chemistry, pregnancy, urine) among others. These data are captured in the trial database and tracked by participant study number and visit number.

## Format and scale of the data

All the trial data will be entered in REDCap. The datasets will be downloaded in Excel/STATA format and analysed using STATA.

# Data collection/generation

## Methodologies for data collection/generation

The data will be collected through interviews and laboratory measurements and recorded in protocol defined CRFs in the REDCap database by project staff at the CRCs.

The generated data files will be labelled appropriately, filed in labelled folders to ensure easy retrieval, and secured electronically on a Sharepoint, each CRC will only have access to their CRC data.

## Data quality and standards

The study database will be developed in REDCap software, an ICH-GCP compliant platform. The database will be hosted on a server at MUL in Entebbe, Uganda. Data collection and entry will be conducted at study sites in Uganda, Kenya and Zambia. The MUL Clinical Information Technology (IT) department will provide and maintain Internet Protocol (IP) addresses linking the study clinics to the database.

Standardised Trial eCRFs in the trial REDCap database will be used to collect and record data. The project staff who will perform data collection and entry will be trained in principles of ICH-GCP. Project staff will also be trained on the study database, completion of the eCRFs and data quality control procedures. The database will be designed to quality control the data at entry and produce reports including: creation of an audit trail to monitor timeliness and accuracy of data entry as well as track records of data corrections/changes in the database, query checks, tracking missing data, and checking data consistency within individual eCRFs and across visits.

The CRC Data Manager(DM) will extract a dataset from the database on a daily basis, analyse for key indicators of study progress (screenings, enrolments, and retention), produce summary statistics and raise data queries. Queries will be sent via email to relevant CRC staff for resolution on a daily basis. The queries will also be discussed during weekly project staff meetings to identify any trends and staff that require further training. In addition, queries will be raised centrally by the DM at the data coordinating centre through reports that will be assigned to the CRC DMs.

The data will be stored on Share point on the MRC/UVRI & LSHTM Uganda Research Unit server in Entebbe, Uganda.

# Responsibilities and timelines

## Responsibilities

In addition to the Principal Investigator (PI) and the project team, the MUL Data Platform, headed by a Senior DM, will have responsibility for the creation of the database as well as ensuring data quality and security. They will also be responsible for ensuring that data that are to be shared are in an appropriate format and are as confidential as possible.

## Organisational responsibilities

| **Table 2: Organisational responsibilities** | | | | |
| --- | --- | --- | --- | --- |
| **Data Management Responsibilities** | MRC/UVRI & LSHTM | Participating CRCs | IAVI | OXFORD |
| Creation/updating of Data Management Plan |  |  |  |  |
| CRF design and maintenance | X |  |  | X |
| Database design | X |  |  | X |
| Build, test and validate databases | X |  |  | X |
| Data receipt procedures |  | X |  |  |
| Data entry |  | X |  |  |
| Randomisation | X | X |  |  |
| Data query generation | X | X | X | X |
| Timely data query resolution |  | X |  |  |
| Central monitoring | X |  | X | X |
| Data coding | X |  |  |  |
| MedDRA coding |  |  |  | X |
| Reporting of trial data | X |  |  | Independent Data Monitoring Committee (IDMC) |
| Data extraction from trial database | X | X |  | X |
| Data analysis (interim & final) – quantitative |  |  |  | X |
| Database Lock | X |  |  |  |
| **Data Management Responsibilities** | MRC/UVRI & LSHTM | Participating CRCs | IAVI | OXFORD |
| Creation/updating of Data Management Plan | X |  |  |  |
| CRF design and maintenance | X |  |  | X |
| Database design | X |  |  | X |

## Data management activities

The CRC coordinators will have the primary responsibility of ensuring that all collected data are of high quality and integrity. The DM(s) are responsible for the day-to-day data management processes at the sites. The central DM is responsible for the overall day-to-day central data management. The Head of Data Management at MUL and at Oxford have oversight of the data management processes and should, in discussion with the Trial Manager (TM), delegate where appropriate any ad-hoc data management activities. Data management support/oversight is also available from the trial statisticians.

