Ver.: 1.01 SOP No. CHN 45

Data Management Plan Version 1.01 05Sep2016

Approved by: _____ _____

Dr. James Berkley Date
Principal Investigator

Reviewed by: _____ _____

Amos Fondo Date
Data Supervisor (Kilifi)

Reviewed by: _____ _____

Date
Data Supervisor (Nairobi)

Reviewed by: _____ _____

Peace Lamtoo Date
Data Manager (Kampala)

Reviewed by: _____ _____

Emmanuel Chimwezi Date
Data Manager (Blantyre)

Reviewed by: _____ _____

Kalpana Biswas Date
Data Manager (Dhaka & Matlab)

Reviewed by: _____ _____

Imran Ahmed Date
Data Manager (Karachi)

Prepared by: _____ _____

Caesar Olima Date
Central Data Manager

Sponsor: Bill & Melinda Gates Foundation

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

List of Abbreviations

A/C	User account on the study database
CDMS	Central Database Management System
CRF	Case Report Form
CTF	Clinical Trials Facility
DSMC	Data Safety Monitoring Committee
COMREC	College of Medicine Research Ethics Committee, Blantyre, Malawi
CSV	Comma Separated Values
DM	Study Data manager
DMP	Data Management Plan
GCP	Good Clinical Practice
ID	Identification
ICH	International Conference on Harmonization
KEMRI	Kenya Medical Research Institute
KEMRI-WTRP	KEMRI Wellcome Trust Research Programme
LSM	Local Safety Monitor
MUAC	Mid-Upper Arm Circumference
OXTREC	Oxford Tropical Research Ethics Committee
PI	Principal Investigator
SAM	Severe Acute Malnutrition
MAM	Moderate Acute Malnutrition
SOPs	Standard Operating Procedures
R	The R Project for Statistical Computing
QA	Quality Assurance
QC	Quality Control

1 INTRODUCTION

1.1 DATA MANAGEMENT

The CHAIN study will adopt a data management approach of paper based data collection and remote double 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. Any change or correction to a pCRF should be dated, initialled, and explained (if necessary) and should not obscure the original entries. The pCRFs will be double entered into the CDMS at the site using electronic Case Report Forms (eCRFs) via remote data entry by the data entry clerks/ field workers. pCRFs will be printed and stored at each site. 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. The study monitors will ensure that entries relating to eligibility and the primary outcome on the pCRF and eCRF are source verified.

1.2 PURPOSE OF THIS DMP

To define data management and quality control procedures for the CHAIN Study.

1.3 SCOPE / RESPONSIBILITY:

1.3.1 Scope

This DMP applies to all study staff involved in data collection, data management and quality control of the CHAIN study data.

1.3.2 Responsibilities

Study Coordinator, Site Investigator/clinician (at each site), Laboratory Manager and Data Manager (s) are responsible for training study (clinic and laboratory) staff to collect and manage the CHAIN study data in accordance with this DMP.

Study Coordinator, study clinician (at each site), study monitors, Data Manager(s) (designated for QC) and Laboratory Manager (designated for QC of lab data) are responsible for overseeing data QC procedures at the wards, clinic and laboratories related to this DMP.

Data Manager (s) is/are responsible for the overall management of the CHAIN databases, source documents and generation of study progress reports and will work closely with the Study Statistician.

Principal Investigator (s) is responsible for the overall quality of study data and for ensuring that all applicable The CHAIN staff members follow this DMP.

1.4 EQUIPMENT / MATERIALS

- Lockable cabinet (s): at each site for keeping pCRFs and source documents.
- Stationery: Black pens, files etc.
- Computers: PCs/Laptops for data entry at each site.
- Patient folders
- Internet connectivity at each site: configured to access REDCap.

2 STUDY BACKGROUND AND DESIGN

2.1 STUDY OBJECTIVES

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).

2.2 ENDPOINTS

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.

2.3 STUDY DESIGN

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.

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

2.4 STUDY POPULATION

Criteria for inclusion of subjects

- *Children 2-23 months* admitted to hospital.
- Planning to remain within the hospital catchment area and willing to come for specified visits during the 6 month follow up period.

