1.0 Purpose

The purpose of this document is to outline the data management procedures at the site(s) and the related responsibilities for the PB SAM study.

2.0 Study Objectives & Design

The objective of this study is to determine whether treating ill severely malnourished children with pancreatic enzymes or bile acids improves mortality. We will conduct a double blind, randomized clinical trial in a 2x2 factorial design in hospitalized severely malnourished children. We will treat participants with paediatric formulations of pancreatic enzymes, bile acids, both or placebo for 21 days. Participants will be followed up daily during their hospital stay and on day 21 and 60 after enrolment.

3.0 Study sites

Participating sites are:

1. Kilifi County Hospital, Kenya;
2. Coast General Hospital, Mombasa, Kenya;
3. Mbagathi Sub-County Hospital, Nairobi, Kenya;
4. Migori County Hospital and Ombo Mission Hospital, Migori, Kenya;
5. Queen Elizabeth Central Hospital (QECH) Blantyre, Malawi;
6. the International Centre for Diarrhoeal Disease Research Hospital, Dhaka, Bangladesh (ICCDR-B); and
7. Mulago National Referral Hospital, Kampala, Uganda.

4.0 Responsibilities

The roles and responsibilities in relation to data management are outlined in table below. The Data Manager is responsible for all day to day data management processes. The Senior Data Manager has oversight of the data management processes. The Data manager is the first line of contact and will ensure tasks are carried out appropriately, and in consultation with the Senior Data Manager where necessary. The Data Manager will liaise with the Study Statistician to ensure routine reports are produced within the required timelines, providing data as required.

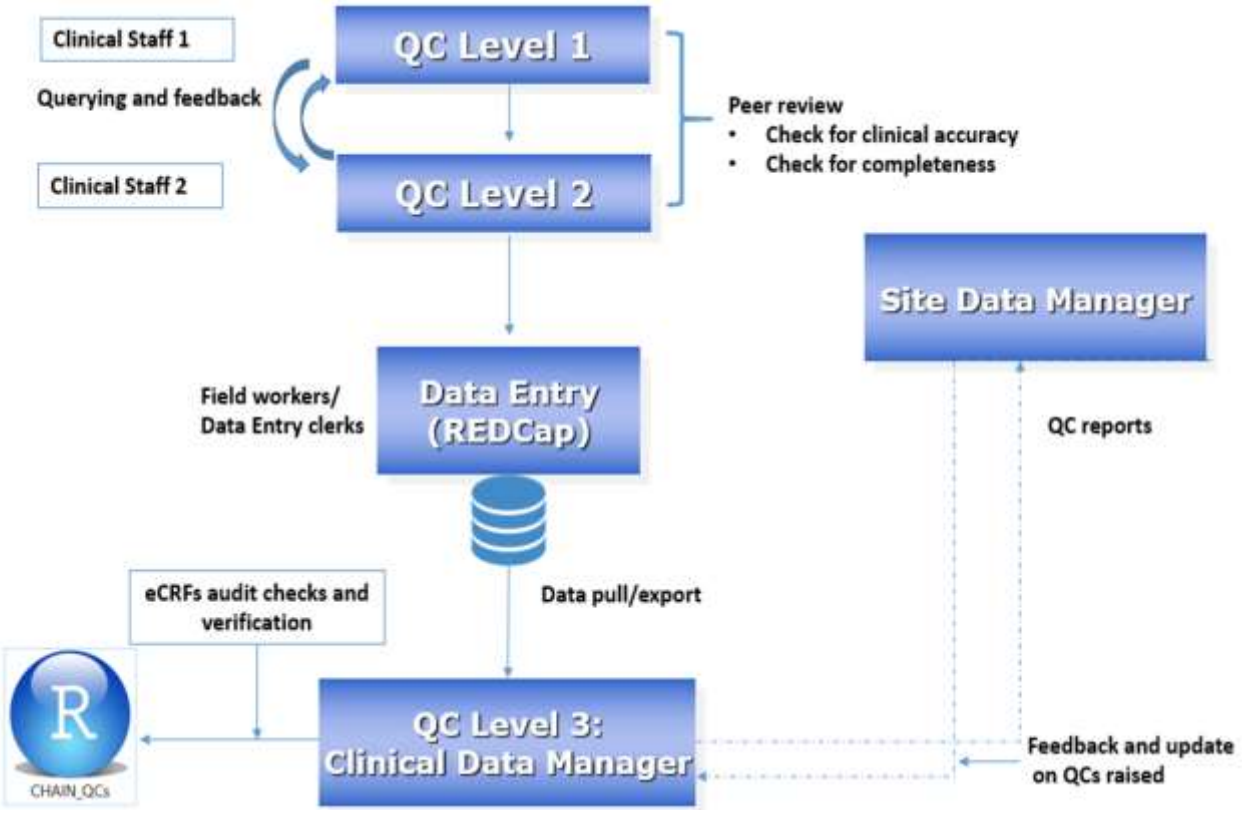
Study Title Roles and Responsibilities for Data management