The central DM is the first line contact person and will, in consultation with the head of data, ensure tasks are carried out appropriately. Details about all personnel involved in data management, and the trial in general, can be found in the HIVCORE006 Delegation Log.

# eCRF design and data definition

## CRF design

The HIVCORE006 study uses eCRFs that are designed in the trial database setup in the REDCap online datamanagement system. Paper backups are created from the database and provided to the CRCs and can only be used due to database failure or internet failure at the CRCs. The CRFs in the database are meant to be the source documents for the study, as the data is meant to be entered in real-time. However, where relevant, the eCRFs will be substanciated by source documents.

Guidance text in both the database and paper CRFs should be easily distinguishable from question text. Branching logic has been implemented in the database as much as possible to facilitate the skip instructions that are required on the eCRFs.If the guidance text is related to more than one question, it will be placed above the section it is referring to.

## CRF review

All versions of all trial CRFs have been, and will be, reviewed and approved by the trial clinical team. The eCRFs will be downloaded from the database by the lead Data Manager who will keep copies of all the different versions of the CRFs for the tracking process.

## Trial CRFs/eCRFs

As indicated, the Form Submissin Schedule(FSS) outlines the required CRFs for each visits includes the visit numbers. It also indicates when and what to evaluate on the forms on each of the visits.

The table below shows all CRFs for the trial and the corresponding visit numbers for each.

| **Code** | **CRF Name** | **Visits** |
| --- | --- | --- |
|  |  |  |
|  | **Eligibility checklist** | **01** |
|  | **Enrollment** | **02** |
|  | **Randomisation** | **02** |
|  | **IP Administration** | **02,04** |
|  | **HIV-1 Risk Assessment** | **01,02,04,06,07,08,09,10,11** |
|  | **HIV Virology/Serology** | **01,02,04,06,08,09,10,11** |
|  | **Concomitant Medications** | **01,02,02A,03,04,04A,05,06** |
|  | **Physical Exam** | **01,02,02A,04,04A,05,06,07,08,09,10,11** |
|  | **Previous Condions** | **01** |
|  | **Interim Medical History** | **02,02A,03,04,04A,05,06** |
|  | **Urinalysis** | **01,06,10** |
|  | **Hematology** | **01,02,02A,04,04A,06,09,11** |
|  | **Chemistry** | **01,02,02A,04,04A,06** |
|  | **Reactogenicity Assesment** | **02,02A,04,04A** |
|  | **Adverse Events** | **02,02A,04,04A,05,06,07,08,09,10,11** |
|  | **Pregnancy Report** | **02,02A,04,04A,05,06,07,08,09,10,11** |
|  | **Child followup** | **02,02A,04,04A,05,06,07,08,09,10,11** |
|  | **Miscellaneous Tests** | **01,02,04,06,08,09** |
|  | **Sample Summary form** | **01,02,02A,04,04A,05,06,08,09,10,11** |
|  | **Visit documentation** | **0,02,03,04,05,06,07,08,09,10** |
|  | **Volunteer Status** | **02,02A,04,04A,05,06,07,08,09,10,11** |
|  | **Deviation report** | **01,02,02A,04,04A,05,06,07,08,09,10,11** |
|  | **Social Impact Assessment** | **11** |

All eCRFs are expected to be completed by designated CRC staff (ie the nurses, laboratory staff, clinicians and/or Data management team weher necessary)*.* All relevant source data is expected to be collected at CRCs including the reactogenicity and memory aids that the trial participants fill and present to the CRCs in the visits 02A and 04A at the CRCs.

## Updating CRFs

* + 1. **Making changes to existing CRFs**

Should updated CRF versions be required during the trial (e.g. due to a protocol amendment), this will be agreed by the Clinical team. It will also be agreed who will make the updates, who will review and approve, and the timelines for this. The HOD will be informed of any potential database changes as early as possible in the CRF update process.

A detailed version history can be found in the CRF Completion Guidelines.