Criteria for exclusion of all subjects

- Unable to tolerate oral feeds while in his/her usual state of health
- Underlying terminal illness that in the opinion of the treating physician is likely to lead to death within 6 months (e.g., cancer, congenital heart disease)
- Diagnosed chromosomal abnormality (syndromically or genetically diagnosed abnormality)
- Primary reason for admission is trauma or surgery
- Diagnosed with a condition that in the opinion of the treating physician is likely to require surgery within 6 months
- Previously included in this study

*Children requiring resuscitation will be defined as those with on-going cardiac or pulmonary arrest or judged to be peri-arrest by the attending physician

Enrolment Stratification:

Group A: MUAC <11.5cm (MUAC <11cm for under 6 month old infants) or kwashiorkor (n=200)
Group B: MUAC 11.5 to <12.5cm (MUAC 11 to <12cm for under 6 month old infants) (n=200)

Group C: MUAC of 12.5cm or more (MUAC 12cm or more for under 6 month old infants (n=100)

MUAC has been chosen as the principle marker of under nutrition because it is generally a better predictor of subsequent survival than other anthropometric measures, is less affected by dehydration than weight-based indices and captures elements of both wasting and stunting – both of which are potentially important determinants of mortality.

Community children:

- Aged 2 to 23 months – purposively sampled in age-stratified groups (under 6 months; 6 to 11 months; 12 to 23 months) (n=100) living in the same community as the acutely ill children that are being recruited.
- Not having an acute illness requiring hospital admission according to WHO and national guidelines
- Absence of known untreated HIV or TB
- Not admitted to hospital within the last 14 days
- Not previously included in the study

2.5 Study Activity Flow chart

	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

3 DATA AND TOOLS

3.1 PRIMARY SOURCE DOCUMENTS

Admission records, daily observation and discharge notes. Admission and discharge notes will be scanned and stored in secure folder at each site. The daily observation notes will be stored in respective participant source document folder at the site.

3.1.1 Case Report Forms

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

3.2 DATA COLLECTION

3.2.1 Screening of the child in the wards:

- The Screening source document is completed
- Consent is obtained
- Routine admission hospital records and daily observation notes to be completed.

3.2.2 If the child is eligible for enrolment:

- Pick the next study ID number in the respective site.

- The Study Number will be entered into the Enrolment Log (see enrolment SOP) by the field worker.
- Primary source documents will be filled by hospital staff/locum nurses.
- Baseline CRFs will be completed by site clinician.
- Enrolment samples will be sent to the lab. No results are expected to be entered on the CRFs.
- Data entry will be done at site. The CRFs will be kept at site until the participant is discharged or dies/withdraws from the study.

3.2.3 For twin or sibling enrolment

Allocate the index sibling subject ID number as usual, then to get the next number for the twin; phone or send an email to study statistician or site data manager who will give the next number.

3.2.4 If the child is ineligible for enrolment:

- The CHAIN's study screening and eligibility log will be used to document the reason for ineligibility of the child.
- No CRFs are filled for ineligible children.

3.3 Study exit

- At discharge, the study conclusion will be completed by the study clinician.
- For the participants who dies/voluntary withdraws/absconds, they will have exited the study and study conclusion should be completed.

4 Data Quality

4.1 Data validation by Data manager

- Manual checks: Visual checks of CRFs with manual review of the data, e.g. medical consistency checks, lab data, inconsistency dates, out of ranges values etc.
- Computerized checks: R scripts will be created by the DM for checking errors after every data extraction before forwarding the data for interim analysis. At the end of the study, the DM will comprehensively validate all the data by using R scripts before forwarding the final data to the study statistician/PI.
- Errors will be reported to the Site Data Manager who will assign the appropriate person for resolution (normally, the person who completed the pCRF)

- Any changes made to the CRF will have to be updated on the Database by the respective site data entry personnel.

