Title	Data Management Plan SOP	
Version	2.0	
Prepared by	Name:	Eunice Kahindi
	Signature:	……
	Date:	02 September 2019
Reviewed by	Name:	Don Obote
	Signature:	……
	Date:	03 December 2019
Reviewed by	Name:	Amos Fondo
	Signature:	……
	Date:	03 December 2019
Approved by	Name:	Narshion Ngao
	Signature	……
	Date	06 January 2020

Sponsor: Bill & Melinda Gates Foundation

Principal Investigator(s): Dr. James Berkley, Dr. Judd Walson, Dr. Robert Bandsma

Table of Contents

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 CHAIN Network study.

2.0 Study Objectives & Design

The objective of the CHAIN Network Study is to improve understanding of underlying characteristics and processes that determine increased risk of mortality, readmission to hospital and inadequate nutritional recovery amongst young children admitted to hospital with acute illness. These include biological (e.g. infection, the immune system and metabolism), nutritional (e.g. breast-feeding and diet), related to the health system (e.g. how children are assessed and managed) and social and behavioural factors (e.g. economic constraints, feeding and health seeking behaviour).

The primary endpoint is:

Participant's status after six months, classified as:

i) Death in hospital
ii) Death after discharge
iii) Poor nutritional status after 180 days (SAM or MAM)
iv) Readmitted to hospital within 180 days
v) Good recovery (no death, readmission, SAM or MAM)

The frequency of these endpoints will be described along with their risk factors.

The study aims to recruit approximately 600 children between 2 month and 23 months of age at each of seven sites:

- At each site, 500 will be acutely ill children admitted to hospital. Children with chronic illnesses that are not related to under nutrition or infection will be excluded. Eligible patients will be approached for consent at admission to hospital:
    - 200 children will have severe acute malnutrition.
    - 200 will have moderate acute malnutrition.
    - 100 will not have acute malnutrition.
- At each site, 100 non-acutely ill children from the community catchment area (non-hospitalized) of each participating health facility.

3.0 Study sites

The study will take place at a network of rural and urban health care facilities in Kenya, Malawi, Uganda, Pakistan, and Bangladesh,

In Kenya, the Kilifi County Hospital, Mbagathi District Hospital and Migori County Hospital will participate. The Kilifi and Mbagathi sites are affiliated and run by KEMRI-Wellcome, and the Migori County Hospital will be run by a UW-Kenya team.