* + 1. **Handling and distributing new versions of CRFs**

When they have been finalised, new CRF versions will be sent to CRCs by email*.* CRCs will be informed that they must use the new CRF versions from a provided effective date, which will be agreed by the TMT prior to release of the CRFs. After this date has passed, all superseded CRF versions will not be entered onto the trial database, but will instead be returned to the clinic CRC with a request for the correct version to be completed.

The CRF Completion Guidelines will be updated and circulated with the updated CRFs.

## Database definition

The main trial database has been programmed in REDCap and was created in accordance with the MUL Database Development SOP.

All project documentation, including project specification, requirements specification, annotated CRF, functional specification, testing, validation report and go-live sign-off can be found on the MUL stats server in the HIVCORE006 Trial folder.

Changes to the database and data cleaning will be carried out only by the DM. The database will be backed-up daily on the server in Entebbe according to the backup SOP. Access rights to the database will be as per request from trial PI/designee.

The HIVCORE006 unblinded data (which includes the randomisation database) has been programmed in REDCap and is to be kept on the same server (but different database) at the MRC/UVRI & LSHTM Uganda Research Unit in Entebbe, Uganda.

The randomization database is hosted on the same server and will be generated from a Stata program written by a blinded MRC/UVRi & LSHTM Uganda Research Unit statistician, and shared with the Oxford unblinded statistician to create the final randomization database. The blinded data managers and other blinded trial staff will have no access to the generated randomisation database.

* + 1. **Database access for Oxford, MRC/UVRi & LSHTM Uganda Research Unit in Entebbe, OXUS Technologies and CRC staff**

Access to the trial and randomisation databases is only granted to authorised members once relevant database training has been completed. Following this, the new staff data checking procedure needs to be completed.

Staff members with database access are listed, along with the relevant training received, on the HIVCORE006 database delegation log. This is stored in the Trial Master File (TMF). For new members of staff, a completed delegation log should be passed to the MUL data team for database access to be granted.

* + 1. **Making changes to the database**

Changes will be made to the trial database as appropriate in the following situations:

* Following any relevant changes to trial CRFs (i.e. excluding minor changes such as formatting)
* In response to faults discovered in the database, occasions where the database is found not to match the CRF, or if it is agreed that any features of the database are not fit for purpose
* In preparation for other database activities such as development of progress reports

All requests for changes, as well as any testing of those changes and related documentation, will be stored on the MUL stats server. All those involved in this process are granted access to HIVCORE006 stats folder. When staff leave MUL, their access will be revoked via the same process.

Development of database changes relating to CRF changes or new CRFs should begin as early as possible in order that they can be released at the same time as the new CRFs. The Head of DataManagement will be informed of any changes as early as possible to allow the required work to be scheduled.

The level of testing required for each set of database changes will depend on the nature and extent of the changes, and will be agreed with the The Head of DataManagement prior to the changes being impletemented.

When all changes are ready to be implemented in the database, all users will be advised that they must not use the live database during the implementation process. All users will be informed when the database can be used again.

# Participant management

## Trial Number/Study Number/Participant ID

Volunteers in the trial shall be given a 9-character participant ID at the screening visit(01) which comprises of the HIVCORE 006 Protocol number (N006), the Clinical Research Center number[KAVI-ICR Kangemi(22), CFHRZ Lusaka(23), KEMRI-Kilifi(17) and MRC/UVRI & LSHTM-Masaka(19)] and the unique 3-digit Volunteer identification number consecutively allocated by the research centre.

For example the first 2 participants to be screened at MRC MASAKA will have the following Volunteer IDs: N006-19-001 and N006-19-002.

## Randomisation

Volunteers will be randomised according to the randomisation schedule prepared by the statisticians at the Data Coordinating Centre (DCC) prior to the start of the study. The ratio of allocation is 72 vaccineess:16 placebo. These will be recruited across 4 sties, each with 18 vaccines and 4 placebo. Volunteers will be automatically assigned a unique allocation number as they are enrolled into the data entry system, corresponding to an allocation number on the unblinding list that will be provided to the unblinded site pharmacist by the DCC, please see Statistical Analysis Plan.