Personnel Responsible	**Role**
Site Data Manager	Data entry, checking and uploading data using [ALEA] and KIDMS databases.
Senior Data Supervisor	Monitoring of query resolution and maintenance of study documents at co-ordination center. First point of contact at the central co-ordination on data entry and queries support.
Data Manager	Both Laboratory and Clinical Data extraction, preparation of data for interim reporting and daily monitoring of data quality including generation of queries for the same. Responsible for cleaning activities with sites and implementing data quality checks/criteria at the central co-ordination team.
Study Statistician	Analysis of primary and secondary outcomes

Senior Data Manager	Data and systems overall co-ordination. Service level agreements with suppliers and ensuring availability of systems for data management hosted by the central co-ordination team. Responsible for all data cleaning, querying and resolution activities.
Principal Investigator	Overall quality of study data and for ensuring that all applicable staff members follow this DMP.

5.0 Data Processes

Figure. Data flow in PB SAM Study



5.1 Data Collection

The PB SAM study will adopt a data management approach of paper based data collection and remote data entry into the study central database. Data collection will be done by completing the paper Case Report Forms (pCRFs) at each site by the site investigator/clinician. The pCRFs will be entered into the Central Data Management System (CDMS) at the site using electronic Case Report Forms (eCRFs) via remote data entry by the data entry clerks/ field workers.

5.2 Data collection instruments

Data will be collected on paper Case Report Forms. The table below highlights the crf names.

CRF name	Time to fill	Who fills	Comments
PB-SAM Enrolment	During recruitment	Site clinician/field workers	Participant meta data, to be collected together with sample request form for samples collected at enrolment.
PB-SAM SAE	Whenever there is a serious adverse event. See protocal for guidance	Site clinician	Details of the SAE and any actions taken
PB-SAM Toxicity	During a toxicity event reported.	Site clinician	Details of the toxicity claim/report
PB-SAM Daily Record	Once everyday for the period of hospitalization	Site clinician/field worker	
PB-SAM Discharge	During discharge event	Site clinician/field worker	
PB-SAM D21 Follow up	During D21 follow up event	Site clinician/field worker	
PB-SAM D60 Follow up	During D60 follow up event	Site clinician/field worker	
PB-SAM Conclusion	When study exit has been triggered either through: death, withdrawal, study completion or lost to follow up after expiration of follow up period	Site clinician/field worker	