4.2 QUERY MANAGEMENT

- The respective site data managers will receive all the discrepancies/queries raised in from the CHAIN's central data repository.
- The central repository data manager will send request for clarification on a data item collected with errors or inconsistency discovered during data validation.
- All query resolution will be tracked and action taken within five working days.
- The closure of a query will be based on information contained in a data clarification form.
- The respective site data manager will keep a record of all queries and resolutions in data query forms.
- After study censure or completion, the participant binder with the filled-in CRFs and source documents will be kept in clearly labelled lockable cabinets at each site until all the data queries have been resolved and the database locked.

5 MOVEMENT OF THE CRF

At the start of the study, the respective site data manager after consulting the study coordinator will prepare CRF binders to be transported to each site. Once the CRFs binders have been prepared, the Study coordinator will arrange how to transport them to each site. The CRFs binders will usually accompany corresponding trial drug.

Request for additional CRFs binders at any site will be received through study coordinator. Once the request has been received, the respective site data manager will prepare the respective CRFs binders and forward them to study coordinator.

5.1 CRF STORAGE AT EACH SITE:

- Each site will have a lockable water proof cabinet for keeping the CRFs and source documents. Keys to the cabinet will be kept at the site investigator's office.

- 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 arching policies.

On a daily basis any documents/ participant binder that leave the data office to be taken to each site will have to be signed in and out on the CRF Transport log.

6 DOCUMENTATION TOOLS

6.1 CRF TRANSPORT LOG

The CRF transport log will show the date of transportation, the participant trial number, items/forms transported, staff code, date of return and comments. Each site will have a CRF transport log which will be completed by site field workers any time They receive new CRFs binders and when they send the completed CRF to central storage facility.

6.2 ENROLMENT LOG BOOK

The enrolment log book will be used to record:

- Child's names
- IP number and or Hospital number
- Study Number
- Enrolment date

The enrolment log book will be updated by field workers after every enrolment. For confidentiality, this enrolment log will be kept under lock and key by each site investigator/clinician.

6.3 PARTICIPANT BINDERS

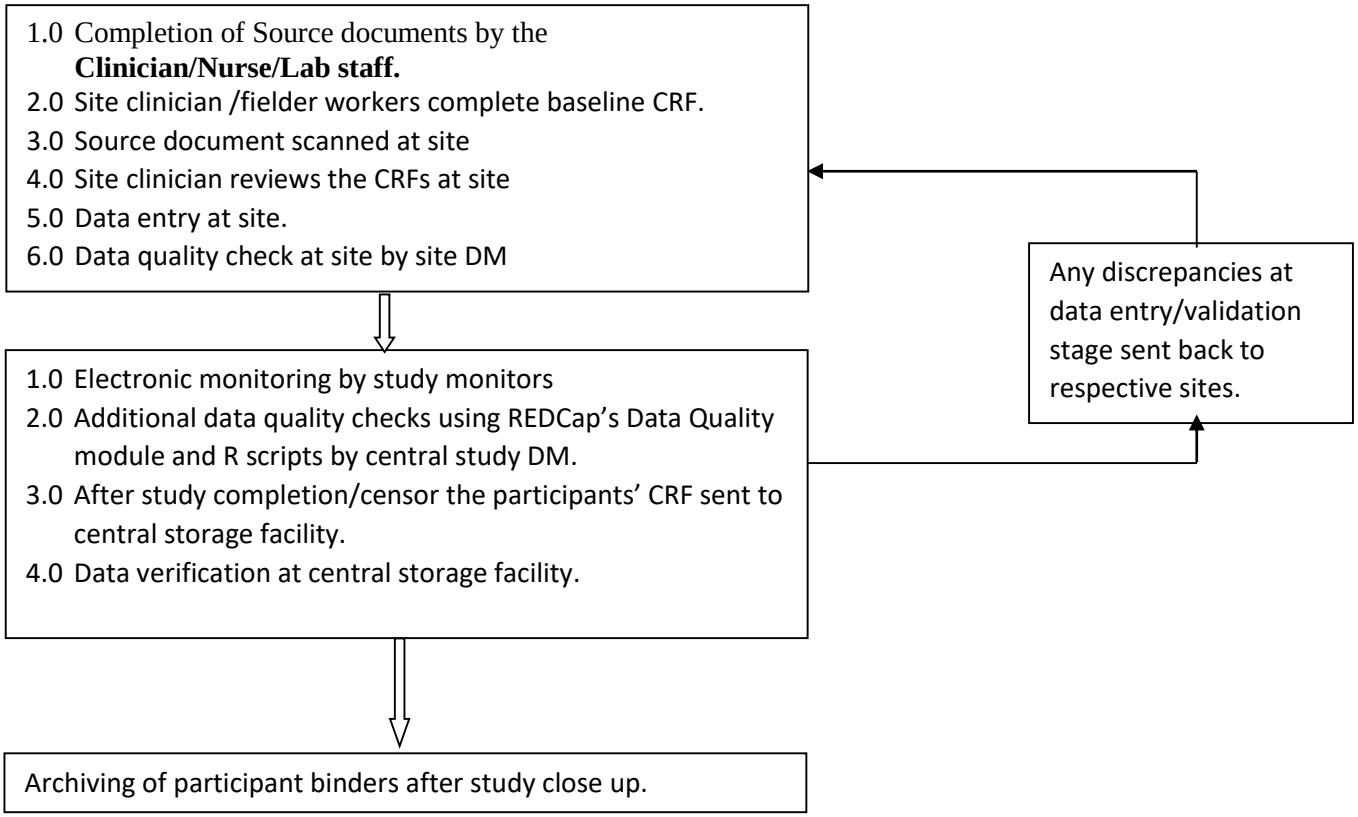
Every enrolled child in the study will have a Participant CRF Binder labelled with the Trial Number assigned to them. The respective site Data Manager will prepare the Participant Binders. The above binders will be stored temporally at the sites up to the study conclusion of each patient. A participant will also have a separate source document binder to hold all the source documents at site. The CRF and source document binders will remain at site up to the time the participant completes follow ups or is censored.