## Registration/randomisation procedures

Prior to randomisation of the trial participants, the study physician should confirm that the consent process has been completed and properly documented and review of all the screening results and check for any changes in medical history and concomitant medication since the last assessment.

If a subject is found eligible for trial participation, the trial team member who is enrolling the participant will log onto the randomisation database using their personal access credentials. They will then input the Volunteer Id, Sex, eligibility checks and CRC of the subject.

The staff member doing the randomiastion shall then be prompted to confirm the sex, eligibility criteria and that they want to randomise a new study participant into the trial. Once confirmed, the participant is then randomised and assigned a unique allocation number that is linked to a Vaccine arm or placeboo arm. The displayed screen with the assigned randomisation code is then printed out and put in the participant binder.

Thereafter, the person who is enrolling the participant will record the randomisation code on the relevant CRFs thereby linking the participant’s study ID to their randomisation number. The web-based system will similarly record that the study ID has been enrolled in the trial and link this to their randomisation number.

Further details of the randomisation procedures are provided in the randomisation SOP for the HIVCORE006 trial.

## Participant tracking

* + 1. **Missed visits**

Refer to the trial SOM.

# Data receipt

Data will be collected on the eCRFs by CRC staff (Nurses, clinicians or laboratory technicians) on the designated database which will be managed by Oxford and MRC/UVRI & LSHTM Uganda Research Unit, Entebbe. Data shall only be entered on the paper CRFs in case of a failure of the REDCap database or internet failure at the sites and later transcribed to eCRFs when REDCap/internet services are restored. This information must be documented, and these CRFs also stored as source documents.

## Data receipt at CRCs

On entering the data into the eCRFs and on the paper CRFs(when necessary), the details of the staff who enter the data in the eCRFs are automatically captured whereas for those that verify (wehere applicable) the data of the eCRFs must enter their initials and the date at which they verified/reviewed the form.

## Data checking on entry

The site Data Managers are respnsible for the checking of all the data that has been entered on the database for completenes and accuracy as required for the Quality control purposes. In case of any variations in the data, the DM at MUL shall send the queries to the site DM to resolve issues arising.

## 7.3 Priority data checking

Certain forms shall be manually checked by the data mangers as soon as possible to ensure their completeness in the database so that omissions or lack of data is related to this are completely and immediately resolved.

The [APPENDIX 2](#_APPENDIX_2:_Priority) below shows the priority forms for the HIVCORE trial.

# Data entry

## Data Entry Procedures

Data will be entered in real time by the nurses, clinicians and laboratory technicians, where applicable, into the eCRFs on the REDCap database within the presence of the participant. Only data that is from the ELISpots and some other laboratory data shall be entered from the other source documents. As mentioned in the section 7 above, data will only be recorded on the paper CRFs when only in emergency events such as a server fail or internet fail. No abbreviations (apart from those specified in the CRF Completion Guidelines) or summaries of text data will be permitted and ambiguous abbreviations or wording provided by CRC will be queried.

The clinicians, nurses, laboratory technicians and the rest of the trial team required to enter data into the database will initially be trained on the HIVCORE006 CRFs by the Data Management team. They will then also be trained on the use of the HIVCORE006 database use, with regards to logging into the database, completion of the eCRFs, reporting of queries and creating of reports for those required to report. During the data entry, ***utmost clarity*** shall be required and there is required to be a reviewer that can mark the data entered as completed after reviewing of entered data. In case of any discrepancies and out of context values, the reviewer shall be responsible for correcting those before marking the eCRF as complete in the REDCap database. This appies to the forms that involve assesing the eligibility of the participants in the study.

The personnel entering data in the REDcap database shall be capable of marking the data as “Incomplete”, “Complete” and “Unverified” as determined by their user rights for the database. “Incomplete” CRFs are those that are missing some data to be enetered by the data entrant, “Unverified” CRFs are those that have all the data entered by the person enterin g data but have not yet been reviewed by the person incharge of the rewiew. “Complete” eCRFs on the other hand are those that ahave been complete data and have also been reviewed by the personnel incharge of the review process.