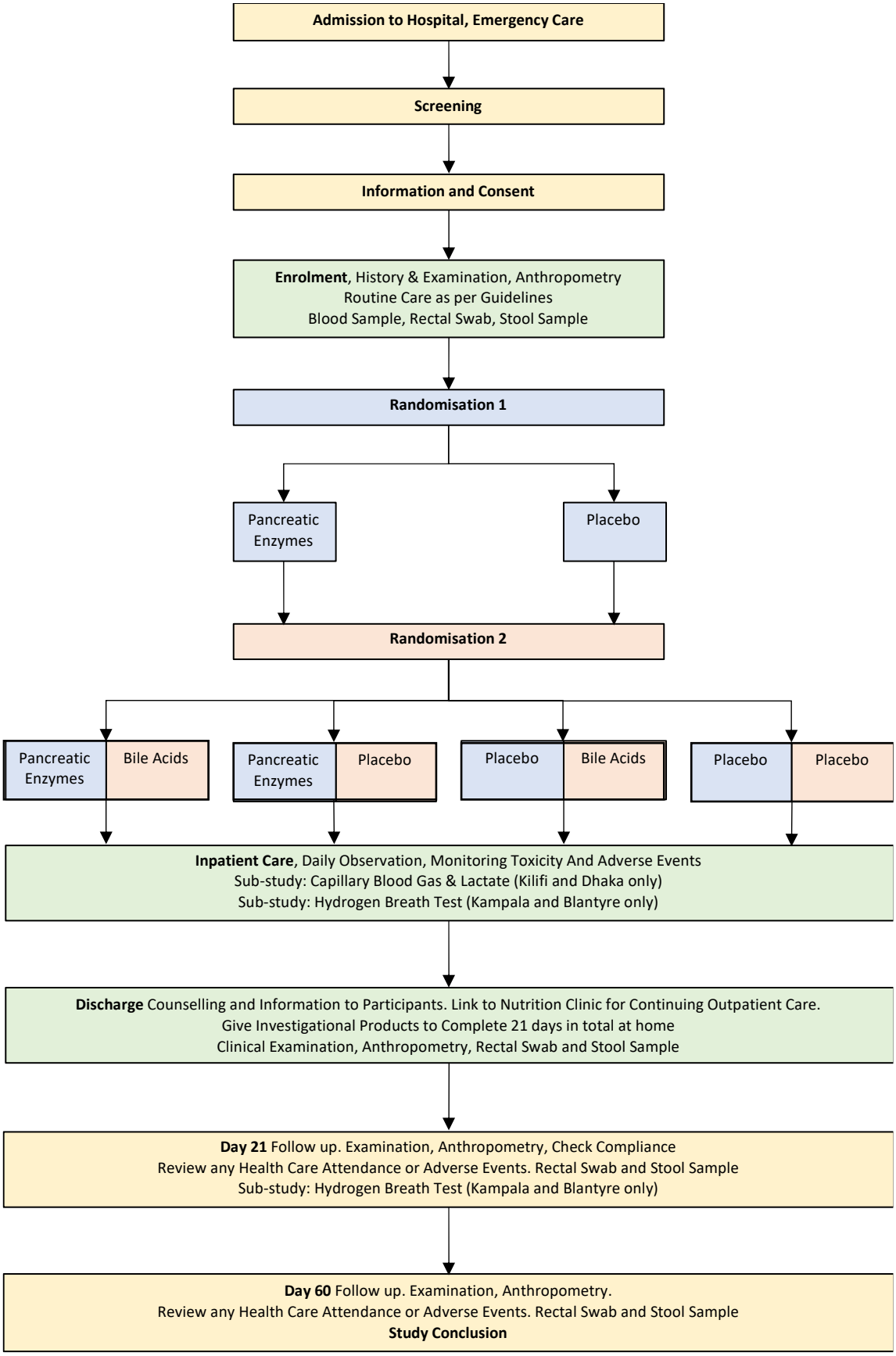
5.3 Accessing Database online

The link to the online database will be: _____.
Only authorized persons will be granted access to the right site/data area.

5.4 Data Entry Procedures

Data entry will be done at each site. Once the data entry to the eCRF is completed, the respective form should be marked as “Complete” by selecting “Complete” from the form status variable. Data entry should be done after each study event.
Data entry should be done within 24hrs of collection. Any backlogs will be monitored and sent to the site as queries.

5.5 Study activity flowchart



6.0 Quality Management

6.1 QC of pCRFs at site

At site, the study clinician will review the pCRFs checking for missed items, inconsistencies, outliers or other errors. Following the QC process on a pCRF, if any additional errors are identified during the data entry process or later, the appropriate staffs who identify these errors will flag the error with a coloured tab. These errors will be corrected by the person who made the original entry by striking across the original entry, entering the correct information followed by her/his initials and current date.

6.2 Central QC of eCRFS

The quality of data will be checked at each of the participating sites, using appropriately chosen tools by the site. Inbuilt data validation checks will be used for screening quality of data entry. Additional validation checks will be performed from the central data repository using data quality modules and routine data checks on data extracted using R statistical software.

7.0 Query Handling

7.1 Sending/Receiving queries

- All Queries will be generated by the Data Manager (Clinical and Lab) using R scripts and visualised in a shiny Dashboard.
- The queries on the dashboard will be posted in a Task Management Tool (Orange scrum) by Senior Data Supervisor for sites to resolve.
- Sites will review and resolve all query on a weekly basis. Any concerns the Data Manager may have with the efficiency of this process will be noted and discussed at the Data Managers weekly meetings.