6.4 CONSENT FORMS FILE

All the signed consent forms will be filed in the consent forms file. They will be filed in a reverse temporal sequence with the recent consent forms on top. This file will be kept at each site up to the end of study. In case one file will not have adequate space, a volume2, volume3; volume4 etc. files will be created.

6.5 SCANNED SOURCE DOCUMENTS FOLDERS

The source documents of all the CHAIN eligible participants will be scanned by field workers and copies stored electronically at each site. Each participant will have a file named after their Study number (e.g. CH [10] [001] [001]). These folders will be backed up weekly on an external hard drive kept away from site. Additionally, the CHAIN data folder will be created on backup server (for Kenyan sites) and the participants' folders backed up there weekly.



DATA FLOW DIAGRAM

6.6 CRF Forms and Variables

The Study site clinician/field workers will complete the CRFs. Separate CRF completion notes have been provided. CRFs will be duly completed and if for a known reason the data is not available, a separate variance form (See Appendix 3) will be filled and filed in participant binder for faster resolution of discrepancies arising from the validation checks.

6.7 Data Entry

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.

6.8 QUALITY CONTROL/MONITORING

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. This correction will be recorded on the variance form. Monitoring by study monitors will be done according to study monitors schedule. (See the monitoring SOP)

7 MANAGEMENT OF DATABASE/DATASETS

The CHAIN study data entry and management platform will be available at http://redcap.kemri-wellcome.org, accessible to users accessing outside the KEMRI-WTRP's LAN. User a/c and passwords will be created for all staff with authority to access the database. The DM, who will design the database and host it on the server in Nairobi, will have administrative rights on the entire study database. Data extraction queries will be created that will generate datasets of interest for the purposes of monitoring the quality of the data and the reporting progress of the trial. REDCap provides data export in a variety of formats however for this study we shall export the data using CSV format. The data file will be saved in CSV format and imported into a statistical analysis of supported within the KWTRP. These files will be generated by DM with the study statistician's assistance, and stored in a user-restricted shareable file space. The R scripts will also be generated for producing monthly progress report. The study's central DM will extract data for interim analysis according to Protocol schedule or as directed by PI.