The Data entry process at the CRCs shall be monitored by the Data Manager. The teams at the CRCs shall be able to use REDCap’s Field Comments tool to make any notes during data entry; these are stored in REDCap’s audit trail, separate from study data and viewable by the Data Manager and others. Any unresolved discrepancies or instances where the data on the CRF does not comply with the range checks will be logged as queries and added to the data query log in the database, for any particular visit.

On completion of the data entry, the database captures the username of the personnel who etered the data and saves in the audit trail where it is tracked. Where necessary, members shall be required to input their initials after completion or review of data on a form.

All the fields on the eCRFs are required to be entered and in case a field is found and was not clarified or included in the CRF completion guidelines, the data manager at the CRC must be notified as soon as possible so that the Data Management team at MRC/UVRI and LSHTM Uganda Research Unit Entebbe is notified to instil the change to the database and/or to the Completion Guidelines.

For a CRF to be completed successfully in the database at the specific visits, this can be checked via the REDCap trial database which shall only be accessed by the member with the rights as assigned by the database administrator.

## Unavailable Data Items

It is not acceptable for sites to report certain data items as unavailable or not known. If on the REDCap dashboard an eCRF is not reported as “Incomplete” or “Unverified”, the CRC will continue to be queried. The quering will only stop when the CRC provides an explanation that is accepted by the trial clinical team for why the data is not known or not available. A comment can also be added in the comments field on te eCRFs where applicable.

# Quality management

## eCRF validation

The validation of the CRFs will be done on the test version site of REDCap prior to the final release on the live version site ensuring that the eCRFs match the requirements of the protocol. In case of a change of requirements, only the eCRFs affected will be revalidated.

Validation and approval will be recorded on database validation form.

Each data field on the eCRFs will be tested by entering data to determine that:

* Data field accepts different values as required
* Validation flags appear when out of range values/invalid data are input
* All data types function properly
* Skip rules/branching logics function properly (fields may be hidden if not applicable as answered in a previous question)
* Calculation functions work efficiently
* All other parameters are fulfilled appropriately

|  |  |  |  |
| --- | --- | --- | --- |
| **Database** | **System** | **Database build by** | **Database validation by:** |
| Randomisation | REDCap | MRCU Uganda | Project Manager  Senior Statistician  Oxus Technologies |
| CRFs | REDCap | MULS DCC in Entebbe | Project Manager  Senior Statistician |

## Data entry

* + 1. **New staff**

New staff with responsibility for data entry and review in the database will have received database specific training before being added to the delegation log and being granted access to the database to begin data entry and data review. Ideally the new member of staff will have had Good Clinical Practice, Data Protection, and Data Management training prior to the commencement of data entry, however if this has not been possible it will be scheduled to occur within the first month of starting at the site.

## On-site quality control of CRFs

Sites are provided with CRF Completion Guidelines, which is a detailed document that includes details on the quality checks sites should perform before entering data on the CRFs in the trial database. The document should also be used by CRC personnel as an aid when completing CRFs on the database to avoid common errors.

The CRF Completion Guidelines are sent to all CRCs and are expected to be read by all staff authorised to complete CRFs before they begin doing so. An updated version of the CRF Completion Guidelines will be circulated to all sites after any major changes to the trial CRFs.

The Data Managers at the CRCs will be responsible of ensuring that every member that is involved in the entry of data into the database is very conversant with the HIV CORE CRF Completion Guidelines.

# CRF tracking

## Generating/identifying overdue CRFs and/or missed visits

Missed visits will be identified using a do file written using STATA code (from the Volunteer status form). These will then be compiled in the quality control report giving the details of the CRF, participant ID, visit number, the query description and a provision for CRC comment/resolution plus the date of resolution. The quality control report will be generated every week.

## Sending reports of overdue eCRFs

Any overdue CRFs will be documented in a quality control report that will be shared with the sites using Sharepoint. The sites will then access the report using the shared link.