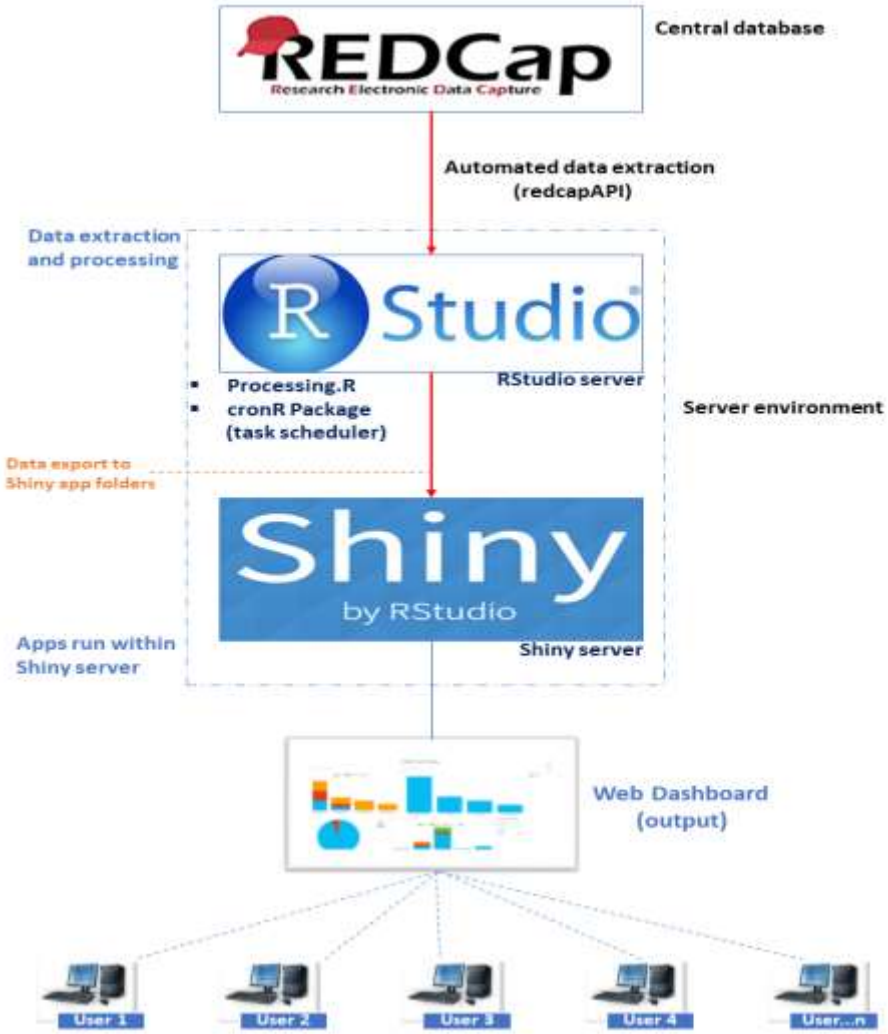
7.2 Handling query responses

- The respective site data managers will resolve all the discrepancies/queries raised in the Orangescrum platform.
- All query resolution will be tracked and monitored and action taken within five working days.

8.0 Reporting

8.1 Regular Reports

Reports will be automatically generated on the shiny dashboard. Accessing the dashboard is through the following link: http://reports.chainnetwork.org**.**

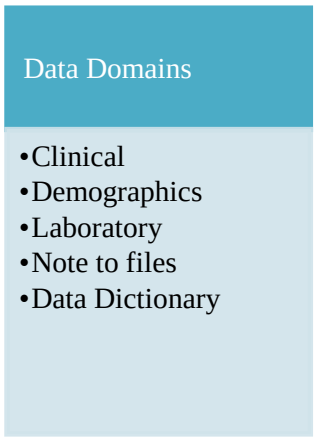


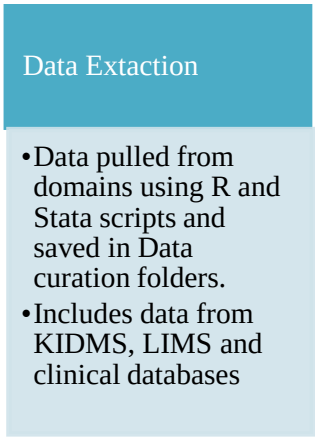
8.2 Reporting Tools

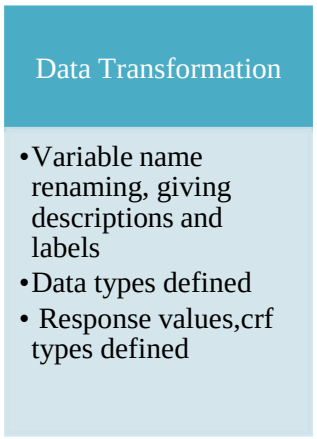
Redcap API	To connect to the Redcap database for automatic extraction
Rmysql	For direct connection to Mysql database in the case of non-redcap implementations
R/Rstudio	Writing business rules and reports
Shiny server	Putting together the output in a web
Apache	Hosting the webapp through proxy server (adding a layer of security