4.0 Responsibilities

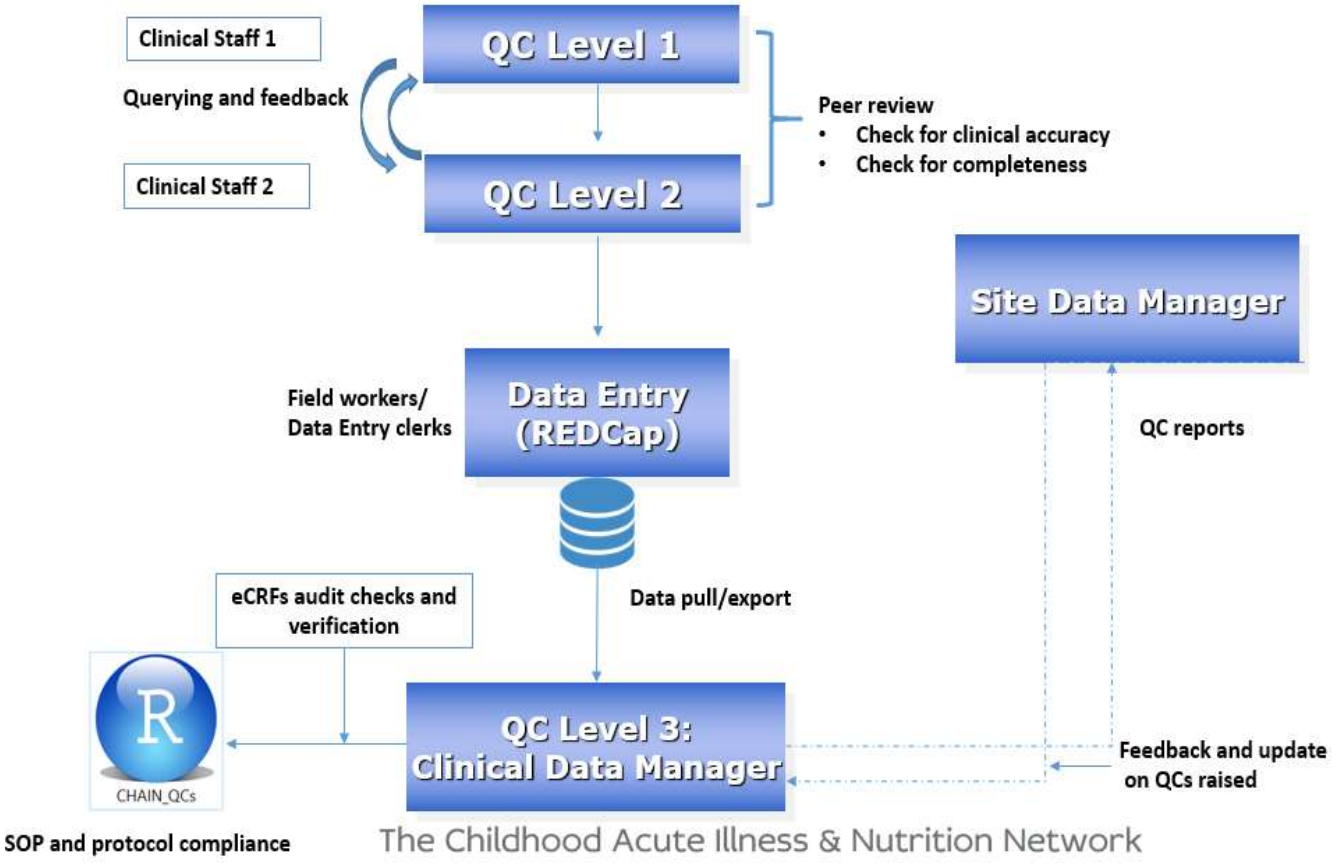
The roles and responsibilities in relation to data management are outlined in table below. The Data Manager Clinical is responsible for all day to day data management processes. The Project Manager has oversight of the data management processes. The Data manager is the first line of contact and will ensure tasks are carried out appropriately, and in consultation with the Project 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 REDCap database.
Senior Data Supervisor	Monitoring of query resolution and maintenance of study documents
Clinical Data Manager	Clinical Data extraction, preparation of data for interim reporting
Lab Data Manager	Laboratory Data extraction, preparation of data for interim reporting
Study Statistician	Analysis of primary and secondary outcomes
Senior Project Manager	Data and systems overall co-ordination
Principal Investigator	Overall quality of study data and for ensuring that all applicable CHAIN staff members follow this DMP.

5.0 Data Processes

Figure. Data flow in Chain Network Study



5.1 Data Collection

The CHAIN 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 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
Baseline	After consent	Site clinician/field workers	Scan primary source documents at discharge
Daily Review	Every day during hospital admission.	Site clinician/field workers	
Follow Up (Day 45,90,180)	At every follow, up visit to hospital or participant's residence.	Site clinician/field workers	
Study conclusion	On discharge/death/voluntary withdrawal/absconded/lost to follow up	Site clinician	Ensure every study exit is captured on this CRF

5.3 Accessing Redcap Database online

The online version of the database can be accessed using the following link
https://production.chainnetwork.org

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. Each study activity is designed as study event on the database (Admission, Daily Review, follow up and Study conclusion), data entry will be done after each study event.

5.5 Study activity flowchart

	SCREENING & ELIGIBILTY	**ENROLMENT**	**DAILY INPATIENT REVIEW**	**48 HR review**	**DISCHARGE**	**Day 45**	**Day 90**	**Day 180**	**Day 14 post-discharge**	**Re-Admission to Hospital**
Standard case management	X	X	X	X	X	X	X	X	X	X
Child admitted to hospital	X									
GIVE STUDY information	X									
Informed Consent		X								
Anthropometry	X	X	X	X	X	X	X	X		X
data collection		X	X	X	X	X	X	X		X
home visit					X					
Faecal sample		X		X	X	X	X	X		X
Blood sample		X		X	X	X	X	X		X

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, if not using REDCap, and REDCap's inbuilt data validation checks for those that will use REDCap for data entry. Additional validation checks will be performed from the central data repository using REDCap's data quality module and routine data checks on data extract 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.