## Managing missed eCRFs

Missed eCRFs that are required for a scheduled visit will be identified in the REDCap database and compiled in the quality control report that will be shared with the sites using Sharepoint. The sites will address the missed CRFs by having them entered onto the Trial database and indicating in the QC report the resolution/comment and the date the query was resolved. In instances where the missed CRF data was not collected this will be indicated by the site and documented in a Note to File.

# Query Handling

## Generating/identifying queries

* + 1. **Internal to the trial database**

The trial database has programmed validations, which check that the entered data are as expected. If data fails a validation, for example if data are outside an expected range, a warning is automatically generated, which will require the person entering the data to enter it correctly, or else to mark the field as validated and add a comment to it. They may also consult the Data manager if it’s a database issue in order to carry out database changes.

* + 1. **External to the trial database**

Where possible, all data querying should be done through the trial database either by automated warnings, and through creating queries on the database, which will be posted on the REDCap Dashboard daily.

If an eCRF cannot be completed onto the database because, for example, it is lacking valid participant details then the CRC should immediately contact the QC/clinic team for a resolution to the query.

Queries meant for the CRCs by the monitor, PI or data management team must be resolved immediately, on the database within 24 hours of being reported so as they can be solved since the data entry is real-time and not many errors are required at this point.

Updates on the CRFs in the database, due to queries, shall be made directly by the Data Management team at the DCC and these changes shall be recorded in the Audit trail in REDCap. Changes to the CRFs in the database shall be communicated to the CRCs immediately after they have been implemented.

Other queries may be generated on data receipt after manual checking of certain priority forms. Resolution of these queries should also be sought immediately by contacting the relevant QC/clinic CRC staff.

* + 1. **Queries raised by the trial lead Data Manager**

The Trial Data Manager will perform consistency checks on the trial data at various points during the trial, including prior to periodic Protocol Trial Safety Reports, DMEC reports and final Trial analyses. If queries are identified within these checks, they are sent to the CRC DM. These queries should then be raised with sites through the database. Once the query has been resolved, this detail should be added to the database by adding a comment to the response of the query in the database.

## Sending/receiving queries

All queries are sent to clinical staff at each site weekly. Clinical staff will have 3 working days to respond to queries, and all queries are to be resolved within one week.

## Handling irregular query responses

If a query is not clear to the CRC staff, the query should be edited by the personnel who raised it to make it clearer in the REDCap trial database. Alternatively, clarity can be provided over the phone or through email to ensure all queries are resolved.

# Data management, documentation and curation

## Managing, storing and curating data.

Paper documents such as informed consent documents, locator forms, and laboratory documents will be filed and stored in lockable cabinets in designated rooms at the CRC study clinics. Study documents that contain participant identifiers (informed consent documents, locator forms) will be stored separately from other documents. Site staff will be assigned user-controlled level access rights that will allow them to perform data entry. Electronic data will be stored on the secure server at the DCC (MUL) in Entebbe, Uganda.

The database will be backed up automatically every day. The backup process is automated and not required daily user intervention. To ensure data is secured and backed up, validation is being performed at the end of every month, and any failed backups are being investigated and resolved with the highest priority.

Database backup and recovery procedure is designed based on:

* Database size
* Backup media available
* Data management system used
* Recovery requirements
* Error detection

## Metadata standards and data documentation

A data dictionary will be compiled for the study dataset. The data dictionary will give a complete description of each table in the database with a list of variable names, variable descriptions, and the codes used. Additional metadata will include the study title, investigators, dates, and locations of data collection, methods, study procedures, key words, and definitions.

## Data preservation strategy and standards

Electronic data will be stored on the database server and is only accessible by authenticated users through the local area network. Physical access to the database server is only restricted to the IT Manager and Data Manager. Data is backed up daily to a secure, off-site location.

The paper (where applicable) and electronic study data will be retained for at least 15 years after the study ends, at which point it will be securely destroyed.

# Storage of participant related documents

## Prior to data entry

All paper documentation containing participant data, for example consent forms, CRFs (where required), emails and query forms, should be kept under lock and key when not in use.

## After data entry

All participant paper information shall be kept in participant binders. Filing of other entered source documents will be done by visit number and the different visits separated using separators where available.