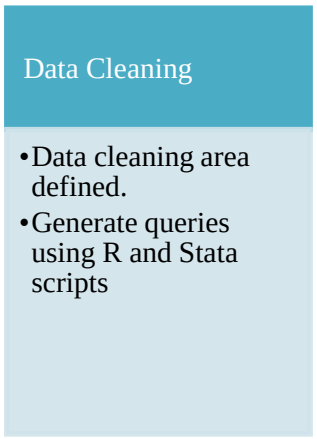
9.0 Data Cleaning

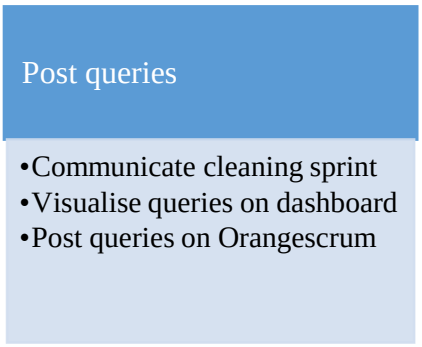
The Data cleaning team will extract data from database from in select domains. Interim data extractions will be undertaken on monthly basis or as required by investigators. The table below highlights the data cleaning activities conducted on every extraction.

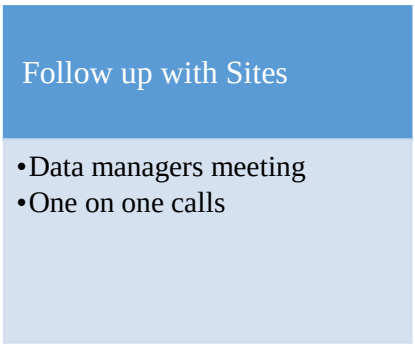


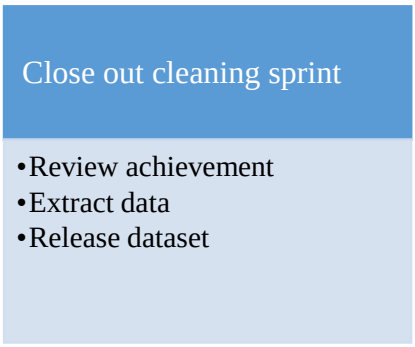












The interim and final analyses will be performed by the Study Statistician assigned using STATA and R software. Database lock will be agreed in advance with the Study Analysis Group to ensure this is done when data entry and query resolution is as complete as possible. Password protected copies of all CSV files will be stored electronically within the curation folders.

10.0 Communication procedure

Queries	These will be communicated via Orangescrum and Shiny Dashboard.
Skyping	One on One calls between Site Data managers and Central team will be made via skype to discuss any challenges or feedback.
Emails	Emailing list for all site data managers will be available for quick communication on urgent information.
Weekly meetings	Weekly meetings for all site data managers will be done to give weekly updates, feedback and discuss challenges.
TGHN Platform	All updated CRFs, SOPS will be centrally available on the TGHN platform. https://chain.tghn.org/study_resources/NEOBAC-crfs/
Reports to Leadership	Reports will be communicated via dashboard and also through Data analysis calls

11.0 Database Closure/Lock

The study database will be locked before the final analysis. All data will be cleaned prior to database lock and queries resolved where possible. Any un-resolvable queries will be closed as 'closed-unresolved/unobtainable'.

12.0 Data Backup and Recovery Procedures

The KWTRP have implemented a database replication, which keeps a real-time copy of the main database on a remote/slave server. With this in place, if something happens to the primary database, it will be much easier to get the database back up and running with current information.

The fail-over/slave server will be maintained for restoration of the clinical system in the event of a disaster that destroys the primary database server. The server will immediately take over normal data query operations. There is very minimal data loss, with this setup in place, as up to the last 5 minutes of data can be recovered.

13.0 Data Storage and Archiving

The study database will be archived on the Kemri-wellcome Trust servers according to the unit ICT policies after database lock**.** Guidance will be provided on the correct procedures to undertake to be able to access the archived data.

After study conclusion, the binders will be transported to the central storage facility (CTF in Kenya). Other participating sites will store and archive their source document according to participating institutions laid out clinical study data storage and archiving policies. Copies of all completed eCRFs/CRFs and source documents will be archived for a minimum of XXXXX of years following completion of the Study. The Documents should be stored in such a way that they are complete, accurate and remain legible. Any alterations made should be traceable.

CRFs and source documents should be archived in an appropriate locked room, cupboard or filing cabinet with adequate fire protection (sprinkler systems), protection from water, humid conditions and pests. The room, cupboard or filing cabinet should have secure access by authorised personnel only.

14.0 Document history

Version	Author	Approved by	Dated
1.01 PB SAM Data Management SOP	Narshion Ngao		

15.0 Site training record

All sites are required to maintain a master copy of this SOP that documents the site staff that have been trained on this SOP.

Document History

Version No.	Trained staff initials	Signature of trained staff	Date	Trainer's Initials
1.01	KDT	Example row	1st Nov 2020	DM