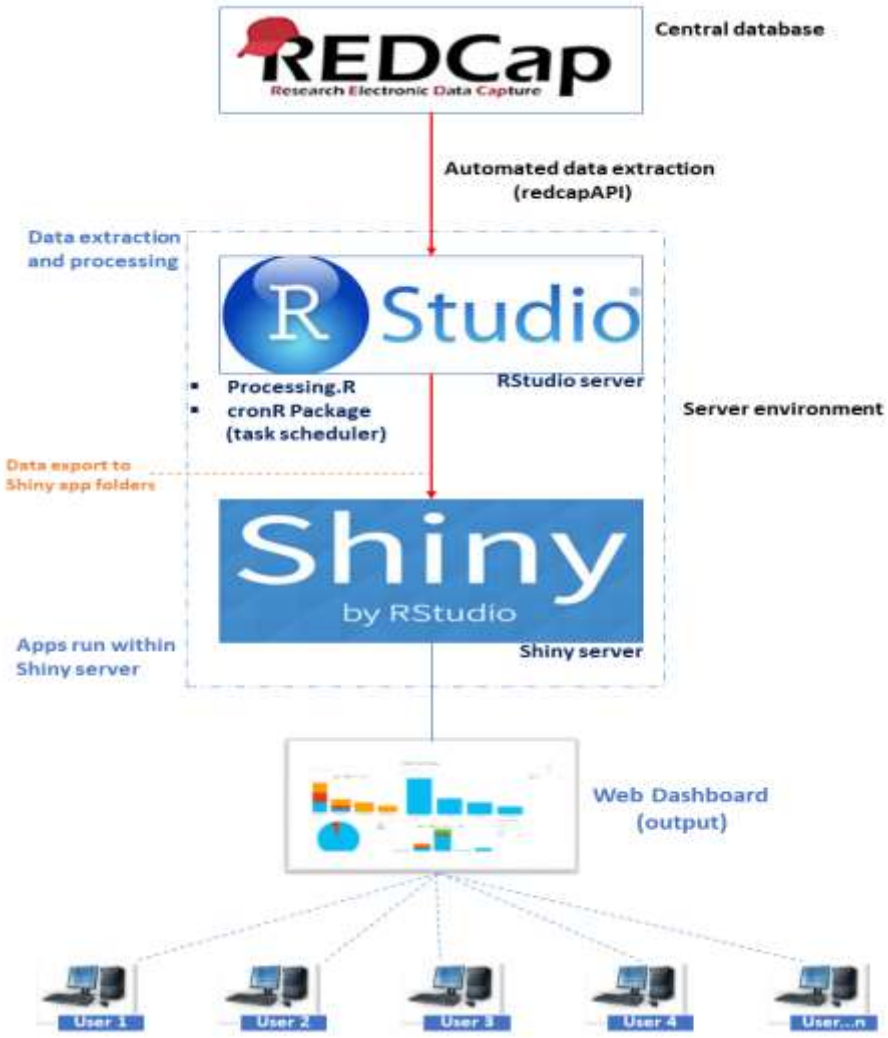
7.2 Handling query responses

- 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 Curation and Analysis

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

Data Domains

- Clinical
- Demographics
- Social&Household
- Laboratory
- Note to files
- Data Dictionary

Data Extaction

- Data pulled from domains using R and Stata scripts and saved in Data curation folders.

Data Transformation

- Variable name renaming, giving descriptions and labels
- Data types defined
- Response values,crf types defined

Data Cleaning

- Data cleaning area defined.
- Generate queries using R and Stata scripts

Post queries

- Communicate cleaning sprint
- Visualise queries on dashboard
- Post queries on Orangescrum

Follow up with Sites

- Data managers meeting
- One on one calls

Close out cleaning sprint

- Review achievement
- Extract data
- Release dataset

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/chain-crfs/
Reports to Leadership	Reports will be communicated via dashboard and through Data analysis calls
Annual meetings	Annual meetings will be held annually to announce results

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 REDCap system in the event of a disaster that destroys the REDCap's 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 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.


The Childhood Acute Illness & Nutrition Network