## Unblinding

### Scheduled unblinding

Scheduled unblinding will occur at the end of the trial; at this point both trial staff and participants will be unblinded.

### Unscheduled unblinding

Alternatively, circumstances may arise in which unscheduled unblinding is required for one specific participant. Examples of this include when a subject in a blinded study has a Serious Adverse Reaction (SAR) or requires medical intervention which would be influenced by whether or not they have received the investigational product/control.

In this circumstance, the site Principal Investigator will make the decision to unblind after consultation with the site Medical Monitor, and the un-blinded information should be restricted to a small group of individuals involved in clinical management/medical treatment of the volunteer (e.g., treating physician) and the blind must be maintained for those responsible for the study assessments.

Once the decision to unblind has been made, this will be communicated to the DMEC, the study statistician, or, if they are not available, the data manager.

Details of any case of emergency unblinding will be recorded in the comments section on the respective form e.g. AE form

# Data extraction

Statistical analyses will be performed using the dataset extracted from the trial database and read into STATA statistical software.

Data extracts function from REDCap will be restricted only to personnel with appropriate access. All extracts are to be password protected and saved in a secured location.

Only specified personnel will have ability to extract data from REDCap and access to date extract folder on site.

The statistician will extract the data regularly to prepare for the reports and the data extract will be saved on the secured location, which has limited access to the study statisticians only.

Data extracts should not be re-circulated. Personnel who wish to have access to data extracts should contact the data manager(s) or the individual who did the data extract directly for the latest and most appropriate extract.

## 

## Analyses

## Final analysis

Data extraction prior to the final analysis will be performed after database lock.

## 

## Data quality checks prior to analysis

On the agreed date of data extraction for analysis, the statistician will check with the Data Manager at the DCC that the data is ready. When all parties have agreed that the data are sufficiently accurate and complete for analysis, and that no further querying or chasing of data are required, the lead Data Manager will download the data and share the downloaded data with the trial statistician.

## Data extraction procedure

On the agreed date of data extraction, the statistician will agree with the trial team a time when the data will be extracted. No other users should access the database during this time.

All data will be stored in named and dated network folderswhich are accessible only to the trial statistician.

As part of each analysis, the trial statistician will run routine consistency checks on the trial data. These will be resolved according to the process described in section 11.1.3 of this document.

Each analysis and associated consistency checks may produce a list of items, which do not work as they should in the trial database. This may necessitate changes to both the database eCRFs.

# Monitoring

All sites shall be periodically monitored by IAVI and monitoring reports shall be written. Details of the overall monitoring plan for the trial can be found in the Monitoring Plan/Quality Management Plan approved by the trial clinical team.

## Regular Reports

Weekly study progress reports and safety reports will be generated using STATA and will be available to the trial clinical and PTSR team through email. The Safety reports will include listings of serious adverse events and events that informed a decision to interrupt or discontinue vaccination. Only new and open events will be included.

## Ad-hoc Reports

Ad-hoc reports will be created by the trial statistician and shared with the relevant parties.

# Data freeze and database lock

## Data freeze

Data freeze is the step carried out before data lock that informs the CRC investigators and monitors that the data management team has validated the data and are confident that they are in a position to carry out data lock. Before data freeze is carried out, the study monitors should confirm that they have completed monitoring, including source document verification as specified in the monitoring plan, and have been in contact with the Data manager at the DCC and are confident that all the queries that they have raised have been resolved. Data freeze will then be carried out, after which the study CRC coordinators and study monitors then have the opportunity of raising any outstanding queries that they believe may not have been resolved.

## Data quality checks prior to lock

As a locked database can only be updated in extraordinary circumstances, comprehensive efforts need to be made to ensure that data are as clean as possible. The procedure for database lock should not be commenced until the following data quality conditions have been met:

Before the database can be frozen or locked, the study data management team will carry out a final review of the study database in order to ensure data quality.