7.1 DATABASE SECURITY

REDCap supports a combination of table-based and Lightweight Directory Access Protocol (LDAP) user authentication. Login query to authenticate users against an LDAP/Active Directory (AD) server for KWTRP staff and table-based authentication for other users ensure only authorised access and single user accounts across the multiple sites used in the study.

7.2 ACCESS RESTRICTIONS

The study PIs will have study director role in the entire study. The DMs will have data management roles in their respective sites; however main DM who will design the database will be in charge of the entire database with rights to add and assign new users in the entire study. Site respective DM will add and assign users to their respective sites. The study statistician will have data specialist roles. Data entry clerks will have data entry person roles in their respective sites with rights to enter data and raise discrepancies. Monitors will have monitors roles. For rights and privileges of each user, see Appendices A4.

7.3 AUDIT TRAILS

REDCap has features to keep a study audit log which keeps history of actions on study subjects and events. The audit log documents all changes made on the database, keeping both previous and current values when changes are made.

Any change or correction to the pCRFs will be dated, initialled, and explained without obscuring the original entry as required by GCP.

7.4 DATA EXTRACTION FOR (DATA MONITORING AND SAFETY COMMITTEE) DMSC

Data extraction queries will be created specifically for the purpose of data required for the DMSC reports. The focus of the DSMC is primarily safety of the study participants; The overall integrity of the study, its continued relevance and ability to answer the primary objective. The Central Data Manager, at Nairobi, will extract the specified data, clean and forward it to study statistician as specified in the protocol or directed by PI or DMSC.

7.5 DATA DICTIONARY

The data manager who will design the database and create a data dictionary explaining and defining all the CHAIN study variables. This will include all the CRFs' variables and all the codes. There will also be an R script file for labelling the data after extraction.

7.6 DATABASE LOCKING

After all eCRFs have been completed and verified, all the data entry forms will be locked by the Central Data Manager following an instruction to do so from either the PI or his designee, after which no data can be added to or edited in the study database. Once this has been done it will be communicated to all relevant parties.

7.7 CONFIDENTIALITY

The use of participant names to identify study related documents will be minimized as much as possible. All source documents bearing the names of participants will be stored in a binder and stored in a locked cabinet. All files will be stored in locked cabinets with access limited to study staff. During working hours, the data centre will be unlocked when the data personnel are present to allow staff access to files to conduct participant visits and perform other required data management functions. Access to the CHAIN database will be limited to the study staff through the use of password protections. No participant names will be appearing on study CRFs or on the database.

8 DATA STORAGE

8.1 PHYSICAL DOCUMENTS

The study pCRFs will be stored temporarily at the respective sites. Respective site DM in collaboration with site investigator (s) will ensure pCRFs are safe at their respective sites. Each site needs to have a water and fire proof lockable cabinets for storing the pCRFs.

8.2 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.

8.3 DATA ARCHIVE PLANS

- In Kenya, pCRFs from the Kilifi and Mbagathi sites will be archived for the numbers of years specified in the protocol at the CTF archive.
- The Malawi site will also archive the pCRFs. Upon conclusion of the study, site study coordinator and DM will ensure both pCRFs and source documents have been archived for the numbers of years specified in the protocol.
- The study database will be archived on the KEMRI/Wellcome Trust ICT servers according to the unit ICT policies after database lock.

9 APPENDIX:

A1 Document change History

A2 Document Copy Control

A3 Variance form

A1 DOCUMENT CHANGE HISTORY

This section is to be completed by The Quality Management or designee

Version Table:

Version 1: Title: Data Management Plan	Dated: 05/05/2016	DMP No.: 1.00	No. Pages: 16
Version 2: Title:	Dated:	DMP No.:	No. Pages:
Version 3: Title:	Dated:	DMP No.:	No. Pages:

DMP Review and Updating Logs

DATE	NAME OF REVIEWER	SIGNATURE	REASON FOR REVIEW

A2 DOCUMENT COPY CONTROL

DMP DISTRIBUTION

DATE	SECTION	RECEIVED BY

A3 Variance Form

The CHAIN Study Number

C	N								

Date	CRF page/Section	Discrepancy description	Remarks	Initials