In brief, completeness checks will be carried out in the databases firstly to ensure that all the eCRFs have been completed (i.e. that all CRFs for a particular visit have been entered if the volunteer attended that visit), secondly by ensuring that there are no missing fields in any of the records corresponding to all of the CRFs and finally carrying out a formal quality control check on 100% of the critical data (consisting of all reactogenicity data, all adverse events, data relating to HIV serology, data from the post-screening visit contact forms, IP administration forms, risk assessment forms, and socio-demographic data).

Data quality will be improved using the usual methods (see sections 10, 11 and 12 of this document for more details), and repeated until the above conditions are met.

## Database lock

The trial database will be locked before unblinding and the final analysis in accordance with the MRC/UVRI & LSHTM Database Lock SOP.

## Database lock procedure

Once the data quality conditions set have been met and this has been documented, the statistician will request database lock by completing a Database Lock request form and submitting this to the lead data manager, who will then arrange the lock as agreed in that document. The statistician can then access the extracted data for the final analyses from the locked database according to the procedure detailed in the Database Lock SOP.

# Archiving

After the study is completed and the final report has been prepared and approved, all relevant paper documentation including CRFs (where necessary) will be archived in accordance with the Long-Term Storage Procedures for Paper Documents.

The Trial manager will work with the trial Data Manager to arrange the archiving of the trial database when they have determined that accessing the database is no longer required for any future analyses, following MRC/UVRI and LSHTM Uganda Research Unit procedures.

# Data security and confidentiality of potentially disclosive information

## Formal information/data security standards

The MRC/UVRI and LSHTM Uganda Research Unit does not currently use any formal information/data security standards. The study staff will ensure that the participants’ anonymity is maintained. The study will comply with the Data Protection Act 2018, which requires data to be anonymised as soon as it is practical to do so.

## Main risks to data security

The main risks to data security are data loss, unauthorised access, and breach of confidentiality. The following measures will be undertaken to safeguard against these risks:

* Paper documents will be stored in lockable cabinets in dedicated rooms at study sites with controlled access to the central trial team and authorized personnel.
* The database will be backed up daily. The database will be password protected and access will be limited to relevant project staff. The head data manager will have exclusive rights to the database. Database access for other members of the study team will be determined by their responsibilities on the study (e.g. data entry, query resolution, data monitoring etc.).

Study monitoring will be conducted by IAVI, study investigators will permit inspection or audit of the study facilities and all study-related records by relevant regulatory authorities, the local IRB/IEC and/or representatives of the Sponsor.

**Data security arrangements:**

* Anonymised Data extracted by Sponsor (Personal Identifiable Information (PII) protection)
* Electronic Data Transfer rules (Password-protected files, no PII data in emails)
* All systems for the study will be access controlled
* Password protected access to REDCap (Raising password confidentiality awareness)
* Electronic Data Capture (EDC) User Agreement (Appendix 1)
* Daily database backup
* The trial database will log who has entered what data for a participant to provide a security audit trail for the data

.

# Data sharing and access

The Sponsor (University of Oxford) owns trial Data and can share the study data upon request:

* Data can by requested with Data Manager (written request)
* Decision about sharing data will be taken by Chief Investigator.

# APPENDICES

## APPENDIX 1: Data Management Plan Review History

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Reviewed By**  **(Print name)** | **Revision type** | **Detail**  **(significant changes from previous version)** | **Previous**  **Version** |
|  |  |  |  |  |

## APPENDIX 2: Priority forms

| **CRF/Database** | **Reason why considered critical** |
| --- | --- |
| AE | Primary |
| Child Follow up | Primary, Clinical Data |
| Reactogenicity | Primary |
| Sample Summary | Primary, Exploratory |
| Visit Contact Documentation Follow up | Primary, Exploratory |
| Visit Contact Documentation IP Administration | Primary |
| Visit Contact Documentation | Primary |
| Concomitant Medications | Protocol Compliance |
| Eligibility Check List | Protocol Compliance |
| Deviation Report | Protocol Compliance |
| Enrolment | Protocol Compliance |
| Hematology | Protocol Compliance |
| HIV Risk Assessment | Protocol Compliance |
| HIV Virology Serology | Protocol Compliance |